

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS
Departamento de Zoología y Antropología Física



TESIS DOCTORAL

**Historia evolutiva de la población andaluza basada en su herencia
materna y su relación con el poblamiento humano del espacio
mediterráneo**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Candelaria Lucía Hernández de la Fuente

Directora

Rosario Calderón Fernández

Madrid, 2016

Universidad Complutense de Madrid
Facultad de Ciencias Biológicas
Departamento de Zoología y Antropología Física



UNIVERSIDAD
COMPLUTENSE
MADRID

**HISTORIA EVOLUTIVA DE LA POBLACIÓN ANDALUZA
BASADA EN SU HERENCIA MATERNA Y SU RELACIÓN CON
EL POBLAMIENTO HUMANO DEL ESPACIO
MEDITERRÁNEO**

Memoria para optar al grado de Doctor presentada por

Candelaria Lucía Hernández de la Fuente

Directora

Rosario Calderón Fernández

Madrid, 2015

Universidad Complutense de Madrid
Facultad de Ciencias Biológicas
Departamento de Zoología y Antropología Física



**UNIVERSIDAD
COMPLUTENSE**
MADRID

**HISTORIA EVOLUTIVA DE LA POBLACIÓN ANDALUZA
BASADA EN SU HERENCIA MATERNA Y SU RELACIÓN CON
EL POBLAMIENTO HUMANO DEL ESPACIO
MEDITERRÁNEO**

Tesis Doctoral

Presentada por

VºBº de la Directora

Candelaria L. HERNÁNDEZ DE LA FUENTE

Rosario CALDERÓN FERNÁNDEZ

Madrid, 2015

**ROSARIO CALDERÓN FERNÁNDEZ, CATEDRÁTICA DE ANTROPOLOGÍA FÍSICA
(DEPARTAMENTO DE ZOOLOGÍA Y ANTROPOLOGÍA FÍSICA) DE LA UNIVERSIDAD
COMPLUTENSE DE MADRID,**

CERTIFICA:

Que la presente memoria titulada "**HISTORIA EVOLUTIVA DE LA POBLACIÓN ANDALUZA BASADA EN SU HERENCIA MATERNA Y SU RELACIÓN CON EL POBLAMIENTO HUMANO DEL ESPACIO MEDITERRÁNEO**", que presenta CANDELARIA L. HERNÁNDEZ DE LA FUENTE para optar al grado de Doctor (Mención Europea), ha sido realizada en el Departamento de Zoología y Antropología Física de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid bajo mi dirección. Así mismo, autorizo su presentación a la Junta de Facultad al considerar que reúne las condiciones científicas necesarias de un trabajo de Tesis Doctoral.

Para que así conste, firmo el presente certificado,

Madrid, octubre de 2015

Fdo.: Rosario CALDERÓN FERNÁNDEZ



La presente Tesis Doctoral ha sido financiada por el Ministerio de Economía y Competitividad a través de los Proyectos de Investigación **CGL2006-04749/BOS**, **CGL2010-15191/BOS**, **CGL2010-09060-E/BOS** y una Acción Integrada Hispano-Portuguesa (**PRI-AIBPT-2011-1004**) concedidos a R. CALDERÓN como Investigador Principal.

C. L. HERNÁNDEZ ha sido beneficiaria de una **Beca Predoctoral** de la Universidad Complutense de Madrid (2011-2015).

Esta Tesis Doctoral está basada en las siguientes publicaciones:

Hernández C.L., Reales G., Dugoujon J.-M., Novelletto A., Rodríguez J.N., Cuesta P., Calderón R. (2014) Human maternal heritage in Andalusia (Spain): its composition reveals high internal complexity and distinctive influences of mtDNA haplogroups U6 and L in the western and eastern side of region. *BMC Genet* 15(1): 11.

Capítulo III

Hernández C.L., Soares P., Dugoujon J.-M., Novelletto A., Rodríguez J.N., Rito T., Oliveira M., Melhaoui M., Baali A., Pereira L., Calderón R. (2015) Early Holocenic and historic mtDNA African signatures in the Iberian Peninsula: the Andalusian region as a paradigm. *PLoS One* (*aceptado para su publicación*, 17 sept. 2015).

Capítulo IV

Hernández C.L., Dugoujon J.-M., Novelletto A., Rodríguez J.N., Cuesta P., Calderón R. (2015) Mitochondrial DNA clade H in southern Iberia supports ancient human genetic exchange in the western extreme of the Mediterranean (*manuscrito en preparación*).

Capítulo V

Durante mis estudios de Doctorado en el Dpto. de Zoología y Antropología Física (Facultad de C.C. Biológicas, UCM), he participado directamente en las investigaciones que han tenido como objetivo avanzar en el conocimiento sobre la estructura genética de las poblaciones actuales y meridionales de la Península Ibérica (región de Andalucía). Esta actividad investigadora ha sido financiada a través de Proyectos de Investigación I+D+i del Plan Nacional (MINECO) en los cuales he formado parte oficial como miembro del equipo de investigación.

Ambrosio B., Dugoujon J.M., **Hernández C.**, de la Fuente D., González-Martín A., Fortes-Lima C.A., Novelletto A., Rodríguez J.N., Calderón R. (2010) The Andalusian population from Huelva reveals a high diversification of Y-DNA paternal lineages from haplogroup E: Identifying human male movements within the Mediterranean space. *Ann Hum Biol* 37(1): 86–107.

Ambrosio B., **Hernández C.**, Novelletto A., Dugoujon J.M., Rodríguez J.N., Cuesta P., Fortes-Lima C., Calderón R. (2010) Searching the Peopling of the Iberian Peninsula from the Perspective of Two Andalusian Subpopulations: A Study Based on Y-chromosome Haplogroups J and E. *Coll Antropol* 34(4): 1215–1228.

Ambrosio B., Novelletto A., **Hernández C.**, Dugoujon J.M., Fortes-Lima C., Rodríguez J.N., Calderón R. (2012) Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain). *Forensic Sci Int Genet* 6(2): e66–e71.

Fortes-Lima C., Dugoujon J.-M., **Hernández C.L.**, Reales G., Calderón R. (2014) Immunoglobulin genes in Andalusia (Spain). Genetic diversity in the Mediterranean space. *C R Biol* 337(11): 646–656.

Reales G., **Hernández C.L.**, Dugoujon J.-M., Novelletto A., Cuesta P., Fortes-Lima C., Rodríguez J.N., Calderón R. (2014) New insights into the distribution of APOE polymorphism in the Iberian Peninsula. The case of Andalusia (Spain). *Ann Hum Biol* 41(5): 443–452.

Calderón R., **Hernández C.L.**, Cuesta P., Dugoujon J.M. (2015) Surnames and Y-chromosomal markers reveal low relationships in Southern Spain. *PLoS One* 10(4): e0123098.

Agradecimientos

Muchas personas han contribuido directa o indirectamente a esta Tesis Doctoral. Quiero expresar mi gratitud a todas ellas en este breve 'prólogo'.

En primer lugar, me gustaría agradecer sinceramente a la Profa. Rosario Calderón por ser una guía constante desde que entré a colaborar con su equipo. Charo, gracias por tu dedicación, en muchas ocasiones más allá de lo profesional, por tu continuo esfuerzo y por empujarme siempre a dar lo mejor de mí.

Je voudrais remercier Prof. Jean Michel Dugoujon pour ses conseils et son aide à propos de la recherche. Prof. Andrea Novelletto has been also helpful in many scientific discussions. Dr. Luisa Pereira and her team hosted me several times in Porto and made me feel always welcome. I appreciate very much their help with data analysis of Chapter IV. In particular, I would like to acknowledge Pedro Soares and Teresa Rito.

Gracias al Dr. Juan N. Rodríguez por sus esfuerzos en la organización para la recogida de muestras de sangre. También querría agradecer a todos los donantes, que generosamente se prestaron a colaborar con nosotros, y sin los cuales nuestro trabajo no sería posible.

Gracias al Dr. Pedro Cuesta por estar siempre disponible para resolver dudas acerca del tratamiento de datos y análisis estadísticos.

Gracias a los profesores de la U.D. de Antropología Física por su cercanía a lo largo de todos estos años. Un agradecimiento especial se merecen los, en su día doctorandos, y ahora casi todos doctores, que he tenido el placer de conocer y con los que he compartido tantos buenos momentos. Gracias a Beatriz Ambrosio, Labib Drak, Jorge Román, Pedro Barrio y Verónica Alonso. Gracias a Susana Vergara y a M^a Ángeles Sánchez, por su ayuda para cuestiones administrativas, o de cualquier tipo, del Departamento.

Gracias a mi familia, a mis padres y hermanos, por ser un apoyo fundamental. Gracias a Aníbal por encargarse del diseño de la portada.

Cualquier agradecimiento que pueda expresar en esta página resulta escaso para Sergio, mi marido. Gracias por creer en mí en todo momento, por estar siempre cerca, y por regalar mucho de tu tiempo para mí.

Me gustaría dedicar esta Tesis a dos personas unidas a través de las generaciones por su genoma mitocondrial. En primer lugar, a la memoria de mi abuela Ángeles, de la que he heredado orgullosamente tantos rasgos, me gusta pensar que no solo son físicos. Por último, a alguien cuya llegada esperamos ansiosos y que ha inspirado mi trabajo y mi vida durante estos últimos meses: a Elena.

Índice

Resumen	7
Abstract	13
Capítulo I. Introducción general	19
1. Los fundamentos	21
1.1. Variabilidad genética humana: conceptos generales	22
1.2. Procesos que modelan la variación genética	23
1.3. Marco teórico para el estudio de la diversidad genética humana	28
2. Marcadores moleculares	30
2.1. Organización del genoma humano y variabilidad genómica	30
2.2. La herencia uniparental humana	33
3. El genoma mitocondrial como marcador poblacional	35
3.1. La mitocondria: biología celular, origen y características	35
3.2. El ADN mitocondrial: estructura y propiedades	36
3.3. El ADN mitocondrial y la historia evolutiva humana	42
4. La Península Ibérica dentro del espacio mediterráneo: análisis de su diversidad matrilineal	49
4.1. Primeros estudios y principales resultados	49
4.2. Análisis de las poblaciones meridionales de Iberia	50
4.3. Alcanzando la máxima resolución: el análisis de mitogenomas	51
5. La diversidad genética del sur de Iberia desde el estudio de otros marcadores	52
6. Los objetivos del presente trabajo	56

Capítulo II. Material y Métodos	57
1. La población de estudio	59
1.1. Localización geográfica.....	59
1.2. Características demográficas.....	60
1.3. Apuntes históricos.....	64
2. El proceso de muestreo poblacional	71
3. Obtención de material genético	74
3.1. Aislamiento y extracción del ADN genómico	74
3.2. ADN. Su cuantificación, control de calidad y estandarización de las muestras.....	76
3.3. Almacenamiento de las muestras.....	76
4. Criterios de filtrado y selección de muestras	76
5. Técnicas de análisis de marcadores moleculares mitocondriales	77
5.1. Secuenciación de la región control del ADNmt.....	77
5.2. PCR-RFLPs.....	79
5.3. Ensayos de PCR a tiempo real	81
5.4. Secuenciación completa del genoma mitocondrial	84
5.5. Genotipado múltiple por SNaPshot.....	86
6. El tratamiento de los datos a nivel de secuencias de ADN, estadístico y filogenético	89
6.1. Análisis de secuencias de ADN	89
6.2. La estimación de edades de coalescencia	95
6.3. Análisis fundador (Founder Analysis, FA)	96
6.4. La estimación de tamaños efectivos (N_e) poblacionales.....	98
6.5. Análisis de estructura poblacional	98
6.6. Análisis multivariantes.....	99
6.7. La evaluación de la correlación entre geografía y variantes genéticas	100

7. Uso de bases de datos y selección de poblaciones para análisis comparativos.....	103
--	------------

Capítulo III. Herencia materna en Andalucía: estructura y diversidad

(Human maternal heritage in Andalusia (Spain): its composition reveals high internal complexity and distinctive influences of mtDNA haplogroups U6 and L in the western and eastern side of region) **105**

1. Introducción.....	108
2. Material y Métodos.....	110
2.1. Poblaciones y muestras.....	110
2.2. Análisis moleculares de ADNmt.....	110
2.3. Análisis estadísticos y filogenéticos.....	111
3. Resultados.....	111
3.1. Diversidad de las secuencias de ADNmt en Andalucía.....	111
3.2. El flujo de los linajes mitocondriales en Andalucía.....	113
3.3. Reconstruyendo árboles de secuencias mitocondriales.....	116
3.4. Estructura genética poblacional.....	119
4. Discusión	120

Capítulo IV. Linajes africanos en Iberia: señales de un flujo génico antiguo

(Early Holocenic and historic mtDNA African signatures in the Iberian Peninsula: the Andalusian region as a paradigm) **127**

1. Introducción.....	130
2. Material y Métodos.....	132
2.1. Poblaciones y selección de muestras.....	132
2.2. Secuenciación de mitogenomas y análisis estadísticos.....	133
2.3. Análisis fundador (FA) y Bayesian Skyline Plots (BSPs).....	134
3. Resultados.....	136
3.1. Linajes mitocondriales africanos en Iberia y áreas próximas. Patrones filogeográficos.....	136
3.2. Secuenciación completa del ADNmt. Filogenias y estimas de coalescencia.....	138

3.3. <i>Linajes mitocondriales africanos en el Mediterráneo: marco temporal y dispersiones</i>	144
4. Discusión	149
Capítulo V. La historia del clado H en el Mediterráneo occidental	
<i>(Mitochondrial DNA clade H in southern Iberia supports ancient human genetic exchange in the western extreme of the Mediterranean)</i>	155
1. Introducción	158
2. Material y Métodos	160
2.1. <i>Poblaciones y selección de muestras</i>	160
2.2. <i>Caracterización molecular del clado H</i>	160
2.3. <i>Análisis estadísticos y filogenéticos</i>	161
3. Resultados	162
3.1. <i>Disección del haplogrupo mitocondrial H en Andalucía</i>	162
3.2. <i>Estructura genética-poblacional en el Mediterráneo</i>	169
3.3. <i>Estimación de edades de coalescencia</i>	173
4. Discusión	173
Capítulo VI. Discusión general	181
1. La diversidad mitocondrial en Iberia y en otros territorios y poblaciones del espacio mediterráneo	183
1.1. <i>Los linajes maternos en Andalucía</i>	183
1.2. <i>Andalucía como puente entre Europa y África</i>	185
1.3. <i>Barreras y/o corredores geográficos en el Mediterráneo</i>	190
2. Consideraciones metodológicas	194
3. Los patrones genéticos en Andalucía: otros enfoques moleculares .	196
3.1. <i>Linajes paternos en el sur de Iberia</i>	196
3.2. <i>Algunos resultados sobre marcadores autosómicos</i>	200
4. Perspectivas de futuro	201
Capítulo VII. Conclusions	205

Referencias **211**

Anexos **239**

Resumen

El estudio de la diversidad genética actual representa una ventana hacia el pasado de las poblaciones humanas. Nuestro genoma y su variación proporcionan claves para comprender los complejos acontecimientos evolutivos y demográficos que ha experimentado nuestra especie en su historia, tales como constricciones y expansiones poblacionales, episodios de migración y flujo génico, colonización de nuevos territorios y adaptación a su ambiente. Uno de los marcadores genéticos que más han contribuido al conocimiento de la historia del hombre moderno es el ADN mitocondrial (ADNmt). Las propiedades inherentes a esta molécula han permitido datar eventos evolutivos, establecer tiempos de migración y desvelar fenómenos de mezcla y dinámica poblacional.

El Mediterráneo es una de las áreas del mundo más relevantes en términos de mezcla poblacional. Dentro de esta cuenca geográfica, la Península Ibérica ha sido un foco principal de atracción para diversas culturas durante su prehistoria e historia hasta tiempos recientes. La franja cantábrica peninsular representa la región más frecuentemente analizada en términos de su diversidad mitocondrial, mientras que destaca la ausencia de análisis monográficos dedicados a evaluar la composición genética materna de la población meridional de Iberia, Andalucía. Su localización geográfica convierte a las poblaciones autóctonas del sur ibérico en un objetivo prioritario para ampliar el conocimiento acerca de las relaciones entre Europa y África y de los procesos de poblamiento del espacio mediterráneo.

El objetivo central de la presente Tesis Doctoral ha sido evaluar la composición y estructura genética matrilineal de poblaciones del sur de la Península Ibérica. Más concretamente, nuestro trabajo pretende analizar las características genéticas de Andalucía, y comparar sus patrones mitocondriales con los detectados en otras áreas y poblaciones de Iberia. Conocer el papel del sur de Iberia como receptor de genes maternos de diversos orígenes continentales, el significado de las rutas marítimas en las migraciones humanas y la permeabilidad de las barreras geográficas al flujo génico en el Mediterráneo, representan otros de los objetivos importantes de esta Tesis.

Para todo ello, hemos realizado un estudio global de la herencia materna de dos poblaciones autóctonas andaluzas, pertenecientes a las provincias de Huelva (en el oeste andaluz) y de Granada (en el este), mediante secuenciación de la región control mitocondrial y el análisis de ciertos SNPs de la región codificante por PCR-RFLPs. Posteriormente, fueron seleccionados ciertos haplogrupos de interés poblacional, los cuales se analizaron a un nivel de resolución filogenética más profundo, empleando diversas técnicas moleculares. Las trazas maternas africanas en Iberia fueron estudiadas mediante la secuenciación del genoma mitocondrial de 32 andaluces y 30 bereberes de Marruecos pertenecientes a los haplogrupos U6, M1 y L. Por último, el linaje mitocondrial más frecuente en Europa, el clado H, fue filogenéticamente diseccionado gracias a un genotipado múltiple por tecnología SNaPshot. Para nuestros análisis se genotiparon un total de 750 andaluces autóctonos, sanos y no emparentados biológicamente.

Nuestros resultados han demostrado que Andalucía es un territorio genéticamente heterogéneo tanto para la composición mitocondrial global como para linajes maternos particulares. Los dos extremos geográficos de Andalucía se

encuentran caracterizados por un amplio espectro de haplogrupos que revelan patrones diferentes y, en cierto modo, complementarios. Mientras que la población occidental de Huelva muestra una elevada frecuencia y gran diversificación de los linajes de origen africano (U6 y L), Granada, en cambio, se caracteriza por la diversidad interna del clado europeo por excelencia (H). Globalmente, ambas poblaciones andaluzas registran altos valores de diversidad como consecuencia de numerosos y continuados episodios de flujo génico desde la prehistoria hasta tiempos históricos, un patrón que permite distinguir a nuestras poblaciones de las que habitan la franja cantábrica.

Además, la información mitocondrial nos ha revelado que los acontecimientos migratorios que unen Iberia con el noroeste de África han sido sostenidos en el tiempo, de notable profundidad temporal y de carácter bidireccional. De ahí que, la influencia africana en la Península (señalizada por los linajes U6, M1 y L) no puede ser explicada exclusivamente por el peso de las migraciones históricas, las cuales encuentran su referente en la prolongada dominación musulmana. También nuestros datos apuntan a que la mayoría de las señales africanas en Iberia se deben a contactos transcontinentales centrados en el Holoceno temprano. Los resultados sobre linajes africanos nos señalan también una entremezclada historia evolutiva de las dos orillas del Estrecho de Gibraltar. Esta conclusión puede ser trasladada al análisis del clado mitocondrial H en alto nivel de resolución. Iberia y el Magreb comparten, además de elevados valores del linaje H, paralelismos en la estructura interna de este clado, siendo el H1 y el H3 los sub-clados principales en ambas regiones. Los análisis filogeográficos y filogenéticos confirman la hipótesis según la cual estas dos radiaciones habrían llegado al Magreb por un movimiento masivo poblacional desde Iberia tras el último glacial máximo.

Los linajes maternos muestran la existencia de una estructuración significativa de las poblaciones mediterráneas, preferentemente en un eje longitudinal (oeste-este). A este escenario se une la remarcable discontinuidad genética que existe en el norte de África. La existencia de áreas desérticas, las cuales han hecho muy difícil los contactos poblacionales, habría sido una importante razón causal que ha conducido a la modelación de la estructura genética actual de las poblaciones norteafricanas.

Todos los resultados obtenidos en la presente Tesis Doctoral, nos llevan a proponer que tanto el Estrecho de Gibraltar como sus áreas marítimas circundantes deben considerarse como un puente transitado entre ambos continentes, al menos, desde el punto de vista de la herencia materna. El mar Mediterráneo ha actuado como una *superficie de transporte*, es decir, como una barrera permeable a las migraciones humanas durante tiempos prehistóricos e históricos. La composición genética de la población mediterránea se estructura longitudinalmente, lo que evidencia eventos pasados en su propia historia evolutiva, incluyendo episodios como la colonización Paleolítica, las expansiones poblacionales post-glaciales y la transición Neolítica.

En definitiva, esta Tesis Doctoral supone una contribución al conocimiento sobre los procesos que han modelado la historia genética humana reciente en el

espacio mediterráneo y, más concretamente, sobre las dinámicas poblacionales que han experimentado los habitantes del sur de la Península Ibérica.

Palabras clave: *ADN mitocondrial, Península Ibérica, Diversidad Genética Humana, Flujo Génico, Estructura Poblacional, Filogenia, Filogeografía, Haplogrupos U6 y L, Linaje H.*

Abstract

Present-day genetic diversity is a window into the evolutionary past of human populations. Our genome and its variation provides clues to understand the complex evolutionary and demographic events that our species has experienced in its history such as population constrictions and expansions, episodes of migration and gene flow, colonization of new territories and adaptation to the environment. The mitochondrial DNA (mtDNA) has significantly contributed to the knowledge of modern human evolutionary history. The underlying characteristics of this molecule allow dating evolutionary events, setting migration times, admixture processes and unveiling population dynamic phenomena.

The region of the Mediterranean Basin is among the most demanding admixture zones in the globe, due to the possible occurrence of many migrations in both directions, along thousands of years. Within this geographic Basin, the Iberian Peninsula has been a focus of attraction for several cultures and civilizations from its prehistory and history until recent times. While its Cantabrian fringe represents the most regularly analyzed area in terms of its mitochondrial diversity, the absence of monographic surveys on maternal genetic composition of southern Iberians (i.e. Andalusians) is striking. Because of its geographic position, Andalusians are a target population to reach a better knowledge about the relationships between Europe and Africa, and peopling processes of the Mediterranean space.

The main aim of this PhD Thesis has been to evaluate the matrilineal genetic composition and structure of defined populations from southern Iberia. More specifically, our work has tried to analyze the genetic continuity within Andalusia and to compare the observed mitochondrial profiles with those found in other areas of Iberia, Europe and Mediterranean space. The role of southern Iberia as a recipient of maternal genes from different continental origins, the involvement of maritime routes in human migrations and the permeability of geographic barriers to gene flow across the Mediterranean are other of the relevant aims addressed in the present Thesis.

Firstly, we carried out a global study of the maternal heritage in two Andalusian populations, from Huelva and Granada provinces, respectively located in the west and east of region. The survey was mainly based of mtDNA control region sequencing together with the screening of some coding region SNPs by PCR-RFLPs. Afterwards, candidate haplogroups were selected and analyzed, in a deeper phylogenetic resolution level, by using several experimental approaches. African genetic signatures into the Andalusian gene pool were studied by complete sequencing of the mitochondrial genome of 32 Andalusians and 30 Berbers –from Morocco– harboring haplogroups U6, M1 and L. In addition, we dissected the most commonly detected maternal lineage in Europe, the clade H, by multiple genotyping (SNaPshot technology). In our study, 750 healthy, unrelated and autochthonous Andalusian individuals were genotyped.

Western and eastern Andalusians show a significant genetic heterogeneity for both the global mitochondrial composition as well as for specific maternal

lineages. Andalusians are characterized by a broad spectrum of mitochondrial haplogroups that reveal opposite and, in a certain extent, complementary profiles. While western Andalusians from Huelva harbored a high frequency and diversification of African lineages (U6, L), the neighbor relatives from Granada presented a great internal diversity of the typical European clade H. Interestingly, both southern Iberian populations registered high diversity values as the consequence of numerous and continuous gene flow episodes from prehistory onwards. The detected mtDNA profiles allow distinguishing these populations from those settled at the far end of the Peninsula (Cantabrian cornice).

Mitochondrial findings revealed that population contacts between Iberia and northwestern Africa have been continued and bidirectional through time, and they present indeed a relevant temporal depth. Therefore, the African influence in the Peninsula, through lineages U6, M1 and L, cannot only be explained by the role of the historic episodes, being the most paradigmatic example the long lasting Muslim invasion of Iberia. Our data fairly have shown that the bulk of the mtDNA African lineages in Iberia is due to early Holocene transcontinental contacts, revealing an interleaved evolutionary history of both shores of the Strait of Gibraltar. This interesting conclusion could be transferred when studying mitochondrial clade H in a high level of genealogical resolution. Iberia and the Maghreb share high frequencies of clade H. Besides, parallelisms in its internal structure, mainly when exploring the major sub-clades H1 and H3, have been observed as well. In this line, phylogeographic and phylogenetic analyses confirmed that H1 and H3 would have reached the Maghreb by means of a massive population movement from Iberia after the last glacial period.

Maternal lineages from Mediterranean populations show a significant structure in a longitudinal axis (west-east). This scenario is complemented by the pronounced genetic discontinuity along the North of Africa. The presence of desert areas, which have prevented population contacts, would have been a relevant causal reason to shape the present genetic structure of North African populations.

The whole results provided in this PhD Thesis lead us to propose that both the Strait of Gibraltar and its surrounding maritime region must be considered as a bridge between continents, at least for the maternal heritage. The Mediterranean Sea has acted as a *transport surface*, that is, as a permeable barrier to human migrations from prehistoric and historic times. The genetic composition of the Mediterranean population is structured in a longitudinal axis, which reveals past events in the evolutionary history of human populations, including episodes as the Paleolithic colonization, post-glacial expansions and the effects of Neolithic transition.

In conclusion, this research work is a contribution to the knowledge on the processes that have shaped the recent human genetic history in the Mediterranean and, more specifically, the population dynamics that have experienced the inhabitants of southern Iberia.

Keywords: *mitochondrial DNA, Iberian Peninsula, Human Genetic Diversity, Gene Flow, Population Structure, Phylogeny, Phylogeography, Haplogroups U6 and L, Lineage H.*

Capítulo I

Introducción general

1. Los fundamentos

Los patrones de diversidad genética que presentan las poblaciones humanas contemporáneas son una ventana que permite inferir el pasado de las mismas. Los genes que portamos y sus variaciones geográficas nos proporcionan claves para entender los complejos acontecimientos demográficos y evolutivos que ha experimentado nuestra especie a lo largo de su historia. Constricciones poblacionales, fenómenos migratorios y de flujo génico, colonización de nuevos territorios y adaptación a ambientes específicos son ejemplos de episodios que pueden ser revelados a través del estudio de nuestro material genético. Señales de esos eventos poblacionales se encuentran en el genoma humano y su análisis, combinado con otras fuentes de información como la Arqueología, la Paleoantropología y la Lingüística, nos pueden permitir esbozar una completa visión de nuestro pasado (ver **Figura I-1**).

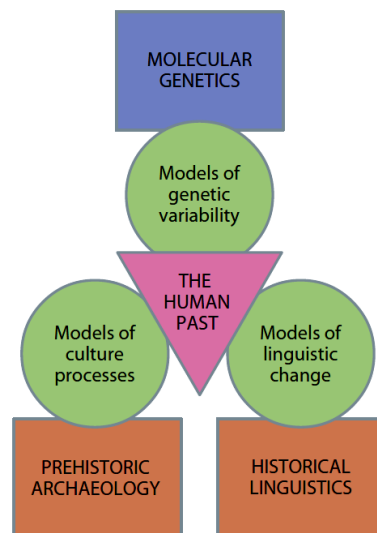


Figura I-1. Fuentes de información sobre el pasado humano.
Fuente: Jobling et al., 2013.

En consecuencia, centrarnos en la variación genética de las poblaciones humanas contemporáneas no solo nos aporta conocimientos en el campo de la Genética de Poblaciones y de la Antropología Física, sino que también tiene un gran alcance para otras disciplinas como la Genética Médica, la Historia y la Demografía. También, en el contexto de las conexiones existentes entre Biología y Medicina, se pretende responder a cómo la diversidad genética que nos encontramos en la actualidad queda reflejada en la presencia de alelos causantes de susceptibilidad a enfermedades. El detallado conocimiento de la estructura poblacional y la variación genética humana es esencial para la correcta interpretación de la relación entre material genético y enfermedad humana (Tishkoff & Verrelli, 2003).

1.1. Variabilidad genética humana: conceptos generales

Numerosas evidencias empíricas han revelado que la mayor parte de la diversidad genética en el hombre se encuentra dentro de las poblaciones (es decir, entre los individuos que componen una población) y no entre poblaciones diferenciadas. Como valor medio, un 6% de la variación genética corresponde a diferencias entre poblaciones, aunque la proporción varía dependiendo del marcador genético empleado (Jobling et al., 2013). Estas constataciones han sido la base para que el concepto de 'raza' sea apartado completamente del panorama científico desde hace años (Tishkoff & Kidd, 2004; Madrigal & Barbujani, 2006). De hecho, las clasificaciones raciales tradicionales han estado basadas en gran medida en características como la pigmentación, que no resultan buenos indicadores, ya que son el resultado probable de una adaptación a condiciones ambientales y, por tanto, son caracteres sujetos a una evolución convergente.

En Antropogenética, el término 'población' es básico. Desde un punto de vista de la Demografía, una población se define como un grupo de individuos que vive en un área específica, en un momento específico y durante generaciones (Murdock & Ellis, 1991). La Biología incorpora a esta definición la reproducción o transmisión efectiva de los genes (la población es un grupo de individuos con reproducción potencial que ocupa un área geográfica). En una población ideal, sus miembros tienen la misma probabilidad de emparejamiento con cualquier otro miembro de sexo opuesto, lo que se denomina '*panmixia*'. Sin embargo, en la práctica nos enfrentamos a que esta situación ideal en las poblaciones humanas se ve modificada (*assortative mating*) por factores como el ambiente, la estructura geográfica o prácticas socio-culturales (lingüística, religión, etc.) (Mielke & Fix, 2006).

Se ha observado que la variación genética humana es principalmente '*clinal*', esto es, esa diversidad se distribuye según gradientes de frecuencias (Handley et al., 2007). A escala global, no se detectan discontinuidades genéticas entre continentes, sino que la diversidad humana presenta más bien escenarios de variación continua (Madrigal & Barbujani, 2006). Asimismo, la diferenciación genética entre pares de poblaciones correlaciona con la distancia geográfica entre las mismas. Diversos estudios han mostrado que la geografía explica al menos el 75% de la varianza entre poblaciones humanas. El modelo que se establece para explicar esta situación se conoce como modelo de aislamiento por distancia (*Isolation by Distance*, IBD) y se encuentra en clara relación con el concepto de 'gradiente' entendido como el cambio gradual lineal en un carácter (frecuencia alélica) con el incremento de la distancia física (Handley et al., 2007).

La distribución de las variantes genéticas se interpreta habitualmente desde un punto de vista *filogeográfico*. La 'filogeografía' es el campo de estudio relativo a los principios y procesos que gobiernan la distribución geográfica de linajes genealógicos, especialmente, dentro de especies y entre especies relacionadas. El concepto imbrica componentes históricos y filogenéticos de la distribución espacial de los linajes, por lo que el tiempo y el espacio son los dos ejes claves de la filogeografía (Avice, 2000). Este campo de estudio busca la interpretación de los procesos históricos en demografía poblacional (colonización y poblamiento,

cambios en tamaño poblacional, etc.) que han dejado huellas evolutivas en la distribución geográfica actual de ciertos caracteres génicos. En este contexto, se define '*poblamiento*' como un proceso dinámico y continuo, que implica movimientos de personas a larga distancia y a gran escala, y que ha sido particularmente frecuente en tiempos recientes (Arredi et al., 2006).

Por tanto, la variación genética humana no es simplemente el resultado de la acumulación de cambios al azar o no necesariamente estocásticos, sino que se encuentra gobernada por variables biológicas y, también, por factores que solo pueden ser explicados por nuestra naturaleza social (*co-evolución entre genes y cultura*).

1.2. Procesos que modelan la variación genética

Existen fuerzas microevolutivas que intervienen modificando las frecuencias génicas, y por tanto, genotípicas, de las poblaciones (Falconer & Mackay, 2006). Estos procesos son de dos tipos: (1) sistemáticos, que tienden a cambiar la frecuencia génica de forma predecible tanto en cantidad como en sentido –mutación, migración y selección natural, y (2) dispersivos, que surgen en poblaciones pequeñas debido a los efectos del muestreo gamético. Este último mecanismo se conoce con el nombre de *deriva genética* (cambio aleatorio de las frecuencias génicas) y tiene como consecuencia la diferenciación entre sub-poblaciones, la uniformidad dentro de las sub-poblaciones (por disminución de la varianza genética) y el incremento de los niveles de homocigosis.

Además de estos fenómenos, la variación humana puede verse modelada a su vez por prácticas socio-culturales. En este marco pueden ser citados los patrones matrimoniales o de emparejamientos preferenciales, donde el parentesco biológico está presente.

Mutación

La variación genética se genera por una serie de factores que introducen nuevas variantes en la población (Dermitzakis, 2010). Los cambios en la secuencia de ADN pueden englobar desde un único nucleótido hasta varios miles. En la sección 2.1 se analiza la naturaleza de las variantes genéticas más conocidas y estudiadas en las poblaciones humanas.

Migración

Este fenómeno se refiere a los intercambios de información genética entre sub-poblaciones parcial o totalmente aisladas. De este modo, la transmisión de material genético no solo se sucede verticalmente (de padres a hijos) sino también horizontalmente entre poblaciones (Dermitzakis, 2010). Se trata de una de las variables demográficas claves que determinan cambios en la composición y el tamaño poblacional. Es interesante destacar que, al contrario que otros

factores demográficos como la fertilidad o la mortalidad, el flujo génico no puede ser caracterizado, en la mayor parte de las situaciones, mediante eventos únicos.

La magnitud, duración y patrón del proceso migratorio viene determinado por una serie de factores (Mielke & Fix, 2006):

- a. Distancia geográfica y el papel que representan las barreras físicas.
- b. Tamaño y estructura de los grupos migrantes.
- c. Factores socio-culturales. La consanguinidad es una variable clave que afecta a la movilidad humana. De manera más general, los patrones de residencia post-marital varían significativamente entre sexos, por lo que nos encontramos ante diferencias específicas en los comportamientos migratorios. Alrededor del 70% de las sociedades humanas son '*patrilocales*', lo que implica que la mujer se traslada al lugar de residencia del varón. Este proceso se revela de forma clara en el análisis de los marcadores uniparentales (Jobling et al., 2013).

La migración tiene consecuencias genéticas que oscilan entre la dispersión de nuevas variantes mutantes o la mezcla de poblaciones con una composición genética más o menos afín, o muy distintiva. La tasa a la que estas mutaciones alcanzan altas frecuencias o incluso se fijan en la población receptora –por la acción combinada de la migración y la deriva o selección– determina el grado de diferenciación genética interpoblacional.

Selección

Bajo el modelo de evolución neutral, la variación genética humana se acumula mediante procesos neutrales. La pérdida de alelos o su fijación vendrá determinada por la deriva genética y la tasa de evolución (o cambio) es función exclusivamente de la tasa de mutación, μ asociada a ese nuevo alelo mutante. Sin embargo, en la práctica, ciertas diferencias genéticas y fenotípicas que separan a las poblaciones naturales (entre las que forma parte la humana) pueden ser explicadas como adaptaciones favorecidas por la selección natural (Tishkoff & Verrelli, 2003).

La selección natural, por tanto, es la fuerza que incrementa o disminuye la probabilidad de herencia de determinadas variantes genéticas debido a diferencias en eficacia biológica (*fitness*). Las 3 formas básicas de selección se ilustran en la **Figura I-2**.

La selección purificante (*purifying selection*) provoca que alelos deletéreos o con *fitness* reducido se pierdan en la población. Por el contrario, la selección positiva (*positive selection*) se basa en que alguna de las mutaciones que ocurren en secuencias funcionales puedan conferir ventajas adaptativas a los individuos que las portan. El efecto de la fijación de esta mutación ventajosa puede arrastrar otras variantes, en su mayoría neutrales, que se encuentran asociadas y que también se fijan, lo que tiene como consecuencia un proceso conocido como

'barrido selectivo' (*selective sweep*). Por otro lado, mediante la selección balanceadora (*balancing selection*) una región mantiene altos niveles de variación y un gran número de variantes a frecuencias intermedias (Dermitzakis, 2010). La presencia de ventajas antagónicas entre alelos lleva a un punto de equilibrio óptimo de las frecuencias en la población.

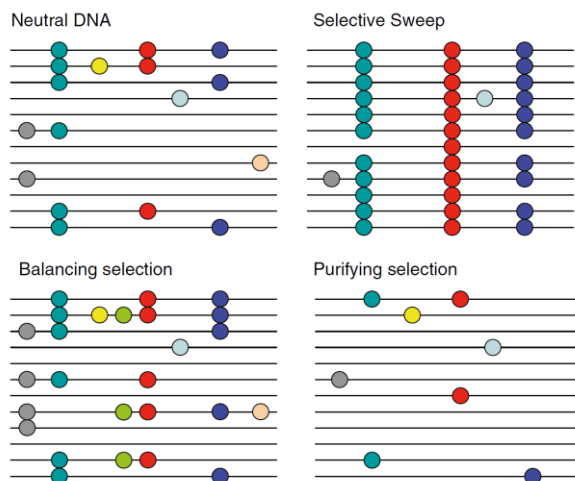


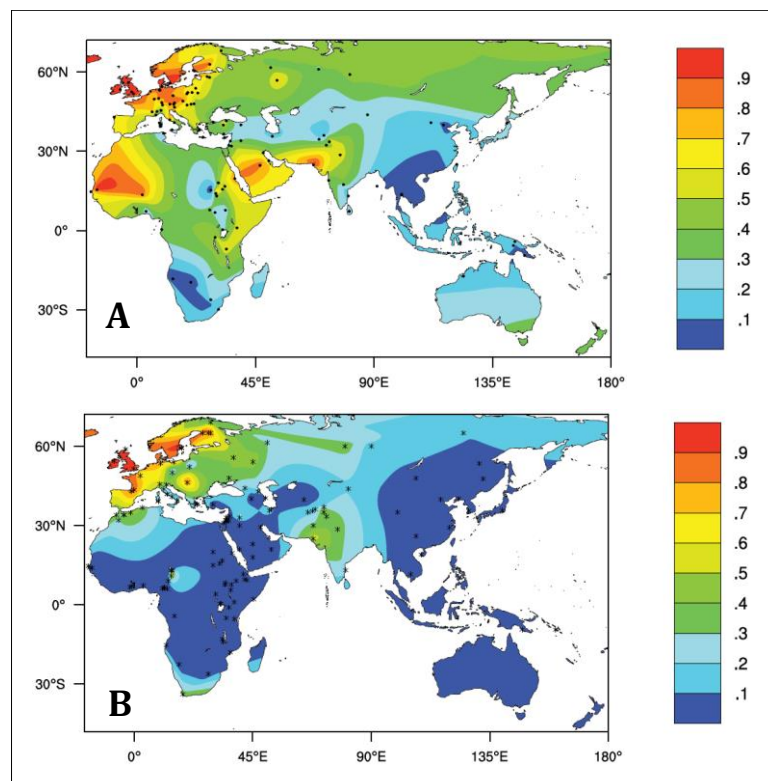
Figura I-2. Patrones de variación de la secuencia nucleotídica en función de experimentar evolución neutral o diferentes manifestaciones de la selección natural. Fuente: Dermitzakis, 2010.

La '*ventaja del heterocigoto*' es un ejemplo de esta última manifestación selectiva. Este fenómeno explica las altas frecuencias de un alelo *a priori* deletéreo pero que, mantenido en heterocigosis, puede suponer una adaptación a un determinado estrés ambiental. Uno de esos factores ambientales es la infección de parásitos del género *Plasmodium*, lo que representa la evidencia más paradigmática de una *adaptación molecular de las poblaciones humanas* frente a un agente selectivo (la infección por malaria). Así, variantes como el alelo Hb^S del gen de la β -globina (que causa la anemia falciforme); otras mutaciones en el mismo gen (que generan desórdenes como las β -talasemias); alelos deficientes de la enzima glucosa-6-fosfato deshidrogenasa (G6PD) o el alelo FY*B^{ES} del *locus* Duffy, se encuentran a altas frecuencias en regiones endémicas para la malaria (Tishkoff & Verrelli, 2003; Jobling et al., 2013). La superposición entre la distribución geográfica preferencial de la enfermedad y de las variantes deficitarias de estos *loci* ha llevado a proponer que estas últimas confieran resistencia a la infección o una atenuación de los síntomas derivados de la misma.

Otras presiones selectivas ambientales han provocado no solo una adaptación a nivel molecular, sino también una *adaptación fenotípica* en el hombre. Estos agentes ambientales están en relación con aspectos clave de nuestro pasado evolutivo, entre los que destacan la colonización de nuevas áreas continentales, el desarrollo y la expansión de la agricultura y los cambios en la dieta derivados de este último proceso (Jobling et al., 2013). Una de las formas más evidentes de

variación fenotípica entre las poblaciones humanas es el color de la piel. La pigmentación humana muestra una correlación con la latitud y el nivel de radiación UV (Jeong & Di Rienzo, 2014). Las pieles oscuras protegerían de las quemaduras y la degradación de folato. Sin embargo, frente a una menor radiación solar, un exceso de protección podría tener como consecuencia una síntesis insuficiente de vitamina D. El gen *M1CR* –que codifica para el receptor de la melanocortina-1– ha sido ampliamente estudiado, ya que algunas de sus variantes afectan a la pigmentación. Cabe destacar las diferencias en la distribución de estas mutaciones que se observan entre África y Europa. La presión selectiva ha sido tan potente que se han detectado algunas variantes asociadas con la pigmentación que son específicas de europeos o de población asiática. Estos resultados apoyarían la evolución convergente de la pigmentación clara en esas áreas geográficas (Arredi et al., 2006).

Figura I-3. A. Frecuencias de la LP. **B.** Frecuencias fenotípicas estimadas basadas en la frecuencia del alelo -13.910^*T . Fuente: Itan et al., 2010.



La revolución que supuso la transición Neolítica y, por tanto, la asimilación de prácticas agrícolas y ganaderas, también, ha modificado nuestra variación genética y ha contribuido a la acumulación de características adaptativas locales. La ‘persistencia de la lactasa’ (*Lactase Persistence*, LP) es una condición que se refiere a la expresión continuada de esta enzima hidrolasa, codificada por el gen *LCT*, la cual degrada la lactosa. Este carácter se encuentra en altas frecuencias en poblaciones que han practicado tradicionalmente el pastoreo. Sobre la base de

esta correspondencia, se ha propuesto que la *LP* confiera una ventaja selectiva basada en la dieta. Además, el análisis molecular revela que esta adaptación, del mismo modo que la pigmentación humana, muestra evolución convergente. Varios polimorfismos están involucrados, siendo el alelo -13.910*T el más frecuente en europeos (ver **Figura I-3**), mientras que el -13.907*G, -13.915*G y -14.010*C se detectan en población africana o del Oriente Medio (Itan et al., 2010). Todas estas variantes se encuentran topográficamente muy próximas entre sí dentro del gen y muestran, además, señales de selección positiva reciente (Jeong & Di Rienzo, 2014).

Deriva genética

La deriva es definida como la fluctuación al azar de las frecuencias alélicas de una generación a otra debida al muestreo aleatorio de los gametos (Tishkoff & Verrelli, 2003). Los efectos de este proceso están en función del tamaño efectivo (N_e) o del número de individuos con reproducción efectiva en la población. El N_e representa el tamaño poblacional ideal que posee los mismos parámetros genéticos que la población observada y suele ser menor al tamaño real (N_{BP} , *Breeding Population Size*) observado (Dermitzakis, 2010).

Por tanto, los efectos de la deriva son particularmente fuertes en poblaciones que hayan experimentado ‘cuellos de botella’ (*bottleneck effect*, reducción drástica en el tamaño poblacional) o ‘efecto fundador’ (*founder effect*, pequeño grupo que se separa de otro más grande y coloniza una nueva región) (Tishkoff & Verrelli, 2003). En estos escenarios demográficos de constricción del tamaño poblacional, una mutación que se produzca tiene una alta probabilidad de fijación (ver **Figura I-4**).

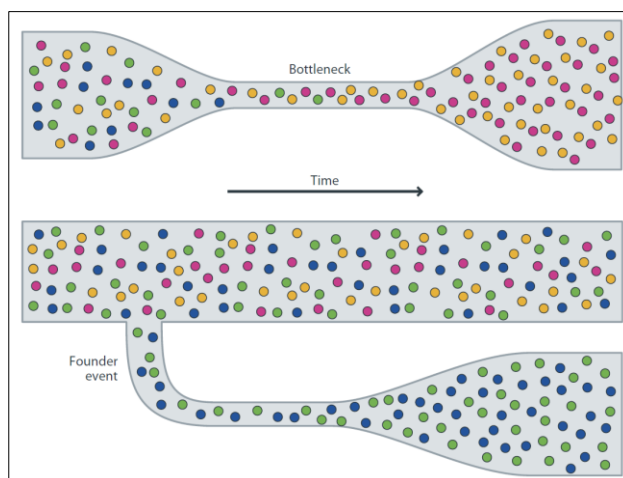


Figura I-4. Cambios en la frecuencia de los alelos (representados por colores) como resultado de fenómenos de cuello de botella y efecto fundador. Fuente: Jobling et al., 2013.

Otros factores

Además de los fenómenos discutidos anteriormente, la estructura de la variación genética humana depende de otros procesos en los que interviene el comportamiento socio-cultural de las poblaciones. El equilibrio Hardy-Weinberg incorpora la premisa de un tamaño poblacional infinito y emparejamiento o cruzamientos al azar. Este escenario no es lineal en el caso del hombre, ya que la elección de una pareja está fuertemente influenciada por factores como la geografía, tamaño poblacional, etnicidad, religión, educación o estatus económico (Bittles, 2010). Por tanto, nos encontramos frente a emparejamientos selectivos (*assortative mating*) o tendencia de los individuos con parecido fenotípico o relaciones de parentesco biológico a emparejarse entre sí (Tishkoff & Kidd, 2004). Esta situación refuerza el aislamiento genético de las poblaciones, incrementado a su vez por mecanismos como la deriva. El bajo tamaño poblacional y, por tanto, la limitación en el número de parejas potenciales (*angustia loci*) motiva el emparejamiento con familiares, situación que tiene como consecuencia que los niveles de homocigosis en la población aumenten (Dermitzakis, 2010). Esto significa que la pareja tiene una probabilidad mayor de heredar alelos idénticos por descendencia (*Identical By Descent*, IBD) en una proporción de *loci*. Las uniones consanguíneas pueden tener diversas consecuencias, tanto relativas a la fertilidad de la pareja como a morbilidad e incidencia de enfermedades en la descendencia (Bittles, 2010).

1.3. Marco teórico para el estudio de la diversidad genética humana

Para explicar las relaciones genealógicas entre especies o genes (en nuestro caso, entre poblaciones e individuos pertenecientes a una especie) se emplean 'filogenias' o 'árboles filogenéticos'. Las ramas (*branches*) de estos árboles representan la persistencia de un linaje genético a través del tiempo, y los nodos (*nodes*), el nacimiento de un nuevo linaje. Las terminaciones o nodos externos (*tips*) indican las especies o individuos del presente, mientras que los nodos internos habitualmente simbolizan ancestros extintos. El patrón de ramificación de un árbol se denomina 'topología' (Yang, 2006). La **Figura I-5** nos muestra la terminología habitualmente empleada en las representaciones filogenéticas. Las longitudes de rama indican la cantidad de divergencia de secuencia, y se transforman en tiempo mediante la aplicación de tasas de cambio evolutivo. Si la tasa de mutación es constante a lo largo del tiempo, esta asunción se conoce como '*reloj molecular*' (Yang, 2006).

Estas filogenias se infieren a partir de una información molecular, habitualmente secuencias de ADN. Existen diversas metodologías de reconstrucción filogenética y principalmente se dividen en: (1) métodos basados en distancias (*distance-based methods*), los cuales se centran en el cálculo de distancias entre cada par de secuencias. Un ejemplo es el *neighbor-joining*; (2) métodos basados en caracteres (*character-based methods*), como por ejemplo la máxima parsimonia, máxima verosimilitud o inferencia Bayesiana. Estas últimas aproximaciones comparan simultáneamente todas las secuencias en el

alineamiento, considerando un carácter (un sitio en el alineamiento) para calcular una puntuación (*score*) para cada árbol. La comparación de los *score* de los árboles ofrece la mejor estimación de la filogenia (Yang & Rannala, 2012).

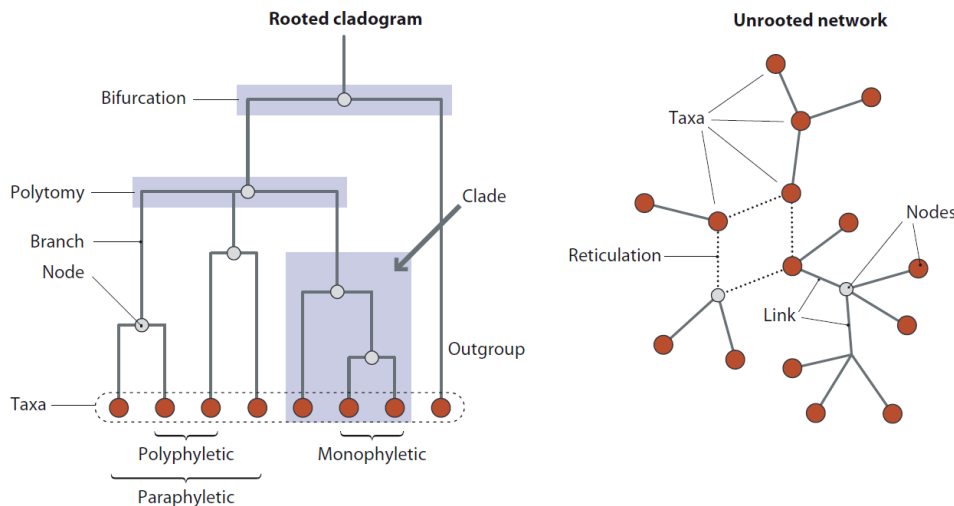


Figura I-5. Nomenclatura para árboles y *networks*. Fuente: Jobling et al., 2013.

La reconstrucción filogenética se basa en la '*teoría coalescente*', que permite la modelización de la genealogía de un conjunto de individuos hacia atrás en el tiempo así como la inferencia de las propiedades y parámetros poblacionales genéticos de la población global (Nordborg, 2003). El marco coalescente predice que esos $2N$ cromosomas comparten un antecesor común en algún momento del pasado (*Most Recent Common Ancestor*, MRCA). Por tanto, la muestra de cromosomas se puede analizar como un árbol filogenético de secuencias de la misma especie, y las longitudes de ramas del árbol pueden estimarse mediante el modelo esperado según el cual la población ha evolucionado. Los modelos coalescentes tienen tres aplicaciones claves: (1) interpretar los procesos potenciales que han contribuido al patrón de variación observado en una población humana, (2) estimar ciertos parámetros de la población, como por ejemplo, tasas de recombinación, y (3) investigar el papel de la selección natural y su impacto en nuestro genoma. En este último caso es interesante conocer que el patrón de longitud de ramas se correlaciona con el tipo de selección que ha actuado potencialmente en la secuencia. Por ejemplo, un escenario de selección purificante tendría como consecuencia un árbol coalescente con ramas cortas, ya que muchas de ellas se habrán perdido como consecuencia de la aparición de mutaciones deletéreas (Dermitzakis, 2010).

En la sección 1.2 hemos discutido cómo las premisas de los modelos de evolución neutral no se cumplen en la población humana. Por tanto, la aplicación de los modelos genéticos poblacionales está limitada por el incumplimiento de

estas asunciones. Así, el uso de un reloj molecular no es tan lineal ya que hay determinadas regiones del genoma sometidas a tasas de cambio evolutivo distintivas.

A continuación vamos a exponer las fuentes de información genética que se están utilizando hoy para reconstruir el pasado evolutivo de las poblaciones humanas.

2. Marcadores moleculares

2.1. Organización del genoma humano y variabilidad genómica

El material genético humano está contenido en 22 pares de autosomas y un par de cromosomas sexuales (XX o XY); también se encuentra presente fuera del núcleo en forma de ADN mitocondrial (ADNmt). Los más de 3.000 millones de nucleótidos que componen el genoma humano pueden clasificarse por su función. Hay secuencias codificantes (genes o *exones*) que llevan instrucciones para sintetizar proteínas, y secuencias no-codificantes. Estas últimas comprenden ~98,5% del ADN nuclear y, debido a su (aparente) falta de funcionalidad, no se encuentran bajo las mismas restricciones evolutivas que los genes, por lo que pueden exhibir mayor variabilidad (Rubicz et al., 2006; Antonarakis, 2010).

Los 'marcadores genéticos' se definen como caracteres que pueden ser usados para identificar a las poblaciones en base a su presencia, ausencia o alta frecuencia en unas poblaciones y baja en otras (Crawford, 1973). Desde finales de la década de 1980, la generalización en su uso se ha favorecido por los enormes avances tecnológicos, entre los que destacan el desarrollo de métodos para la extracción rápida del ADN, la aplicación de las enzimas de restricción a la genética molecular y, muy notablemente, el descubrimiento de la reacción en cadena de la polimerasa (PCR) y la secuenciación automatizada del ADN (Rubicz et al., 2006).

Habitualmente, se emplean de manera indistinta los términos 'variante genética', 'mutación' o 'polimorfismo'. Sin embargo, de manera estricta el término 'polimorfismo' se define como un carácter mendeliano que existe en la población en al menos dos fenotipos, ninguno de los cuales ocurre a una frecuencia menor del 1% (Antonarakis & Cooper, 2010).

Las variaciones que se detectan en la secuencia nucleotídica se pueden clasificar según su extensión y naturaleza, así como por su tasa de cambio evolutivo (ver **Figura I-6**).

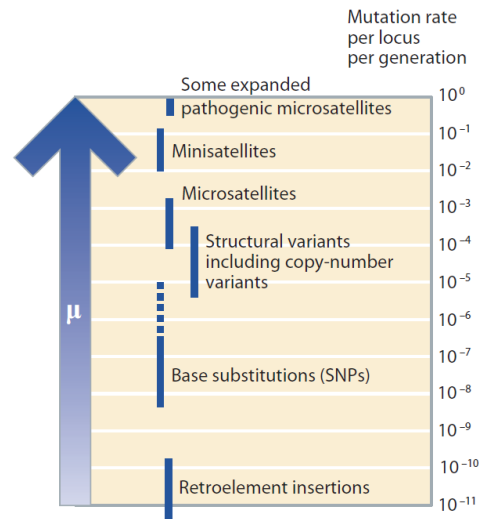


Figura I-6. Resumen de las tasas de mutación (μ) estimadas para los diferentes tipos de polimorfismos. Fuente: Jobling et al., 2013.

Single Nucleotide Polymorphisms (SNPs)

Se conocen como polimorfismos bialélicos, binarios o de un único nucleótido, y evolucionan a una tasa μ relativamente lenta ($2,5 \times 10^{-8}$ mutaciones/nucleótido, Nachman & Crowell, 2000). Estas sustituciones pueden ser ‘transiciones’ –cambio de una base por una de la misma naturaleza, es decir purina por purina o pirimidina por pirimidina– o ‘transversiones’ –cambio de una purina por una pirimidina o viceversa. El primer caso ocurre con más frecuencia que el segundo.

La baja tasa de mutación μ a la que están sometidas estas alteraciones de secuencia hace poco probable que hayan ocurrido más de una vez en el mismo sitio durante la historia del hombre moderno. Por tanto, si los individuos comparten la misma variante, puede asumirse una ancestría común, es decir, que la variante sea IBD (*Identical By Descent*) (Rubicz et al., 2006). Esta situación es algo más compleja para ciertas regiones genómicas (como por ejemplo la del ADNmt), lo que se abordará en la sección 3.3.

Los SNPs, son, en la actualidad, la variación de secuencia que se estudia con más profusión. De hecho, nuevas tecnologías y aproximaciones metodológicas están basadas en un barrido simultáneo desde cientos de miles a millones de SNPs a través del genoma humano completo (*Genome-Wide studies*). Estos análisis GWs pueden proporcionar visiones más completas de la historia poblacional humana ya que están basados en múltiples *loci* independientes. En concreto, son útiles para inferir tiempos de divergencia poblacional, migración y mezcla, cambios en tamaños poblacionales y otras facetas demográficas (Pugach & Stoneking, 2015). Sin embargo, estos avances en el conocimiento no carecen de inconvenientes. El mayor problema es el fenómeno conocido como *ascertainment bias* (que se podría traducir como ‘sesgo de verificación’) o sesgo generado por la selección y caracterización de marcadores en un sub-conjunto de muestras. Es decir, si se seleccionan ciertos SNPs en función de su elevado grado de

polimorfismo en una población determinada, su variabilidad puede ser artificialmente incrementada en esa población, con relación al resto (Handley et al., 2007). Una consecuencia de esta situación es la sobreestimación del nivel de heterocigosis de las poblaciones europeas con respecto a otras poblaciones.

Variable Number of Tandem Repeats (VNTRs)

Se trata de repeticiones en tándem de número variable. La mayor variabilidad de estos *loci* (debido a su estructura) provoca, al contrario que los SNPs, que los alelos con el mismo tamaño y secuencia no tengan por qué ser necesariamente IBD, es decir, ellos pueden ser idénticos por naturaleza (*Identical By State*, IBS), como un reflejo de evolución convergente (Jobling et al., 2013).

Los VNTRs se clasifican en función del tamaño de sus unidades de repetición. Así, los ‘microsatélites’ o STRs (*Short Tandem Repeats*) están compuestos por pequeñas unidades de 1-7 pares de bases (pb) de longitud con un tamaño total de secuencia habitualmente menor de 350 pb. Su tasa de cambio varía, dependiendo del *locus* considerado. Los alelos con mayor número de repeticiones tienden a mutar más rápidamente. La tasa de mutación general estimada de los STRs es de 10^{-4} a 10^{-3} mutaciones por locus y por generación. Aunque se considera que los STRs son selectivamente neutros, hay evidencias de que ciertos trinucleótidos están implicados en enfermedades humanas, por ejemplo, la repetición ‘CAG’ en la enfermedad de Huntington. Los STRs son, junto con los SNPs, las dos fuentes de variación genómica principal empleadas en el estudio de la historia genética humana.

Si las unidades de repetición son mayores (de 8 a 100 pb), estos elementos se denominan ‘minisatélites’. La tasa de mutación de estas variaciones es mayor que la de los propios STRs (10^{-2} – 10^{-1} por locus y por generación). Estos polimorfismos exhiben gran variación entre individuos y, por ello, son particularmente útiles en el campo forense (*DNA fingerprinting*) (Rubicz et al., 2006).

Copy Number Variants (CNVs)

Variaciones estructurales genómicas a gran escala en las cuales hay repeticiones en tándem de 50 Kb a 5Mb de largo y presentes en un número de copias variable. Estas variantes también incluyen duplicaciones y deleciones a gran escala e involucran cambios en la estructura de los cromosomas (Antonarakis, 2010).

Otras variaciones genómicas

Por último, nos encontramos con polimorfismos que no se usan tan frecuentemente, como por ejemplo los ‘satélites’ (repeticiones en tándem a gran escala) o los ‘arrays teloméricos’ (estructuras codificantes localizadas al final de los cromosomas, y frecuentemente clasificadas como minisatélites).

Otro tipo de variación de secuencia son los elementos móviles. Los más frecuentes en mamíferos son los denominados ‘retrotransposones no virales’ y entre los mismos destacan los ‘elementos intercalados cortos’ (*Short Interspersed Nuclear Elements, SINEs*) que engloban ~300 pb (Antonarakis, 2010). En muchas de estas secuencias humanas se detectó una diana de restricción para el enzima *Alu I*, por lo que estos elementos, también conocidos como ‘familia *Alu*’ (Lodish et al., 2003), se han aplicado ampliamente en el estudio de la diversidad genética humana (Batzer et al., 1994; Comas et al., 2000; Nasidze et al., 2001; Calderón et al., 2003).

2.2. La herencia uniparental humana

La mayor parte del genoma humano se hereda de los dos padres y, por tanto, sufre reordenación debido a la recombinación meiótica. Sin embargo, existen dos segmentos genómicos que escapan a esta regla general, siendo transmitidos por uno solo de los padres. El ADNmt se hereda por vía materna y lo poseen tanto mujeres como varones, mientras que el cromosoma Y (C-Y) lo transmiten los padres a los hijos varones exclusivamente.

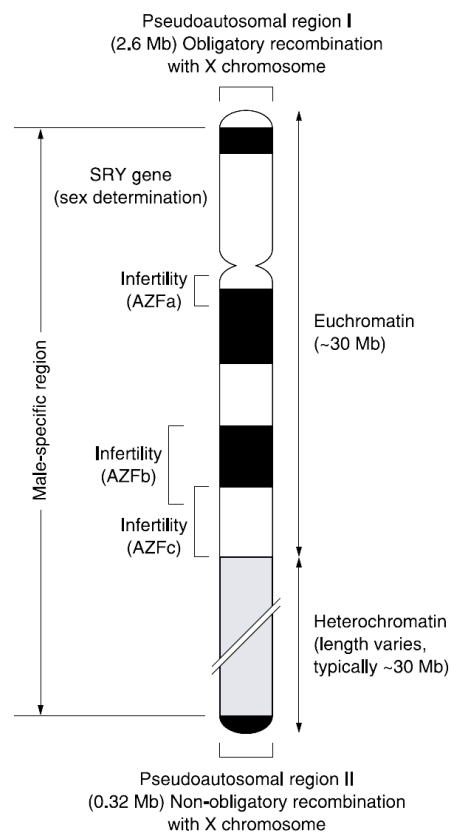
La especificidad sexual en el patrón de herencia implica que los marcadores genéticos incluidos en estas regiones contienen un registro de procesos demográficos y poblacionales específicos de mujeres y varones. De hecho, durante más de 20 años, las cuestiones acerca del pasado evolutivo humano se han intentado responder mediante el estudio de estos marcadores ‘haploides’ o ‘uniparentales’ (Underhill & Kivisild, 2007). El modo de transmisión en bloque de la molécula de ADNmt y de la NRY (escapando a la recombinación) explica que los marcadores haploides sean muy adecuados para la estimación de tiempos de coalescencia (TMRCA) y la reconstrucción filogenética de las poblaciones. Cualquier mutación que ocurra en la línea materna/paterna se transmitirá a futuras generaciones. Además, estos sistemas haploides tienen un *Ne* estimado de $\frac{1}{4}$ con respecto al de los marcadores autosómicos, lo que tiene considerables repercusiones en términos de historias demográficas (ver sección 3.2).

El C-Y comprende unos 60 Mb, de los cuales solo 3 Mb experimentan recombinación con el cromosoma X (región pseudoautosómica, ver **Figura I-7**). El resto se denomina habitualmente *non-recombining region of the Y chromosome* (NRY) o *male-specific region of the Y chromosome* (MSY). Este segmento contiene unos 25 genes, la mayor parte involucrados en la espermatogénesis (Jobling & Tyler-Smith, 2003). Los marcadores más ampliamente usados del C-Y son los STRs y SNPs. Los SNPs se usan para definir los linajes principales del C-Y mientras que los Y-STRs se utilizan para caracterizar la variación dentro de esos linajes (Rubicz et al., 2006). El hecho de que los Y-STRs sean altamente polimórficos justifica su aplicación (habitualmente mediante *kits* que analizan simultáneamente varios microsatélites) en el campo forense y de identificación personal.

La otra región no-recombinante del genoma humano es el ADNmt. En la siguiente sección se expondrán detalladamente las características de este genoma así como sus principales aplicaciones para el análisis de la diversidad humana.

El estudio de ambos marcadores uniparentales ha trazado los orígenes del hombre moderno (*Homo sapiens* 'moderno') en el continente africano, lo que apoya la evidencia fósil y paleoantropológica. Las estimas del TMRCA son alrededor de 172.000 ± 50.000 años (BP, *before present*) para el ADNmt (Ingman et al., 2000) y de 59.000 años BP (40.000-140.000) para la filogenia del C-Y (Thomson et al., 2000). El hecho de que la edad de coalescencia para todos los linajes paternos sea más reducida que para el caso de los linajes maternos ha sido explicado en función de procesos demográficos específicos según el sexo y de fenómenos selectivos (Wilder et al., 2004; Tishkoff & Gonder, 2006; Heyer et al., 2012). La diferenciación entre las poblaciones humanas varía también en función de que los análisis genómicos estén basados en el ADNmt o el C-Y. Habitualmente, se detecta una mayor proporción de diversidad debida a la variación entre poblaciones cuando se analizan los linajes paternos. Esta situación se ha fundamentado en el fenómeno de la patrilocalidad, que tiene como consecuencia que las moléculas mitocondriales migren más que las referentes al C-Y (Jobling et al., 2013).

Figura I-7. Estructura del cromosoma Y humano. Fuente: Rubicz et al., 2006.



3. El genoma mitocondrial como marcador poblacional

3.1. La mitocondria: biología celular, origen y características

La mitocondria es un orgánulo que posee un papel central en el metabolismo de la célula y se encuentra, por tanto, involucrada en procesos celulares fundamentales como la respiración, la fosforilación oxidativa (OXPHOS) y la apoptosis celular (Legros et al., 2004; Crimi & Rigolio, 2008). Estos orgánulos con doble membrana están presentes en todas las células nucleadas de los mamíferos. La densidad mitocondrial varía de un tipo celular a otro, en función de la dependencia del tejido a la OXPHOS para la provisión energética. De este modo, las neuronas y las células del músculo cardíaco y esquelético muestran una alta densidad del orgánulo (Schapira, 2006).

La cadena respiratoria mitocondrial está compuesta por 5 complejos enzimáticos situados en la membrana interna (ver **Figura I-8**). Cada complejo está formado por varias subunidades y en total comprenden 85 subunidades (Schapira, 2006).

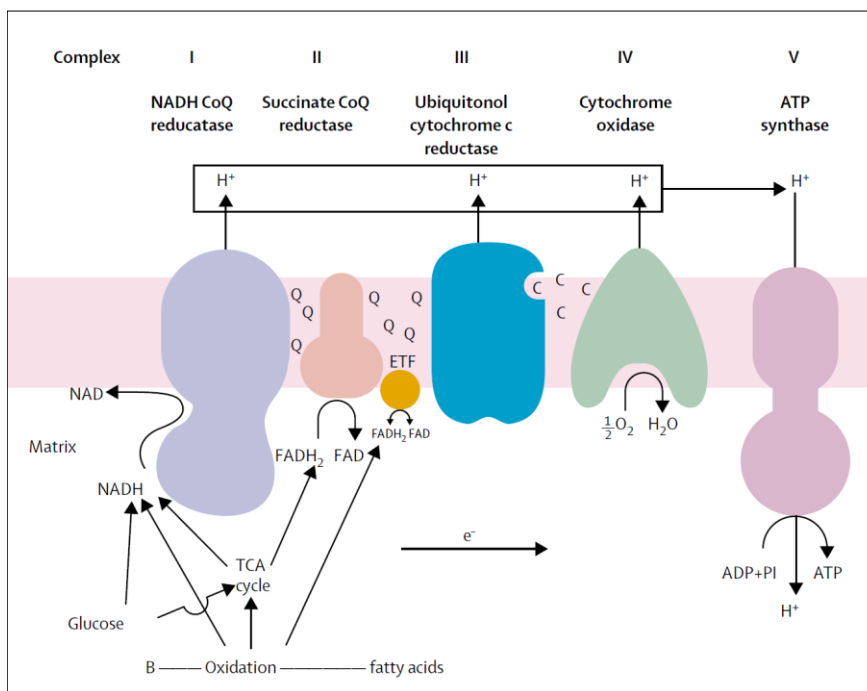


Figura I-8. Estructura funcional de la cadena respiratoria mitocondrial y del sistema de fosforilación oxidativa. Fuente: Schapira, 2006.

El proceso consiste en una serie de reacciones de oxidorreducción a través de los complejos de la cadena. Los cofactores reducidos (NADH y FADH₂) generados en el metabolismo intermedio de carbohidratos, proteínas y grasas, donan electrones al complejo I y II. Los electrones atraviesan la cadena *vía* ubiquinona y citocromo C a los complejos III y IV y se establece un gradiente electroquímico

por el bombeo de protones hacia fuera de la membrana mitocondrial interna. Todo este proceso es utilizado por el complejo V para sintetizar ATP, componente usado en los procesos metabólicos de la célula (Chinnery, 2006; Schapira, 2006).

Dentro de la mitocondria se producen la mayor parte de las especies (formas) reactivas derivadas del oxígeno (*Reactive Oxygen Species*, ROS), las cuales pueden dañar proteínas, lípidos y ADN (Ballard & Whitlock, 2004). Ese deterioro del material genético se ha propuesto como explicación al envejecimiento celular ('teoría de los radicales libres', Harman, 1992) ya que la producción de estos compuestos se incrementa con la variable edad.

Otro aspecto interesante en relación con este orgánulo celular se centra en su origen evolutivo. Actualmente, existe un consenso científico que apunta a que la mitocondria es el vestigio de un organismo procariota que formó parte de una relación simbiótica con otra célula. La hipótesis de un origen endosimbiótico de la mitocondria fue postulada por la bióloga norteamericana Lynn Margulis (Margulis, 1970, 1981) y se basa, fundamentalmente, en la posesión de un genoma independiente, de tipo circular y no aislado del citoplasma por un compartimento. Esta bacteria endosimbionte ancestral pertenecería a la división α de las proteobacterias (Yang et al., 1985). La relación entre la célula hospedadora y la procariota evolucionó con el tiempo, acompañada de la pérdida de genes redundantes y la transferencia de genes desde el organismo procariota al núcleo eucariota (Scheffler, 2001). Como consecuencia de este proceso, la mitocondria actual no es un organismo autónomo sino que es dependiente de su hospedador. De hecho, como se verá en la siguiente sección, la mayor parte de la información genética para la biogénesis mitocondrial y su función reside en el genoma nuclear (Gray et al., 1999).

3.2. El ADN mitocondrial: estructura y propiedades

Junto con los cloroplastos, las mitocondrias son los únicos orgánulos sub-celulares que cuentan con su propio genoma. Aunque la mayoría de las 1.000 proteínas de la mitocondria se codifican en el núcleo de la célula, las 13 proteínas codificadas por la molécula de ADNmt circular son esenciales para la respiración y la OXPHOS. La importancia de estas funciones mitocondriales para el metabolismo celular explica que, en la evolución de los mamíferos, la molécula de ADNmt haya permanecido estable en cuanto a su contenido y organización génica (Scheffler, 2001).

Desde el punto de vista estructural, el genoma mitocondrial humano consiste en una molécula de ADN circular de doble hebra que comprende 16.569 pb. Las dos hebras de la molécula se denominan 'cadena pesada' (H, *heavy*), la cual es rica en guaninas, y 'cadena ligera' (L, *light*) que es rica en citosinas (Jobling et al., 2013). Además de los 13 genes críticos para la producción de energía celular, esta molécula también contiene 24 genes que codifican la maquinaria transcripcional para el propio genoma (22 ARNt y 2 ARNr) (Ballard & Whitlock, 2004).

La mayor parte del ADNmt (~90%) está compuesto por regiones codificantes (ver **Figura I-9**). El fragmento restante se denomina 'región control' debido a que se encuentra implicado en el control de la replicación y transcripción del genoma mitocondrial. Esta región de unos 1.100 pb se conoce también como *displacement loop (D-loop)* y es la parte más variable del genoma mitocondrial (Chinnery, 2006). El carácter polimórfico de la región control justifica que haya sido, durante décadas, el foco de atención en los estudios de diversidad mitocondrial humana.

Este fragmento se extiende desde la posición 16024 a la 576 de la secuencia de referencia (ver sección 3.3). Su variabilidad, sin embargo, no es homogénea a lo largo de toda su extensión, lo que permite definir 3 regiones (Chinnery, 2006):

- Región hipervariable I (HVS-I): nps 16024-16365.
- Región hipervariable II (HVS-II): nps 73-340.
- Región hipervariable III (HVS-III): nps 438-576.

De los tres fragmentos, los más polimórficos son HVS-I y HVS-II. Un reciente estudio estimó que las tasas de mutación para estas regiones son de 1 mutación cada 16.677 años y 1 mutación cada 22.388 años, respectivamente (Soares et al., 2009). Del mismo modo, existen posiciones específicas en el genoma mitocondrial más propensas a sufrir variación ('hotspots mutacionales') (Malyarchuk et al., 2002).

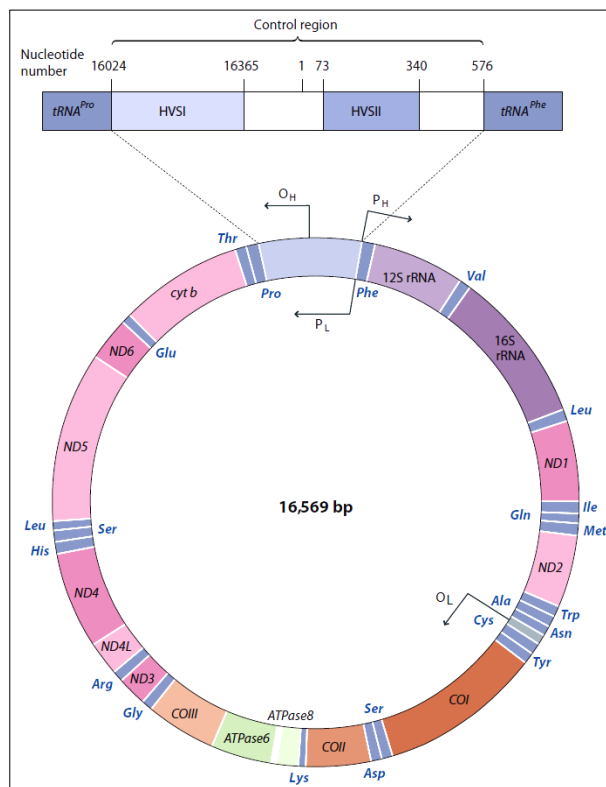


Figura I-9. Representación de la molécula de ADNmt, con la localización de los genes, de los orígenes de replicación de las cadenas ligeras y pesadas (O_L y O_H , respectivamente) y de los promotores de transcripción de ambas hebras (P_L y P_H). En la parte superior se destaca la región control y su estructura. Fuente: Jobling et al., 2013.

El ADNmt, por su localización específica dentro de la mitocondria, posee unas propiedades que diferencian a esta molécula del genoma nuclear. Estas características interactúan entre sí y fundamentan las numerosas aplicaciones del ADNmt como herramienta molecular.

Alta tasa de cambio evolutivo

El genoma mitocondrial tiene asociada una elevada tasa de mutación. Se ha calculado en Primates superiores una tasa de 5 a 10 veces mayor que la del genoma nuclear (Brown et al., 1979). Sin embargo, como se ha comentado anteriormente, esta tasa de cambio no es homogénea dentro de la molécula. La región control evoluciona a una tasa superior a la región codificante, e, incluso, la misma región control es heterogénea a este respecto ya que posee fragmentos más hipervariables que otros (destacando la HVS-I). Este escenario favorece que exista un elevado grado de variabilidad entre individuos en lo que se refiere a su herencia materna (Castro et al., 1998).

En este contexto de cambios evolutivos, se podrían indicar dos fuentes de variación genética mitocondrial. En primer lugar, la que se puede producir en la línea germinal materna, debido al particular patrón de herencia de la mitocondria (ver más abajo). Por otro lado y más centrado en la línea somática, la acumulación de cambios debidos a la edad, en el marco del estrés oxidativo que sufre el ADNmt como consecuencia de los procesos fisiológicos mitocondriales (Crimi & Rigolio, 2008).

En los últimos años se han realizado grandes esfuerzos en la estimación de las tasas de mutación del ADNmt humano. La calibración del reloj molecular es esencial para las posteriores estimaciones de edades de coalescencia, y, a su vez, para realizar inferencias sobre la demografía y la historia evolutiva de las poblaciones humanas. Numerosos estudios han abordado este punto (Excoffier & Yang, 1999; Sigurđardóttir et al., 2000; Mishmar et al., 2003; Kivisild et al., 2006a; Endicott & Ho, 2008; Henn et al., 2009, entre otros). Sin embargo, en la mayoría de los mismos no se tiene en cuenta nada más que una fracción del genoma mitocondrial, o bien no se valoran adecuadamente las diferencias entre la acumulación de sustituciones dependiendo de la región genómica. Estos factores han sido considerados por Soares et al. (2009) para la calibración del reloj mitocondrial y, las tasas incluidas en el citado trabajo están siendo ampliamente utilizadas en la actualidad.

Elevado número de copias por célula

Cada célula humana puede contener desde cientos a miles de moléculas de ADNmt. Esta naturaleza poliploide es la base de una relevante característica del sistema genético mitocondrial que enfrenta los conceptos de '*homoplasmia*' y '*heteroplasmia*'. El primer término indica la identidad de secuencia entre todas las copias del ADNmt, el último hace referencia, sin embargo, a la existencia de

una mezcla de varios genotipos mitocondriales (Taylor & Turnbull, 2005; Strachan & Read, 2011). Esta fracción de copias diferenciales proviene de mutaciones somáticas que surgen y se acumulan con el envejecimiento, producidas principalmente por las ROS (Schapira, 2006). La heteroplasmia, por tanto, se asocia a la elevada tasa de mutación asociada al genoma mitocondrial.

Las enfermedades mitocondriales se caracterizan por la coexistencia del tipo salvaje y mutante del ADNmt (Schapira, 2006) y se considera que existe un umbral en la fracción del tipo mutante que determina la expresión clínica de estos desordenes (Taylor & Turnbull, 2005).

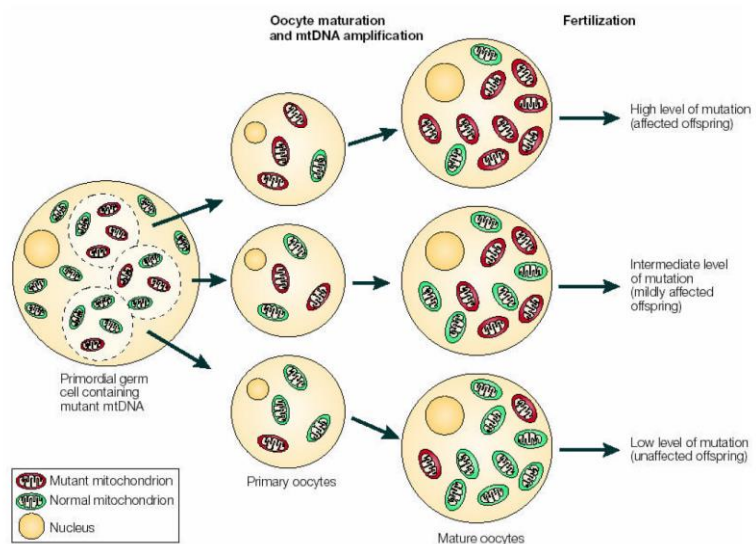
Además, la existencia de numerosas copias de ADNmt por célula es una característica clave que explica el por qué este genoma es una herramienta molecular clave en el campo de la genética forense y el ADN antiguo (ADNa). Ambas aplicaciones se encuentran limitadas fundamentalmente por la calidad o cantidad de la muestra fuente para la extracción de material genético. Por tanto, la posesión de varias copias potencia el éxito de la amplificación de ADN (Budowle et al., 2003; Pakendorf & Stoneking, 2005).

Patrón de herencia

El particular patrón de transmisión materno del ADNmt se demostró en el hombre por primera vez en la década de 1980 (Giles et al., 1980). La base explicativa está en que en el momento de la concepción, las mitocondrias paternas no penetran en el óvulo. Este fenómeno podría tener una base evolutiva, ya que de este modo se evitaría el conflicto entre los genomas mitocondriales materno y paterno (White et al., 2008). Sin embargo, puede haber raras excepciones a esta vía hereditaria, tal como el detectado por Schwartz & Vissing (2002) en el que un varón se encontraba afectado por una patología muscular causada por una deleción en el gen ND2 (NADH deshidrogenasa). Los análisis demostraron que la molécula de ADNmt afectada había sido heredada a través del padre.

En el ciclo experimentado por las células germinales femeninas se manifiesta un fenómeno de 'cuello de botella', mediante el cual un pequeño número de moléculas de ADNmt sirve de origen para las que formarán parte de la descendencia (ver **Figura I-10**). Durante la embriogénesis, hay una gran reducción del número de copias y, durante la ovogénesis, se produce un aumento muy significativo de ellas. Por tanto, solo una fracción del ADNmt original repoblará las sucesivas generaciones (Bergstrom & Pritchard, 1998). Relacionando este fenómeno con otra característica del genoma mitocondrial, se ha propuesto que el cuello de botella contribuya a la alta carga mutacional del ADNmt (Neiman & Taylor, 2009).

Figura I-10.
Representación del cuello de botella mitocondrial durante la producción de ovocitos primarios. Los niveles variables de ADN mutado afectarán a la descendencia. Fuente: Taylor & Turnbull, 2005.



Ausencia de recombinación

Otra particularidad del ADNmt con respecto al genoma nuclear es la ausencia de recombinación. Aunque la maquinaria necesaria para llevar a cabo estos procesos de intercambio genético está presente en la mitocondria, la recombinación, tanto intergenómica (entre diferentes moléculas de ADNmt) como recíproca (la misma molécula) es un fenómeno muy raro en animales. Sin embargo, se ha encontrado alguna evidencia de recombinación en tejido somático humano (Kraytsberg et al., 2004) sin repercusión para la descendencia (Slate & Gemmell, 2004).

Esta falta de recombinación, unida a la vía de transmisión uniparental y escenarios de cuello de botella, provoca que la molécula de ADNmt se vea intensamente afectada por la deriva genética por lo que se favorece la ocurrencia y permanencia de variantes deletéreas. Estas cuestiones pueden resultar paradójicas, debido al papel crucial de la mitocondria en la fisiología celular. Se ha propuesto que la no-recombinación del genoma mitocondrial respondiese a una presión selectiva para el mantenimiento de ciertos bloques de genes co-adaptados (Castro et al., 1998).

Reducido tamaño efectivo

El tamaño efectivo (N_e) de la molécula de ADNmt se considera $\frac{1}{4}$ con respecto al ADN nuclear y coincide con el del cromosoma Y, a igual proporción de sexos (Hartl & Clark, 1989). Esta propiedad conduce a que los marcadores haploides sean más sensibles a la deriva genética y a los efectos demográficos, tales como los cuellos de botella o el efecto fundador (Jorde et al., 2000).

Todas estas consideraciones son básicas para fundamentar la aplicación del ADNmt en el estudio de la historia evolutiva de las poblaciones humanas actuales y del pasado (ver sección 3.3). Sin embargo, el genoma mitocondrial se analiza también en el campo de la Biomedicina, debido a sus repercusiones clínicas y su influencia en el envejecimiento humano.

Las patologías mitocondriales tienen un rango de gravedad y expresión clínica muy variable. Las primeras enfermedades relacionadas con el ADNmt se describieron a finales de la década de 1980 en la investigación de dos síndromes, Kearns-Sayre y Neuropatía Óptica Hereditaria de Leber (LHON) (Holt et al., 1988; Wallace et al., 1988). En ambos casos, se observaron delecciones a gran escala en biopsias musculares de pacientes y un patrón concreto de herencia materna de la enfermedad. El papel de las mutaciones mitocondriales en la patología es complejo, ya que se pueden encontrar diferentes situaciones que van, desde mutaciones con alta penetrancia (las cuales por sí mismas son capaces de desarrollar una patología) a mutaciones con penetrancia débil, más bien consideradas como factores de riesgo que, unidas a un determinado perfil genético poblacional o a unas condiciones ambientales, pueden provocar el desarrollo de la enfermedad (Herrnstadt & Howell, 2004). Ciertas variantes mitocondriales han sido relacionadas con la predisposición a patologías de gran proyección social como la Diabetes de tipo 2, el Alzheimer y la enfermedad de Parkinson. En estos dos últimos casos, se ha propuesto que mutaciones causantes de alteraciones en la actividad OXPHOS pudieran interactuar con el desarrollo de la enfermedad (Schapira, 2006).

Del mismo modo, en los últimos años se han publicado numerosos estudios que asocian la pertenencia a un determinado haplogrupo con la susceptibilidad a padecer un amplio rango de desórdenes (Carrieri et al., 2001; Amar et al., 2007; Chinnery et al., 2007; Hudson et al., 2007; Feder et al., 2008; Gaweda-Walerych et al., 2008). Sin embargo, hay que ser cuidadoso al extrapolar datos extraídos de poblaciones muy generales a otros escenarios consistentes en poblaciones caracterizadas por una estructura genética concreta.

Como se ha comentado anteriormente, se ha propuesto que la acumulación progresiva de mutaciones somáticas en el ADNmt durante la vida lleve a un inevitable decremento en la función mitocondrial (Harman, 1992). En este contexto, actuarían las variantes ROS generadas normalmente a niveles muy bajos en la cadena respiratoria mitocondrial. Hay evidencia de que algunas mutaciones concretas, como la 8993T>G, conduzcan al incremento de ROS. Este incremento provocaría, a su vez, más mutaciones en un ciclo de acumulación de daños oxidativos durante el envejecimiento, lo que resulta en una eventual pérdida de las funciones celulares y tisulares a través de una combinación de insuficiencia energética, defectos de señalización, apoptosis y senescencia replicativa (Taylor & Turnbull, 2005).

Las mutaciones en el ADNmt puedan afectar a la longevidad. Estos estudios se han desarrollado fundamentalmente en población japonesa (Zhang et al., 2003; Niemi et al., 2005) aunque los resultados no son necesariamente convergentes entre poblaciones humanas (Costa et al., 2009).

3.3. *El ADN mitocondrial y la historia evolutiva humana*

A pesar de que el ADNmt representa solo una pequeña fracción del genoma humano, ha sido una de las herramientas genéticas más utilizadas en las últimas tres décadas. Las nuevas variantes mitocondriales aparecen más frecuentemente de lo esperado (debido a la elevada tasa de mutación), y, además, están más sometidas a la deriva genética (debido a la ausencia de recombinación y al bajo Ne). Esta situación provoca que diferentes linajes mitocondriales en una población evolucionen independientemente unos de otros por una acumulación secuencial de mutaciones. Ello hace posible la reconstrucción de las relaciones filogenéticas entre diferentes linajes que están presentes en una población o en diferentes poblaciones.

Estrategias metodológicas de análisis

La publicación de la secuencia del ADNmt humano en *Nature* en 1981 (Anderson et al., 1981) supuso un acontecimiento clave para la genética humana. Esta secuencia se conoce como *Cambridge Reference Sequence* (CRS) o 'secuencia de Anderson'. Unos años más tarde se llevó a cabo una re-secuenciación de este genoma mitocondrial, corrigiendo algunas mutaciones erróneamente determinadas en el estudio original (Andrews et al., 1999), y estableciéndose la rCRS (*revised CRS*), la cual es la referencia en cuanto a la notación de variantes mitocondriales.

Las primeras aproximaciones al estudio de la variabilidad humana basadas en el genoma mitocondrial, se realizaron mediante análisis de restricción (*Restriction Fragment Length Polymorphisms*, RFLPs). Esta metodología implica el uso de enzimas que fragmentan el ADN y la detección de esos fragmentos. Los trabajos pioneros (Johnson et al., 1983; Cann et al., 1987) demostraron una correlación entre los patrones de restricción mitocondriales y los orígenes geográficos de los individuos analizados, detectándose una mayor variabilidad humana en poblaciones africanas. Estos resultados fueron la base para apuntar a África como la fuente de toda la variabilidad mitocondrial humana.

Posteriormente, la atención se centró en las regiones hipervariables mitocondriales, cuyo rendimiento a lo largo de las últimas décadas ha sido notable. Su variabilidad entre individuos y poblaciones ha permitido descender más concretamente a patrones específicos de población e inferir escenarios de flujo génico. La secuenciación de la región control se ha aplicado a poblaciones africanas (Vigilant et al., 1989), asiáticas (Horai & Hayasaka, 1990) y europeas (Di Rienzo & Wilson, 1991; Richards et al., 1996; Parson et al., 1998; Helgason et al., 2001, entre otras investigaciones). Esta información de secuencia, combinada con ciertas posiciones de la región codificante –genotipadas principalmente por RFLPs– ha sido el enfoque más ampliamente utilizado y el que ha ido asentado las bases de los conocimientos actuales acerca de las historias poblacionales de diferentes continentes. Esta dinámica ha permitido enriquecer nuestra

comprensión acerca de la magnitud de la diferenciación genética existente dentro y entre poblaciones humanas (Macaulay et al., 1999).

Otras aproximaciones al análisis de variantes mitocondriales (fundamentalmente de la región codificante) se realizan habitualmente mediante genotipados múltiples. Estas estrategias se emplean cuando la información de la región control no permite la discriminación entre individuos (Brandstätter et al., 2003) por lo que su importancia en el campo forense es indiscutible (Quintáns et al., 2004).

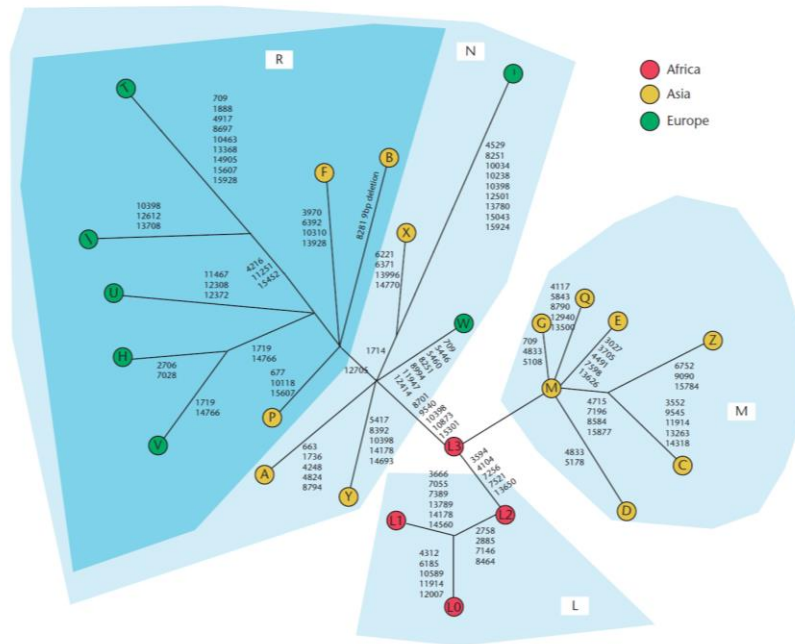
La máxima resolución que se puede alcanzar en el análisis del ADNmt es su secuenciación completa. Además de permitir una detallada clasificación filogenética y la detección de patrones de diversidad previamente ignorados, el estudio de genomas mitocondriales completos (*mitogenomas*) está siendo utilizado en la evaluación del peso de los factores evolutivos así como en la caracterización de la historia demográfica de las poblaciones. En los últimos años se ha convertido en la tendencia habitual, favorecida también, por la aplicación de nuevas tecnologías de secuenciación (que permiten superar limitaciones coste-tiempo). Habitualmente, estos estudios se centran en la disección de linajes mitocondriales específicos (Schönberg et al., 2011; Behar et al., 2012a; Soares et al., 2012; Achilli et al., 2013; Fernandes et al., 2015, entre otros).

Nomenclatura y notación

El concepto de *'haplogrupo'* surge a principios de la década de 1990 y se refiere a un grupo de secuencias mitocondriales filogenéticamente relacionadas, definidas por mutaciones compartidas y que presentan patrones geográficos de distribución espacial concretos (Pakendorf & Stoneking, 2005). El término *'haplogrupo'* se suele utilizar indistintamente al de *'linaje'*, *'clado'* o *'cluster'*. Este concepto se recoge por primera vez haciendo referencia a "*groups of related haplotypes*" (Torroni et al. 1993), siendo un *'haplotipo'* un conjunto de variantes mitocondriales relacionadas.

Los primeros haplogrupos del ADNmt (A, B, C y D) se definieron en población nativa americana (Torroni et al., 1992) y, poco más tarde, el mismo equipo delimitó los linajes H, I, J y K en población americana de ascendencia europea (Torroni et al., 1994). Con la caracterización de los haplogrupos T, U, V, W y X (Torroni et al., 1996) prácticamente la totalidad de la diversidad mitocondrial fuera de África estaba definida. La **Figura I-11** muestra la distribución de los haplogrupos en grandes *clusters* o agrupamientos geográficos.

Figura I-11. Árbol filogenético esquemático de los linajes mitocondriales humanos. Se indican las posiciones que determinan cada una de las ramificaciones. Fuente: Passarino & Rose, 2013.



Los estudios pioneros sobre la diversidad mitocondrial en poblaciones actuales y vivientes sentaron las bases de un campo en expansión que requería unas reglas cladísticas específicas. Richards et al. (1998) establecieron las normas para la estructuración jerárquica de los haplogrupos y sub-haplogrupos. Los linajes (haplogrupos) mitocondriales se designan con letras mayúsculas. Conforme se desciende en la genealogía materna, las ramificaciones se van nombrando con números y letras sucesivas desde la raíz a las ramas más terminales (*tip branches*) (H, H1, H1a, H1a1...). Cada sub-haplogrupo se caracteriza por una o varias mutaciones en la secuencia mitocondrial respecto a la rCRS. La definición de nuevos sub-haplogrupos se debe hacer sobre la base de un polimorfismo (preferentemente más de uno) que no sea *hotspot* mutacional (Kivisild et al., 2006b). Estas consideraciones previenen de ambigüedades filogenéticas y errores en el establecimiento de relaciones entre secuencias mitocondriales. Por último, es muy frecuente en la literatura observar un asterisco (*) a continuación del haplogrupo. Este símbolo indicaría miembros del linaje en cuestión que no pertenecen a ninguno de los sub-haplogrupos definidos dentro de ese linaje. Esta condición se denomina ‘*paragrupo*’.

Esta normativa general se ha aplicado también en el establecimiento de una filogenia del cromosoma Y humano (*Y-Chromosome Consortium, YCC, 2002*).

Consideraciones acerca del uso del ADNmt en filogenia y filogeografía humana

Las particularidades del ADNmt han convertido a esta molécula en un instrumento central en el estudio de relaciones genéticas entre poblaciones humanas. Sin embargo, hay ciertas limitaciones que se deben tener en cuenta en

su aplicación. En primer lugar, se trata de un solo *locus* y, por tanto, su historia no representa necesariamente la historia de una población o especie, sino que se encuentra sesgada hacia un sexo exclusivamente.

Por otro lado, hay evidencias de que el genoma mitocondrial se ha visto sometido a un fuerte efecto de la selección natural, en concreto, la *selección purificante*, la cual ha eliminado de la filogenia humana las mutaciones más deletéreas (Elson et al., 2004; Stewart et al., 2008). En este contexto, se ha observado, paradójicamente, que los genes mitocondriales evolucionan mucho más lentamente que lo que se podía predecir sobre la base de sus altas tasas de mutación. Además, la filogenia mitocondrial tiende a mostrar una mayor proporción de mutaciones sinónimas en ramas más antiguas, es decir, el *ratio* dN/dS (mutaciones no- sinónimas respecto a sinónimas) es menor en secuencias africanas que en europeas. Por tanto, los clados más recientes muestran más frecuentemente mutaciones no-sinónimas. Una hipótesis alternativa que se ha propuesto para explicar este fenómeno se centra en el clima como presión selectiva (Mishmar et al., 2003; Ruiz-Pesini et al., 2004). De acuerdo con esta premisa, los cambios medianamente deletéreos que provoquen una reducción en la eficiencia de la OXPHOS tendrían como consecuencia un incremento en producción de calor en detrimento de la generación de ATP. Esto supondría una ventaja adaptativa para las poblaciones humanas en un ambiente frío. Sin embargo, las ramas recientes mitocondriales muestran un exceso de mutaciones no-sinónimas que es independiente de la localización geográfica o latitud que habiten (Kivisild et al., 2006a). La consecuencia de este efecto de la selección natural es que se eliminan progresivamente las mutaciones levemente deletéreas y, por tanto, la tasa de mutación se amortigua conforme uno va hacia atrás en el árbol. Esto supone la infracción de la premisa del reloj molecular.

Otra limitación relevante con el uso del ADNmt deriva de su alta tasa de mutación, ya que se pueden generar ambigüedades filogenéticas causadas por las mutaciones recurrentes. Estos eventos de mutación convergente se denominan '*homoplasia*' y conducen a que varios árboles filogenéticos puedan ser inferidos a partir de un mismo *set* (conjunto) de secuencias (Macaulay & Richards, 2013).

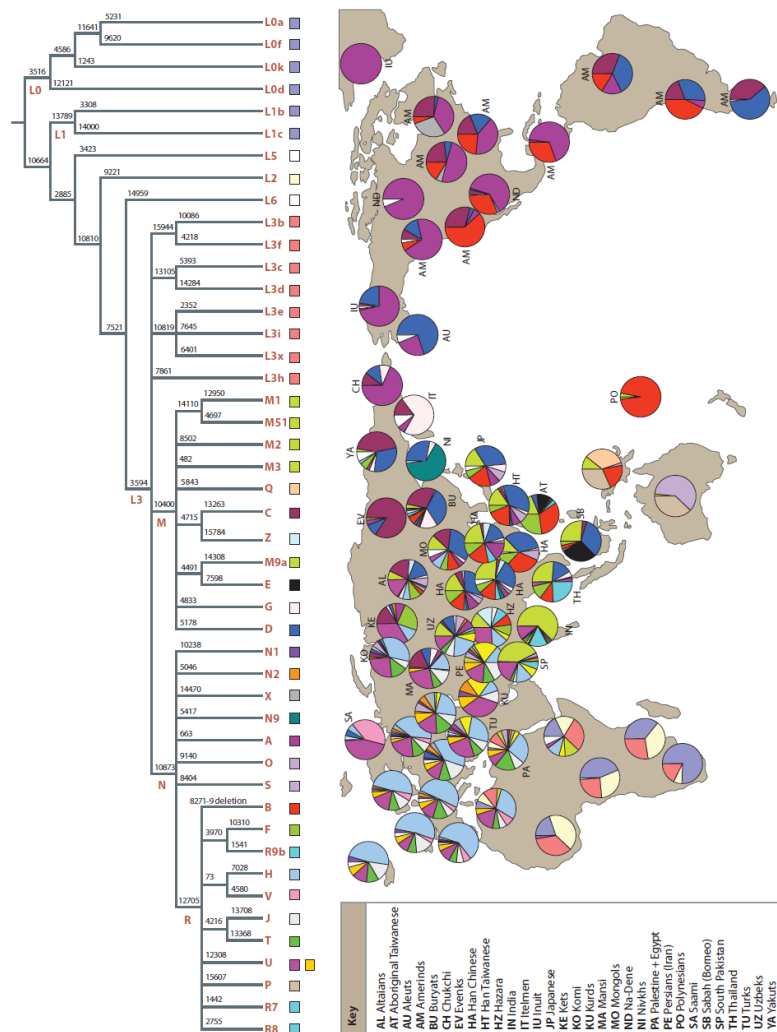
Por último, existen regiones mitocondriales que evolutivamente se han integrado en el genoma nuclear. Esto se conoce como NUMTs (*Nuclear Mitochondrial Pseudogenes*). La problemática surge ya que se pueden amplificar en lugar del ADNmt –fundamentalmente en muestras degradadas o de baja concentración de ADN– y oscurecer así las relaciones filogenéticas reales (Tishkoff & Gonder, 2006).

Breve historia de los linajes mitocondriales: dispersión y flujo génico de las poblaciones humanas modernas

La raíz del árbol mitocondrial –representada por el macro-haplogrupo L– parece encontrarse en África. Dentro de este linaje, los haplogrupos L0-L2 y L4-L6 son exclusivos africanos, mientras que el L3 es el clado parental de toda la diversidad

mitocondrial fuera de África. Como se puede observar en la **Figura I-12**, el linaje L3 tiene como ramas derivadas, además de varios sub-haplogrupos africanos (L3b, L3f, etc.), a los clados M y N. Estos dos grandes *clusters* mitocondriales marcan la salida del *Homo sapiens* 'anatómicamente moderno' de África (50.000-65.000 años BP, Behar et al., 2008) y el poblamiento del resto del globo, distribuyéndose el clado M principalmente por Asia y, el N por el continente europeo.

Figura I-12. Filogenia mitocondrial humana y distribución geográfica de haplogrupos. En el árbol se indican los polimorfismos definitorios de los linajes y sub-linajes. Fuente: Jobling et al., 2013.



El linaje N pronto derivó en el haplogrupo R. Tres ramas principales de R [R0 (que incluye al clado H), JT y U] representan alrededor del 80-90% de la diversidad mitocondrial total de Europa. De forma puntual, el clado H constituye ~40-50% del *pool* materno del Continente (Pereira et al., 2005a).

Tres acontecimientos poblacionales masivos han sido tradicionalmente considerados para explicar la variabilidad genética de las poblaciones europeas

(Pinhasi et al., 2012) (ver **Figura I-13**). En primer lugar, la colonización del Continente por parte del hombre anatómicamente moderno (*Anatomically Modern Humans*, AMH) hace más de 40.000 años (Mellars, 2006; Hublin, 2015). Desde la perspectiva de la herencia materna, el principal clado asociado a este asentamiento inicial ha sido el U, con una edad de coalescencia estimada ~55.000 años y un origen probable en el sudoeste de Asia (Fernandes et al., 2012).

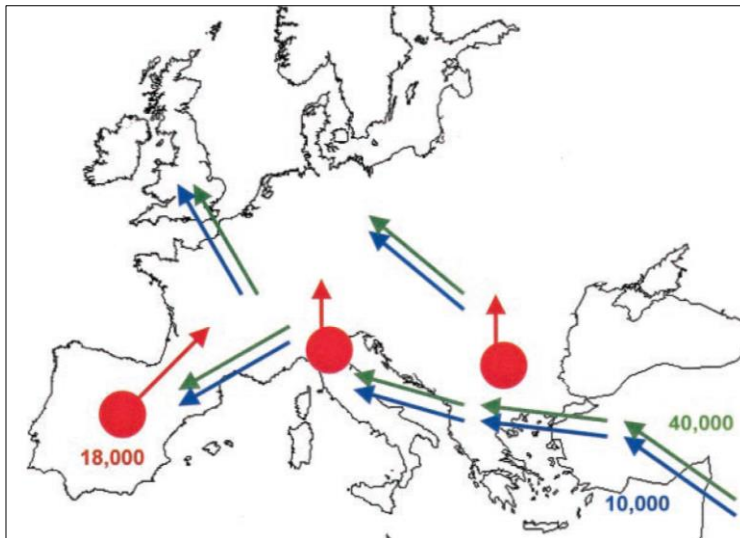


Figura I-13. Principales procesos demográficos en Europa. En verde se indica la colonización Paleolítica, en rojo las expansiones desde refugios del sur, y en azul, la difusión Neolítica. Los números son dataciones estimadas en años BP. Fuente: Barbujani & Bertorelle, 2001.

Cronológicamente, el segundo evento demográfico es la migración masiva de población hacia áreas refugio del sur de Europa como consecuencia de las condiciones climatológicas del último glacial (*Last Glacial Maximum*, LGM, hace unos 26.500-19.500 años, Clark et al., 2009) y las posteriores expansiones de recolonización del Continente. Señales mitocondriales de este fenómeno se han detectado en la actual distribución de los linajes V, H1 y H3 (Torroni et al., 1998, 2001; Achilli et al., 2004; Pereira et al., 2005a). Estos haplogrupos registran sus mayores frecuencias en el sudoeste europeo (Iberia), y por ello se ha propuesto que se re-expandieran desde esta región hacia el resto del Continente con la mejora de las condiciones ambientales tras el LGM. El fenómeno de recolonización no solo involucró a áreas del norte y centro de Europa, sino también al Norte de África, lo que ha sido revelado en un número importante de estudios (Cherni et al., 2009; Ennafaa et al., 2009; Ottoni et al., 2010).

La llegada del Neolítico no implicó solo una transición en el modo de vida, de una economía cazadora-recolectora a prácticas agrícolas y ganaderas, sino también cambios en la composición genética y características demográficas de las poblaciones humanas (Bellwood & Oxenham, 2008; Milisauskas, 2011). Este fenómeno, lejos de ser un proceso simple, consistió en una serie de movimientos puntuales tanto marítimos como terrestres de dispersión desde el Oriente

Próximo hacia Europa (Pinhasi et al., 2012). Los clados del ADNmt habitualmente ligados a la llegada del Neolítico a Europa occidental son el J y T. Sin embargo, un reciente estudio ha propuesto que estos linajes podrían reflejar dispersiones más tempranas (Pala et al., 2012).

Fundamentado en el modo de asimilación del Neolítico en Europa, su impacto en la composición genética de Europa ha suscitado un debate continuado para la Antropología y la Arqueología a lo largo de las últimas décadas. Hay dos modelos sintéticos de dispersión Neolítica: el modelo de '*difusión démica*' (reemplazamiento de los grupos Mesolíticos por la población inmigrante, Ammerman & Cavalli-Sforza, 1984) y el modelo de '*difusión cultural*' (transmisión de avances tecnológicos sin suponer apreciables movimientos de población). La herencia mitocondrial muestra que menos del 15% de los linajes europeos provienen del componente Neolítico y que la gran mayoría de los mismos comparten un sustrato que se data en tiempos peri-glaciales o post-glaciales (Soares et al., 2010). La contribución de los componentes Paleolítico/Neolítico es variable dependiendo del tipo de marcador genómico que se use, ya que los linajes paternos actuales muestran habitualmente una mayor proporción atribuible al Neolítico (Balaresque et al., 2010; Sjödin & François, 2011).

Al margen de los asentamientos y dispersiones humanas a gran escala dentro del Continente, el *pool* mitocondrial actual de Europa se encuentra enriquecido por ciertos linajes que indican episodios de flujo génico y migración desde áreas continentales vecinas. En este contexto, un foco de interés científico reside en el oeste del Mediterráneo como punto de encuentro entre Europa y África. En concreto, numerosos estudios se han centrado en determinar si el Estrecho de Gibraltar y áreas circundantes al mismo (por ejemplo el sur de Iberia) han actuado como una barrera o como un corredor para el flujo génico entre los dos continentes (Currat et al., 2010). Esta cuestión ha sido objeto de reiterada discusión, ya que dependiendo del marcador genético empleado, la proximidad entre las poblaciones europeas y africanas (norteafricanas) es mayor o menor. Se han encontrado evidencias de una fuerte barrera geográfica en el estudio de marcadores clásicos, inserciones *Alu* y de marcadores de los cromosomas X e Y (Bosch et al., 1997, 2001; Comas et al., 2000; Tomas et al., 2008). Sin embargo, otros marcadores como el sistema *GM* (alotipos de las inmunoglobulinas humanas) revelan la permeabilidad del Estrecho al movimiento humano (Coudray et al., 2006). Asimismo, el análisis del ADNmt ha mostrado importantes trazas genéticas africanas en Europa –fundamentalmente en su extremo occidental, Iberia– marcadas por los haplogrupos L y U6. La datación (edades evolutivas) de los acontecimientos que han dado lugar a la distribución actual de estos linajes en Europa es materia de debate en la actualidad (González et al., 2003; Plaza et al., 2003; Pereira et al., 2005b).

4. La Península Ibérica dentro del espacio mediterráneo: análisis de su diversidad matrilineal

Dentro de Europa y, particularmente, como un componente también de la cuenca mediterránea, la Península Ibérica ha sido y continua siendo una región protagonista para los estudios sobre diversidad genética humana. Diversos factores justifican este interés científico. En primer lugar, su localización geográfica es estratégica. Situada en el extremo más sudoccidental de Europa, representa la conexión más cercana del continente europeo con África. A esas particularidades hay que añadir la de estar geográficamente distante de los lugares donde ocurrieron los dos acontecimientos más relevantes de la evolución humana reciente: la salida de los AMH de África y la introducción de la agricultura (ver Calderón et al., 2006). Asimismo, la Península Ibérica se ha visto involucrada en episodios demográficos claves en la historia evolutiva humana, siendo el principal refugio meridional durante el LGM (Soffer & Gamble, 1990). Sus numerosos atractivos, por situación espacial y recursos naturales, han propiciado que la Península haya sido un foco de atracción para diversas culturas a través de la prehistoria y de la historia.

Por todas estas razones, la diversidad mitocondrial de las poblaciones ibéricas ha sido objeto de un análisis científico recurrente durante los últimos 20 años. (Côte-Real et al., 1996; Torroni et al., 1998, 2001; Achilli et al., 2004; Pereira et al., 2004, 2005a; 2005b, entre otros)

4.1. Primeros estudios y principales resultados

Las primeras aproximaciones al estudio de la diversidad matrilineal en Iberia se centraron en el norte del territorio, y, en concreto, en la población vasca. Los análisis basados en la región HVS-I (Bertranpetit et al., 1995) indicaron una baja diversidad de secuencia (comparada con las pocas poblaciones de las que se tenían datos disponibles en esos años) y una distribución de diferencias por pares de secuencias consistentes con una expansión poblacional reciente. En posteriores análisis, las muestras de vascos siguen destacando –en un contexto europeo– por sus bajos niveles de diversidad, lo que las diferencia significativamente del resto del Continente (Richards et al., 1996). Además, esta población se encuentra caracterizada por su elevada representación del linaje denominado inicialmente “*Group 1*” y que corresponde con el haplogrupo H (CRS y derivados). Estos resultados se han interpretado como una respuesta a los presumibles periodos de aislamiento y deriva genética que ha sufrido la población autóctona vasca.

Estudios basados en una más alta resolución filogenética siguen detectando una baja variabilidad genética en vascos (Martínez-Cruz et al., 2012), aunque también han revelado una heterogeneidad en el territorio debido a las presumibles fragmentaciones poblacionales causadas por condiciones orográficas y, por tanto, por un aislamiento local y un flujo génico limitado (Alfonso-Sánchez et al., 2008; García et al., 2011).

Otra de las poblaciones más ampliamente estudiadas desde el punto de vista del ADNmt también pertenece al norte de la Península Ibérica, pero en este caso, en su vertiente más atlántica. La población de Galicia muestra similitudes con la autóctona del País Vasco en cuanto a altas frecuencias de haplotipos CRS y baja diversidad genética (Salas et al., 1998). Sin embargo, en esta región noroccidental se han detectado, al contrario que en el País Vasco, una influencia restringida pero llamativa de linajes ligados al continente africano. Los perfiles subsaharianos (L) alcanzan entre el 2,5-3,0% de la variación mitocondrial total en Galicia (González et al., 2003; Álvarez-Iglesias et al., 2009), aunque es el extremo sudoeste ibérico el que registra las mayores frecuencias de L de toda la Península (sur de Portugal: 8,6%, González et al., 2003) Del mismo modo, la vertiente atlántica norte de Iberia destaca por una relativa elevada presencia del linaje norteafricano U6 (Galicia: 2,2%, norte de Portugal: 4,3%).

Estos gradientes opuestos de influencia africana en el extremo atlántico de la Península (de altas frecuencias de U6 en el norte y prevalencia de los linajes L en el sur) se han observado también en población exclusivamente portuguesa (Pereira et al., 2005b). De todos estos resultados se revela la enorme importancia de la región más occidental de la Península para la asimilación de genes africanos.

El litoral cantábrico ha seguido siendo protagonista para el estudio de la herencia materna, bien sea analizando poblaciones aisladas o muestras de población general de territorios como Asturias o Cantabria (Maca-Meyer et al., 2003a; Cardoso et al., 2010; Pardiñas et al., 2012).

Una población interesante (distante ~100 km del noroeste africano) es la autóctona de las Islas Canarias. El estudio de su herencia materna ha revelado la existencia de una remarcable influencia del norte de África, fundamentado en el sub-clado U6b y la rama derivada U6b1a, la cual se considera autóctona canaria y que podría reflejar la colonización inicial del archipiélago (Rando et al., 1999).

4.2. Análisis de las poblaciones meridionales de Iberia

Andalucía es una región extensa y densamente poblada del sur de la Península Ibérica y posee una larga historia modelada por numerosas migraciones desde distintas regiones geográficas (preferentemente mediterráneas) y a diferentes tiempos. Pese a estas particularidades, de fuerte impacto en el poblamiento de la Península Ibérica, la composición y estructura del ADNmt de la población andaluza no ha sido estudiada nunca en profundidad. En el marco de la investigación formal, esta región meridional ha estado ensombrecida por el papel principal manifestado por la franja cantábrica, en los estudios sobre diversidad genética humana. Sin embargo, relevantes cuestiones como la presencia y orígenes de las trazas africanas en Iberia, el papel del Estrecho de Gibraltar en escenarios de flujo génico, la influencia de otras áreas geográficas y poblaciones –especialmente de la cuenca del Mediterráneo– en el *pool* mitocondrial de la Península, solo pueden ser abordadas mediante el análisis detallado de muestras poblacionales andaluzas bien definidas, provenientes de procesos de muestreo

cuidadosos. Habitualmente, las muestras de Andalucía han tenido una indefinición geográfica dentro de la región (Côrte-Real et al., 1996; Larruga et al., 2001; Plaza et al., 2003) y éstas se han empleado con propósitos comparativos profusamente en la literatura (Pereira et al., 2005b; Coudray et al., 2009; Ottoni et al., 2009; Martínez-Cruz et al., 2012; entre otros).

Andalucía ha destacado por sus elevados niveles de diversidad genética desde los primeros trabajos sobre el ADNmt llevados a cabo a baja resolución filogenética. Por ejemplo, Côrte-Real et al. (1996) mostraron que el valor más elevado del número medio de diferencias por pares de secuencias de todas las poblaciones analizadas (pertenecientes a Iberia, Italia y Argelia) se detectaba en Andalucía; Larruga et al. (2001) revelaron resultados convergentes en cuanto al número de haplotipos diferentes. En este último estudio se muestra también cómo la población andaluza muestra la influencia africana (marcada principalmente por los linajes L y M1) y las trazas ‘neolíticas’ (J, T1, U3) más destacables de Iberia.

Las primeras muestras asociadas a un área concreta dentro de Andalucía se enmarcan en la provincia de Córdoba (Casas et al., 2006). Esta definición geográfica, unida a un apreciable tamaño muestral y al hecho de que se trata de un análisis diacrónico (individuos contemporáneos y provenientes de una necrópolis medieval) nos proporciona abundante información acerca de la diversidad mitocondrial de la región. La población actual analizada muestra una notable contribución de linajes subsaharianos (L), pero una representación menor del marcador norteafricano U6.

4.3. Alcanzando la máxima resolución: el análisis de mitogenomas

En los últimos años, la diversidad de las poblaciones de la Península Ibérica se está abordando desde el estudio de sus genomas mitocondriales completos. Habitualmente, y por motivos técnicos, estos enfoques no se centran en el estudio del *pool* mitocondrial global de una población, sino en aproximaciones ‘basadas en haplogrupos’, es decir, focalizadas en el estudio en alta resolución de determinados linajes de interés, sobre los cuales se pretende conocer su diversificación de manera detallada. Además de inferencias filogeográficas y filogenéticas, estos nuevos enfoques permiten evaluar características del pasado demográfico de las poblaciones humanas.

Como ya ha sido mencionado, el haplogrupo mitocondrial H destaca en Europa por su amplia presencia y diversidad. Por ello, y debido a que este clado –en la mayor parte de los casos– es indistinguible mediante el mero análisis de la región control, numerosos estudios han secuenciado completamente mitogenomas ibéricos pertenecientes al haplogrupo H (Achilli et al., 2004; Álvarez-Iglesias et al., 2009; Behar et al., 2012a). Esta estrategia aplicada a la máxima resolución ha permitido diseccionar lo que *a priori* se consideraba un *cluster* homogéneo y, detectar también patrones específicos de poblaciones y áreas geográficas, dentro de la Península, pero enmarcados en un contexto europeo.

Del mismo modo, el genoma mitocondrial de la población vasca ha sido secuenciado completamente, lo que ha permitido definir linajes ‘autóctonos’ con un origen probable en la región. Ejemplos los encontramos en los sub-clados U8a (González et al., 2006), U5b1f, J1c5c1 y V22 (Cardoso et al., 2013), HV4a1a (Gómez-Carballa et al., 2012) o numerosas ramas dentro del haplogrupo H (H2a5, Álvarez-Iglesias et al., 2009; H1j1, H1t1, H2a5a1, H1av1, H3c2a, y H1e1a1, Behar et al., 2012a).

Sin duda, uno de los focos de interés principal en la aplicación de estos análisis a máxima resolución es la evaluación de la presencia de linajes africanos (L, M1 y U6) en la Península Ibérica. La distribución de esta influencia del continente vecino a nivel peninsular (ver sección 4.1) no es fácilmente reconciliable con interpretaciones reduccionistas de los eventos de flujo génico.

En esta línea, Cerezo et al. (2012), analizando el macro-haplogrupo L, presentaron evidencias de episodios de movimientos transcontinentales que no se restringían únicamente a la invasión musulmana de Iberia –evento histórico ampliamente presentado como único responsable de la herencia africana en la Península– sino que podían explicarse mediante contactos prehistóricos desde hace unos 11.000 años. Del mismo modo, la detección de miembros del sub-clado L3f en población asturiana se ha interpretado bajo la premisa de conexiones temporalmente profundas entre Iberia y el norte de África (Pardiñas et al., 2014).

Por su parte, el linaje norteafricano U6 ha sido también analizado en dos recientes publicaciones (Pereira et al., 2010a; Secher et al., 2014) en las que se perfila la historia de este clado en un contexto, no solo de la Península Ibérica, sino también a nivel global, con el fin de establecer avances en el conocimiento acerca de los procesos de intercambio poblacional entre Europa y África.

5. La diversidad genética del sur de Iberia desde el estudio de otros marcadores

Nuestro grupo de investigación ha venido desarrollando en estos últimos diez años estudios de diversidad genética humana basados en diferentes marcadores clásicos y moleculares (de ADN) en el sur de la Península Ibérica (región de Andalucía), por lo que las mismas poblaciones y muestras son las empleadas en la presente Tesis Doctoral. Esta estrategia nos ha permitido comparar resultados derivados de los linajes maternos con otros sistemas genéticos, con el objetivo de ir obteniendo una visión completa del acervo génico de las poblaciones que habitan el sur de Iberia. Asimismo, gracias a este mosaico de estudios genéticos paralelos, la población andaluza puede ser enmarcada de manera más precisa en un contexto europeo y mediterráneo.

Entre estos antecedentes genéticos de la población andaluza autóctona, cabe destacar la información obtenida sobre el otro sistema haploide humano (además

del ADNmt), esto es, la región no recombinante del cromosoma Y (NRY). El haplogrupo paterno más frecuentemente detectado tanto en el oeste como en el este de Andalucía, es el R-M269, (en ambos casos ~60%) (Ambrosio, 2011; datos sin publicar). Este haplogrupo es el marcador europeo del cromosoma Y por excelencia. En este contexto, el haplogrupo E es el segundo en importancia en la Andalucía occidental (Huelva ~12%). Su composición revela, además, una compleja mezcla de marcadores (linajes) de todo el espacio mediterráneo, con interesantes señales de la Península Balcánica y del Oriente Medio así como la influencia de población bereber. Sin embargo, la contribución bereber (fundamentada por la presencia del linaje E-M81, ~3,5%, ver **Figura I-14**) no parece ser concordante con las regiones (como la andaluza) donde la ocupación islámica fue más intensa y prolongada. Este resultado apoyaría de hecho que los movimientos poblacionales entre el Magreb e Iberia se hubieran iniciado más tempranamente en el tiempo. Otros sub-linajes patrilineales, como el E-V13, señalarían contactos con poblaciones que habrían llegado por mar en períodos protohistóricos, cuando *Tartessos* poseía gran relevancia estratégica (ver **Capítulo II**, sección 1.3), mientras que el E-M34 –detectado también en las muestras poblacionales de varones andaluces analizados– ha sido relacionado con población judía. Todas estas señales apuntan a que las poblaciones de Huelva y Granada han estado sometidas a altas dosis de migraciones (y flujo génico) procedentes del espacio mediterráneo (Ambrosio et al., 2010a).

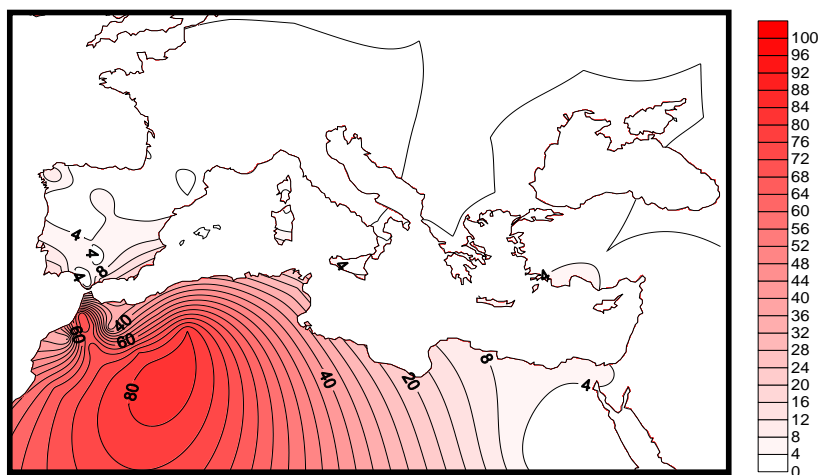
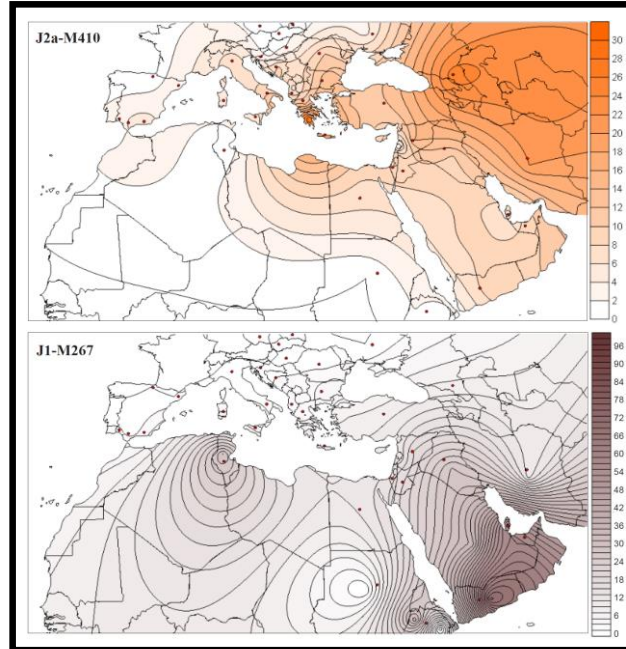


Figura I-14. Mapa de contornos mostrando los patrones de variación espacial del haplogrupo E-M81 del C-Y a través del Mediterráneo.

Otro haplogrupo del C-Y especialmente informativo es el Hg J, con frecuencias ~10% en las poblaciones autóctonas andaluzas de Huelva y Granada. Este linaje registra valores modales en Oriente Medio y en el sudoeste de Asia (ver **Figura I-15**). Sus sub-clados han sido relacionados con la presencia de árabes en la Península Ibérica durante la expansión islámica (J1) y con las colonias griegas y fenicias de Iberia (J2) (Ambrosio et al., 2010b). Por último, la detección del

haplogrupo G en el territorio peninsular (también en Andalucía) puede ser reflejo de la influencia romana (Ambrosio, 2011).

Figura I-15. Mapas de contornos mostrando los patrones genéticos de los haplogrupos J2a-M410 y J1-M267 a través del Mediterráneo y otras áreas geográficas vecinas. Fuente: Ambrosio et al., 2010b.



La caracterización de marcadores microsatélites del C-Y en Andalucía (Y-STRs) ha revelado también una elevada diversidad haplotípica para las poblaciones de Huelva y Granada. La estructura haplotípica detectada muestra una estrecha afinidad genética con otras poblaciones ibéricas. Sin embargo, haplotipos característicos de otras poblaciones y regiones del sur y norte mediterráneo han sido igualmente observados (Ambrosio et al., 2012).

Asimismo, hemos prestado una atención especial al análisis de algunos marcadores autosómicos en el sur de la Península Ibérica. Un ejemplo lo encontramos en el estudio de los alotipos de las inmunoglobulinas humanas (*GM* y *KM*) para Huelva (Calderón et al., 2006) y Granada (Fortes-Lima et al., 2014). El sistema *GM* resulta ser particularmente más informativo al albergar altos niveles de polimorfismo. Ambas poblaciones andaluzas muestran un componente *GM* europeo típico, pero enriquecido por el haplotipo africano *GM*1,17 5** (ver distribución geográfica en la **Figura I-16**).

Por último, hemos abordado la variación molecular de la Apolipoproteína E (APOE) en Andalucía utilizando las mismas poblaciones y muestras (Reales et al., 2014). Las elevadas frecuencias de alguno de los alelos del gen APOE ($\epsilon 4$) en el sur de Iberia (fuera del rango esperado para una población europea de esa latitud) podrían ser explicadas por las múltiples migraciones que ha experimentado este territorio.

Un escenario que se encuentra de manera recurrente en el análisis de la diversidad andaluza, además de su cercanía con poblaciones del norte de África y de regiones distantes del área mediterránea, es una identidad genética entre el oeste de Andalucía (Huelva) y otras regiones occidentales de Iberia, como son Galicia y la vecina Portugal.

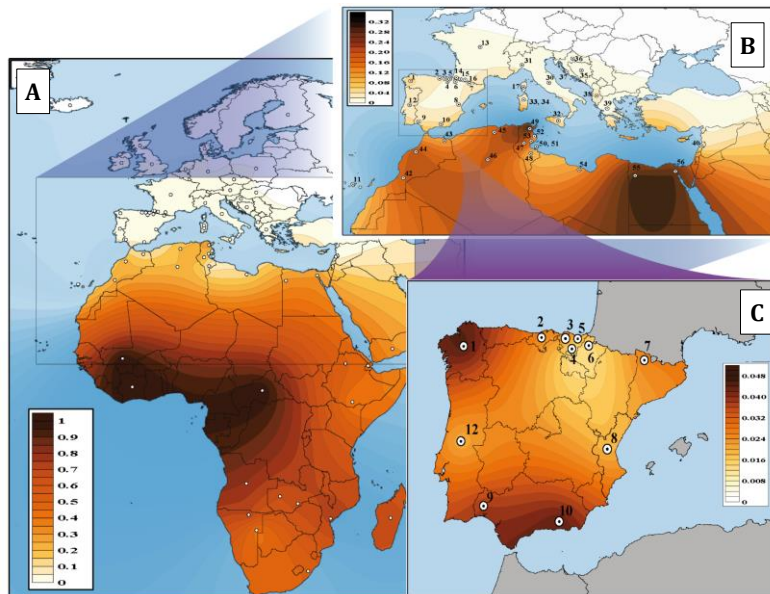


Figura I-16. Variación espacial del haplotipo subsahariano GM 1,17,5* en el continente africano (A), la cuenca mediterránea (B) y la Península Ibérica (C). Fuente: Fortes-Lima et al., 2014.

La relación entre los haplogrupos del cromosoma Y con otros marcadores de herencia paterna (en este caso cultural y no-genética) como son los apellidos ha sido objeto de una reciente publicación de nuestro grupo de investigación (Calderón et al., 2015). Este estudio es pionero en varios aspectos: (1) se ha realizado un análisis comparativo en población general, sin seleccionar apellidos con baja representación, (2) el proceso de muestreo (ver **Capítulo II**, sección 2) llevado a cabo, en los territorios de Huelva y Granada, nos proporciona la información directa de los apellidos de cada individuo estudiado. Este hecho, junto con el elevado tamaño muestral usado, aporta sin duda robustez a nuestros resultados. Los datos presentados en este estudio revelan que la correlación entre apellidos y marcadores del cromosoma Y es baja, independientemente del nivel de resolución alcanzado, cuestionando el carácter monofilético de los apellidos españoles.

6. Los objetivos del presente trabajo

El objetivo central de la presente Tesis Doctoral es el análisis de la diversidad matrilineal en el sur de la Península Ibérica con el fin de desvelar acontecimientos relacionados con el poblamiento del espacio mediterráneo y otros procesos evolutivos de la historia humana reciente. Este objetivo central se ha subestructurado en los siguientes objetivos concretos:

1. Caracterizar la naturaleza y magnitud de la diversidad mitocondrial global de dos poblaciones de Andalucía, situadas en los extremos geográficos de la región, el occidental representado por la provincia de Huelva y el oriental por la provincia de Granada. La información genética obtenida ha permitido evaluar escenarios de continuidad o discontinuidad genética.
2. Analizar las características que mejor definen la herencia materna de Andalucía dentro del marco de la Península Ibérica, fundamentalmente respecto a los patrones detectados en la franja cantábrica.
3. Valorar el papel del sur de Iberia como receptor de genes de diversas áreas continentales (especialmente del continente africano) y examinar el papel de las rutas marítimas –a pequeña y gran escala– en las migraciones humanas y la permeabilidad de las barreras geográficas al flujo de genes.
4. Evaluar las trazas genéticas africanas en la Península Ibérica mediante un análisis en alta resolución filogenética de linajes mitocondriales específicos.
5. Diseccionar el clado mitocondrial europeo (clado H) muy frecuente también en Andalucía y analizar detalladamente sus principales radiaciones.
6. Estimar tiempos de coalescencia de linajes mitocondriales y analizar el peso de los períodos prehistóricos e históricos en la configuración del *pool* génico de las poblaciones ibéricas contemporáneas.
7. Aplicar metodologías experimentales, adaptadas específicamente al estudio de cada fracción de la diversidad materna en la población andaluza.
8. Sobre la base de los resultados obtenidos, aportar inferencias sobre dinámicas de población y eventos demográficos en el contexto de la Península Ibérica y el espacio mediterráneo con especial referencia a su extremo más occidental.

Capítulo II

Material y Métodos

1. La población de estudio

1.1. Localización geográfica

Andalucía se sitúa en el extremo meridional de España, con una superficie (87.598 km²) que representa el 17,3% del total del territorio nacional, y es la única región española que tiene sus costas repartidas entre dos mares: el Mediterráneo y el Atlántico. Su proyección mediterránea ha favorecido la integración y la continua interacción con las culturas que han sido la cuna de la civilización occidental; su vertiente atlántica, por otro lado, le garantizó un papel relevante en el descubrimiento y colonización del Nuevo Mundo. Andalucía es también un puente natural entre Europa y África, gracias a los escasos 14 km que separan ambos continentes en el Estrecho de Gibraltar. Esta estratégica posición geográfica (ver **Figura II-1**) y su significativo papel marítimo a lo largo de su historia, justifica que su población autóctona esté siendo analizada con especial interés desde un punto de vista de la Genética de las Poblaciones Humanas.



Figura II-1. Localización geográfica de Andalucía y de las provincias de Huelva y Granada.

La palabra “*Andalucía*” aparece por vez primera aplicada a un espacio político-administrativo durante el período cristiano-musulmán. Hasta 1833, cuando se creó la división provincial española, y que prevalece actualmente, la definición del territorio andaluz era algo imprecisa (López Ontiveros, 2003a). Los documentos geográficos y administrativos distinguían entre Andalucía y Granada, reminiscencia del *Reino Nazarí de Granada*, último bastión musulmán peninsular. Esa base histórica junto con algunas características de su geografía física, su demografía y su economía, han fundamentado que Andalucía se haya dividido tradicionalmente en dos grandes sub-regiones: Andalucía Occidental (identificada con la *Baja Andalucía* y compuesta por las provincias de Huelva,

Cádiz, Sevilla y Córdoba) y Andalucía Oriental (*Alta Andalucía*, Málaga, Granada, Jaén y Almería) (López Ontiveros, 2003b).

El relieve geográfico (orografía) de Andalucía está constituido por una amplia variedad de paisajes geomorfológicos. Existen tres grandes unidades morfo-estructurales dentro del territorio: *Sierra Morena* (que configura la región septentrional), la *Depresión del Guadalquivir* (como un triángulo abierto al Atlántico) y entre esta llanura y el Mediterráneo, se encuentran los *Sistemas Béticos* (conjunto de alineaciones de sierras que desde el Estrecho de Gibraltar se extienden hasta el Cabo de la Nao, en Alicante) (Ferre, 2000; Moreira & Ojeda, 2003). En la región de Andalucía existen dos vertientes hidrográficas, la atlántica y la mediterránea, que son desiguales en la superficie que drenan y en la longitud de sus litorales respectivos. La cuenca hidrográfica que destaca por su extensión y relevancia es la del río Guadalquivir (Valle, 2000).

Concretamente, en nuestras investigaciones sobre diversidad genética humana nos hemos centrado en dos territorios distantes y bien diferenciados dentro de Andalucía. En primer lugar, la provincia de Huelva, situada en el extremo más occidental de la región y bordeando la vecina Portugal; en segundo lugar, la provincia de Granada, localizada en la vertiente oriental (ver **Figura II-1**). Ambas provincias se encuentran abiertas al mar y han estado históricamente diferenciadas.

La provincia de Granada (12.647 km²) es más extensa que la de Huelva (10.128 km²) (Instituto Nacional de Estadística, INE, <http://www.ine.es/>). Ambas provincias se encuentran separadas ~300 km y presentan también rasgos distintivos en cuanto al relieve. La influencia de las cadenas montañosas (Sistemas Béticos) es superior en Granada mientras que el territorio de Huelva está determinado, en su mayor parte, por el valle del Guadalquivir. También existen diferencias en cuanto a la explotación del territorio, destacando Huelva por su notable actividad minera desde tiempos prehistóricos hasta la actualidad.

1.2. Características demográficas

La población de la Península Ibérica experimentó un continuado crecimiento demográfico desde el asentamiento romano, alcanzando su punto culminante hacia el siglo II a. C. Estimaciones posteriores apuntan a que España tuviese una población de 6 millones a comienzos del siglo primero d.C., un valor notable respecto a los 23 millones de personas que habitaban el Imperio Romano en su área europea (Corvisier et al., 1997). La provincia *Bética* (el nombre romano para Andalucía, basado en el antiguo nombre del río Guadalquivir) fue una de las regiones más pobladas y ricas del Imperio. Dentro de la misma, destacaba un importante núcleo de densidad poblacional: *la Vega de Granada* (Plácido, 1989). En el siglo VIII, el tamaño demográfico de España se estima entre 6-7 millones de habitantes (Guichard, 1976; Dupâquier, 1997), con una importante contribución bereber debido a la conquista musulmana de Iberia. En ese siglo se calcula que

llegaron alrededor de un tercio de los 300.000 bereberes que fueron introducidos en el territorio peninsular (Mackay, 1977). Este período también trajo a Iberia esclavos (~100.000) de procedencia franco-eslava mayoritariamente. La población del *Reino Nazarí de Granada* durante su época de mayor apogeo (siglo XV) fue alrededor de unos 300.000 habitantes, concentrados preferentemente en determinadas áreas como la Alpujarra granadina, una zona poblada por unas 40.000 personas durante esa época. A mediados del siglo XV, la ciudad de Granada tenía una población de unos 50.000 habitantes siendo entonces una de las ciudades más pobladas de la Península (Ladero-Quesada, 1993).

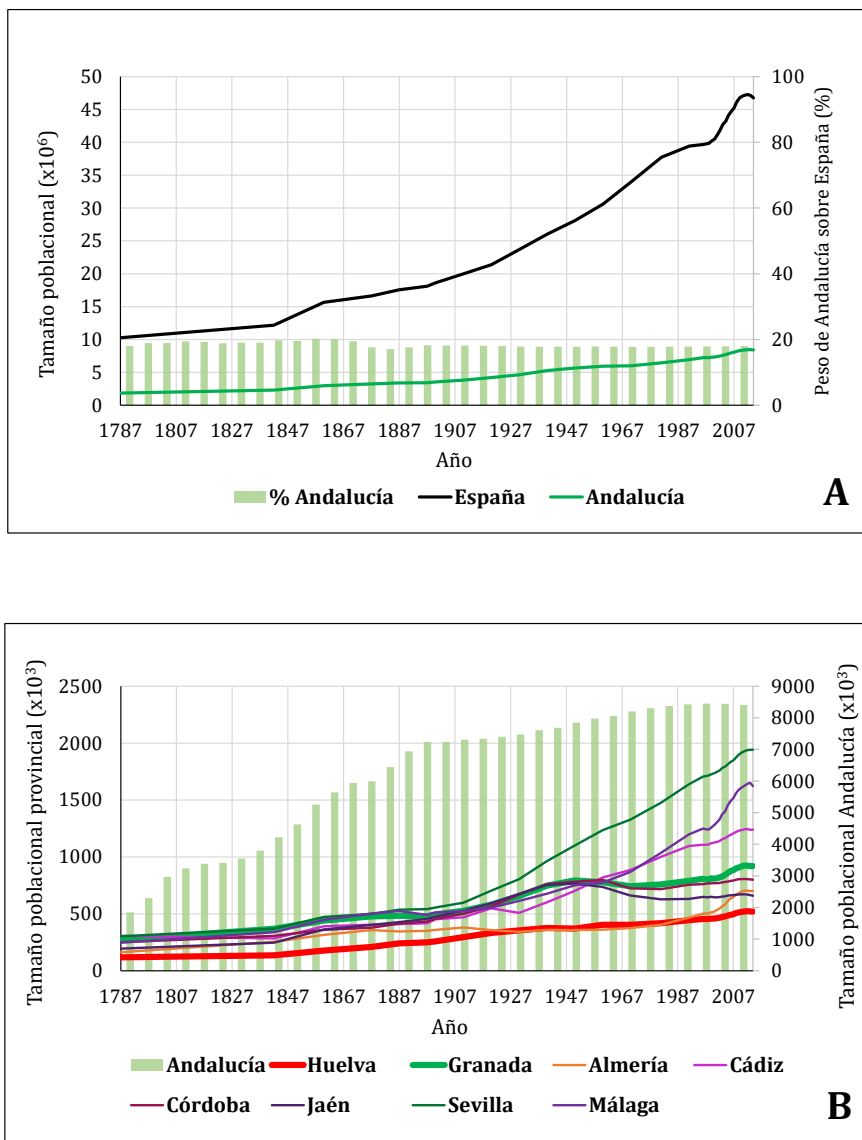
A finales del siglo XV, el territorio andaluz concentraba alrededor de 1.200.000 personas (Pérez & Reher, 1988). Tras la Reconquista (a partir de 1492) entraron en el antiguo Reino de Granada de 35.000 a 40.000 repobladores procedentes de otras partes de la Península, principalmente de la Submeseta Norte (Ladero Quesada, 1988).

La población española experimentó a partir del siglo XV un incremento constante que compensó el declive sufrido a mediados del anterior siglo. El crecimiento poblacional entre los siglos XVI y XVII siguió un patrón similar al del resto de la Europa occidental, alcanzándose una cifra de en torno a los 7 millones de habitantes, con oscilaciones continuadas debido a crisis económicas, epidemias y guerras (Pérez & Rowland, 1997). A partir del siglo XVIII, comenzó a cambiar el patrón de evolución demográfica en España, registrándose un crecimiento en las cifras poblacionales que se frenó ligeramente durante el siglo XIX (Pérez, 1988). El incremento demográfico más acusado sucedió durante el siglo XX, pasando de un tamaño de 19 millones de personas a superar los 40 millones (Cabré et al., 2002) (ver **Figura II-2A**). La población actual de España es hoy alrededor de 47 millones (INE, 2014).

La aparición de los primeros Registros Civiles en España nos permite estudiar más detalladamente la evolución del tamaño demográfico tanto en el marco nacional como regional. La **Figura II-2A** nos muestra el peso demográfico de Andalucía, como un todo, respecto al total de España. Podemos observar cómo esta contribución se ha mantenido por debajo del 20% del total nacional desde el primer censo moderno de población– con registro por personas y no por hogares (*Censo de Floridablanca*, 1787). En la actualidad (datos de 2014), la población andaluza constituye ~18% del total de España (Instituto de Estadística y Cartografía de Andalucía, IEA, <http://www.juntadeandalucia.es/institutodeestadisticaycartografia/>).

La abundante población de Andalucía, sumada a la gran superficie que representa con respecto al territorio nacional, ha provocado que se considere esta área como “síntesis geográfica de la Península Ibérica” (López Ontiveros, 2003a).

Figura II-2. Evolución diacrónica (1787-2014) del tamaño demográfico de Andalucía con respecto a España (A) y de las 8 provincias que componen esta Comunidad Autónoma (B). Datos del INE y del IEA (consultados en abril, 2015).



En la **Figura II-2B** se observa que la provincia de Sevilla es, al menos desde el siglo XVIII, el territorio que más contribuye al tamaño demográfico de Andalucía. La provincia de Huelva ha registrado durante la mayoría de las series históricas, e incluso en la actualidad, el menor tamaño poblacional de la región. Además, el gráfico nos muestra cómo Huelva y Granada han mantenido una tendencia uniforme durante los años analizados, sin incrementos abruptos como los que han experimentado Sevilla, Málaga y Cádiz. En definitiva, la provincia de Granada posee un mayor tamaño demográfico, y una densidad de población más alta que la de Huelva, respectivamente: 72,98 hab/km² y 51,63 hab/km².

Los patrones de población dentro de Andalucía a lo largo del tiempo se observan espacialmente en la **Figura II-3**. El incremento del poblamiento de las áreas costeras es evidente.

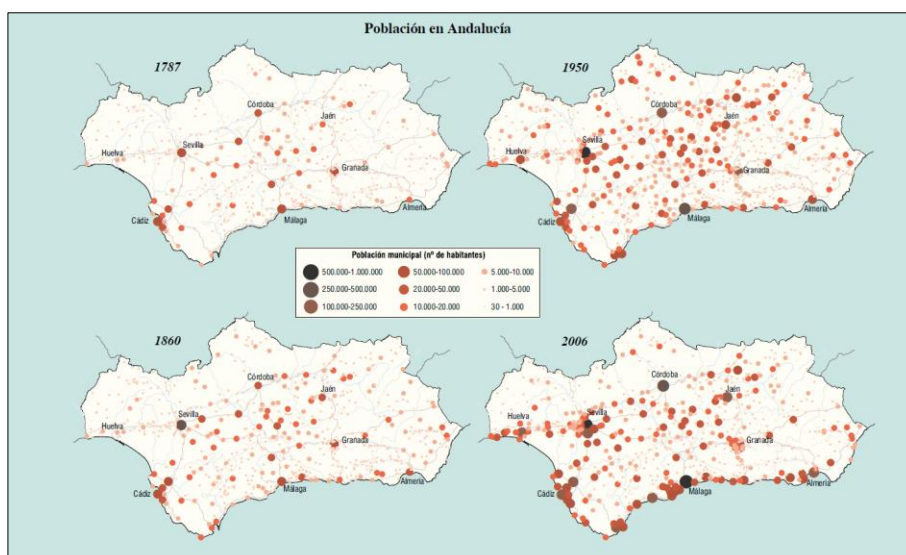


Figura II-3. Evolución de la población en Andalucía y patrones de poblamiento desde finales del siglo XVIII. Fuente: *Atlas de la Historia del Territorio de Andalucía* (IEA, 2008).

En la **Figura II-4** podemos observar que el gran eje de organización interna de Andalucía es el valle del Guadalquivir. El río ha sido navegable en parte hasta la época moderna, cuando los navíos procedentes de América llegaban hasta Sevilla, siendo el valle una gran llanura abierta al mar. La amplia franja geográfica que representa el litoral andaluz ha representado históricamente un elemento de gran importancia en el transporte a corta y larga distancia. En este contexto, cabe señalar que la comunicación con el resto de la Península se ve favorecida en el oeste por el escaso relieve de Sierra Morena, mientras que existe más dificultad en el Sistema Bético (este de Andalucía).

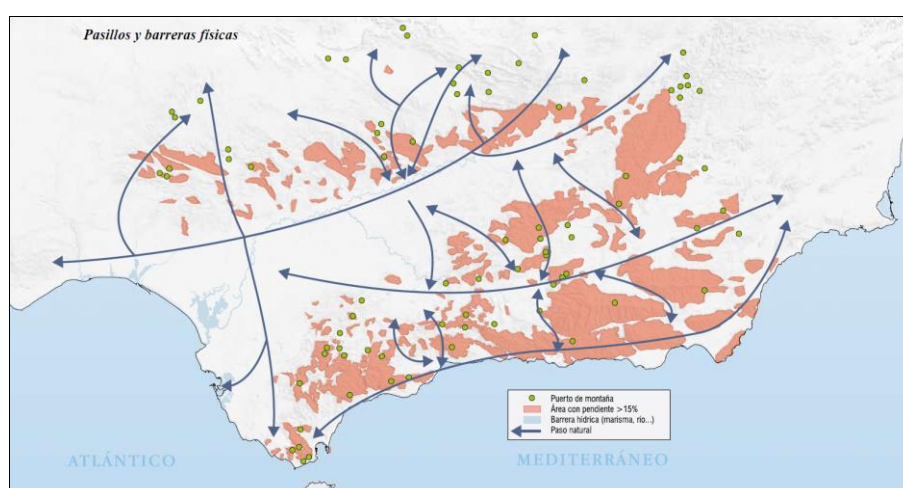


Figura II-4. Corredores y barreras físicas determinadas por la orografía en Andalucía. Fuente: *Atlas de la Historia del Territorio de Andalucía* (IEA, 2008).

1.3. *Apuntes históricos*

La Península Ibérica se ha considerado como un “*continente en miniatura*”, debido a su estructura geográfica, su variado clima, su vegetación y su población (Demangeon, 1907 en Domínguez, 2010). Además de esta complejidad orográfica interna, su situación geográfica y sus recursos naturales han propiciado que desde tiempos muy antiguos haya sido un foco de atracción para la llegada de poblaciones de muy variadas culturas desde Europa (norte), el continente africano (sur) y, con especial intensidad y relevancia, a través del Mediterráneo en general, el cual permitió conectar Iberia con territorios y poblaciones de su parte oriental y del Oriente Próximo, especialmente. Con estos escenarios, la Península ha representado a lo largo de su protohistoria e historia, un polo gravitacional de culturas, entre las que destacan fenicios, griegos, celtas, visigodos, romanos, árabes y bereberes. Las áreas geográficas más destacadas para estas migraciones han sido Levante y Andalucía, territorios abiertos al Mediterráneo, centro del mundo occidental hasta la Edad Moderna.

Andalucía posee un registro rico y bien preservado de su Prehistoria. El sudeste de Andalucía ha sido y está siendo profusamente estudiado (Gilman et al., 1985; Chapman, 1989). La Cueva de Nerja (Málaga) destaca por ser uno de los pocos yacimientos de la España meridional que cuenta con un buen registro de industria del Paleolítico Superior. Además, esa secuencia arqueológica muestra una ocupación humana hasta la Edad del Bronce (Villaverde et al., 1998). Es interesante destacar que el Magdaleniense mediterráneo presenta un proceso de marcada regionalización, y distintivo del observado en el área cantábrica (Villaverde et al., 2012). Restos de culturas Epipaleolíticas en el este andaluz se han encontrado en la Cueva de La Palica (Almería) y en Hoyo de la Mina (Málaga) (Gilman et al., 1985). La distribución de yacimientos pre-Neolíticos es desigual en el sur peninsular, con cuatro focos de concentración: las costas del río Guadalentín; la costa de Mazarrón-Cartagena (Murcia) y la cuenca de Vera (Almería); el área Iznalloz-Piñar (noroeste de Granada); la zona costera de la Bahía de Málaga y los ríos del noroeste de Algeciras y Tarifa (Cádiz) (Chapman, 1989).

El Neolítico en la Península Ibérica está documentado al menos desde ~5600-5500 a.C. (Zapata et al., 2004; Chapman, 2008). El Neolítico ibérico se considera de gran interés por su posición entre las tradiciones atlántica, africana y mediterránea. En un ámbito general, los estudios acerca del origen del Neolítico gravitan principalmente sobre dos pilares: (1) resolver el debate entre continuidad o discontinuidad poblacional y (2) evaluar las tasas de dispersión y la homogeneidad o heterogeneidad de los registros (Cruz Berrocal, 2012).

Las primeras poblaciones Neolíticas en el sur de Iberia se establecieron preferentemente en enclaves previamente utilizados por poblaciones Mesolíticas, las cuales dependían de un amplio rango de recursos costeros. Las muestras más tempranas en esta región se encuentran en las costas de Málaga y en el Algarve portugués (Cortés Sánchez et al., 2012). En Andalucía, restos de cultivos Neolíticos tempranos se concentran en dos cuevas de Córdoba (Los Murciélagos) y Málaga (Toro y Nerja), con dataciones ca. 5600-5400 a.C. La presencia de una

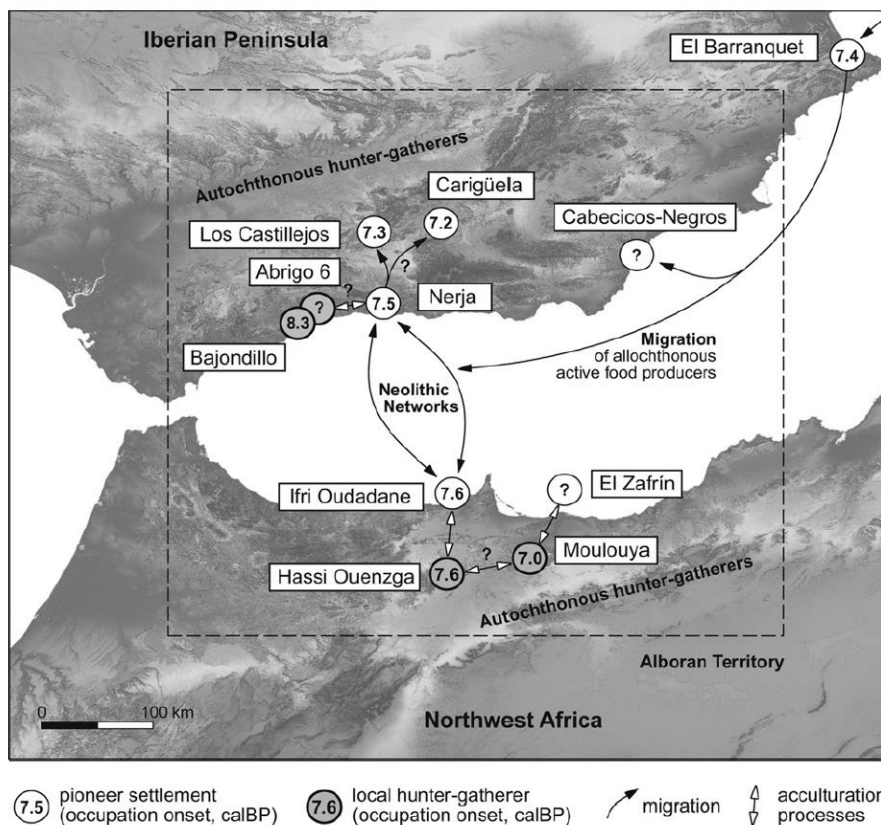
gran diversidad de cultivos en el Neolítico ibérico con diferentes requerimientos, usos y procesamientos, implica que los primeros agricultores importaron o adquirieron rápidamente un amplio rango de conocimiento agrario (Zapata et al., 2004; Cruz Berrocal, 2012).

Varias evidencias apoyan procesos de rápida dispersión del Neolítico en el Mediterráneo occidental y en el sur de Iberia. Zilhão (2001) propuso el ‘modelo de colonización marítima’ que explica la llegada de la agricultura a este territorio. Mediante este modelo, pequeños grupos llegaron desde el Mediterráneo y fueron navegando a lo largo de la costa, asentándose en áreas concretas que servirían como centros de neolitización. Por tanto, la rápida expansión del Neolítico en el Mediterráneo occidental fue facilitada por el uso de tecnología marina y de comercio de materiales basado en focos mediterráneos bien establecidos en el centro y este de la cuenca (Broodbank, 2006). Sin embargo, algunos autores han señalado que el registro heterogéneo observado en Iberia (gran variabilidad regional en manifestaciones cerámicas, sin una clara distribución espacial) contrastaría con lo esperado según un modelo de colonización rápida a gran distancia (Cruz Berrocal, 2012). Estos datos nos revelan que la neolitización de la Península Ibérica no fue un proceso uniforme, sino diverso y complejo que parece reconciliar tanto las explicaciones indigenistas (transmisión de ideas sin intercambio poblacional) como las colonialistas (Bernabeu, 1996).

Parece aceptarse que la mayor parte del paquete Neolítico fue introducido en el sur de Iberia desde el este, en particular, desde la región del golfo de Génova, (Arena Candida). Sin embargo, algunos componentes del Neolítico occidental ibérico no pueden ser explicados exclusivamente por esta influencia oriental. Se han observado paralelismos en los procesos de neolitización en el sur de Iberia y en el Magreb –por ejemplo, similitudes en cerámicas, industrias líticas y en el uso de las mismas especies de plantas y animales domesticados. Estos hallazgos se interpretan como evidencias de las relaciones estratégicas entre orillas opuestas del Mediterráneo occidental (Manen et al., 2007; Linstädter et al., 2012). La región de Alborán es el contexto geográfico fundamental donde esas conexiones podrían haberse registrado, con la pequeña isla del mismo nombre como una etapa intermedia ocasional (ver **Figura II-5**).

Por otro lado, el sudeste ibérico, en concreto, la región de Almería-Murcia, destaca en cuanto a sus sucesivas industrias de la Edad de los Metales, con personalidad propia y numerosos asentamientos. Durante la Edad del Cobre, la cultura de *los Millares* (3200-2250 a.C.) muestra un perfeccionamiento de las técnicas agrícolas y ganaderas, el trabajo del cobre y cerámicas, la construcción de un complejo sistema de murallas y necrópolis megalíticas. Posteriormente, en la misma área geográfica se desarrolla la cultura de la Edad de Bronce temprana de *El Argar* (2250-1550 a.C.). Esta industria se diferencia de la precedente en el uso del metal, organización social y diferentes estructuras urbanísticas (Carrilero, 2000; Chapman, 2008).

Figura II-5. Modelos de neolitización en el área de Alborán. Los asentamientos pioneros costeros de ambas orillas entraron en contacto con sociedades cazadoras-recolectoras locales. Las dataciones se indican en miles de años BP. Fuente: Linstädter et al., 2012.



El período que transcurre entre los años 1000 y 500 a.C. es clave para el mundo mediterráneo occidental, ya que representa la transición entre la prehistoria y los albores de la romanización. Este proceso se desarrolló como consecuencia de la actividad marítima incesante de fenicios, griegos y cartagineses a través de navegaciones hacia Iberia (preferentemente a Andalucía y Levante) lo que conllevó procesos de colonización de nuevos territorios e implantación de un poblamiento propio (ver **Figura II-6**). Los fenicios y cartagineses fundaron importantes ciudades andaluzas como Cádiz (*Gádir*), Málaga (*Málaga*), Almuñécar (*Sexi*) o Adra (*Abdera*) (Fernández, 2003). Estos asentamientos son producto de las iniciativas promovidas por Tiro para canalizar el comercio de metales desde el Mediterráneo occidental hasta oriente (Carrilero, 2000). Existe unanimidad en otorgar a los metales el papel de elemento básico de atracción (Tuñón de Lara et al., 1980) de ahí que, el mar haya desempeñado un papel creciente como medio de comunicación entre poblaciones humanas distantes (Delgado, 2013).

La llegada de los fenicios-cartagineses y griegos al Mediterráneo occidental provocó un impacto considerable en procesos de aculturación locales. Como resultado de ello, ciertos grupos desarrollaron culturas específicas gracias a los estímulos recibidos de esas avanzadas civilizaciones. Por ejemplo, el surgimiento de la *Cultura Tartésica* estaría relacionada con este fenómeno (Almagro-Basch et al., 1974; Acquaro, 1999). *Tartessos* (también referido como *Tarshish* o *Tarschisch*

para las fuentes bíblicas) fue el nombre que fuentes griegas dieron a una zona del extremo mediterráneo occidental, concretamente al tramo final del valle del Guadalquivir y territorios circundantes (Carrilero, 2000) (ver **Figura II-7**). Esta cultura se considera un eje esencial en la protohistoria ibérica, el foco a través del cual la Península Ibérica entró de manera definitiva en la historia mediterránea (Acquaro, 1999). La importancia de esta civilización la resumió el arqueólogo alemán Adolf Schulten: “*Tartessos fue la primera ciudad comercial y el más antiguo centro cultural de Occidente*” (ver Schulten, 2006).



Figura II-6. Rutas de comercio fenicias y griegas a través del Mediterráneo. Fuente: *Atlas de la Historia del Territorio de Andalucía* (IEA, 2008).

Los abundantes restos de la Cultura Tartésica encontrados en el área de Huelva –correspondientes al período posterior al siglo VIII a.C.– se encuentran marcados por intensos contactos con civilizaciones del este del Mediterráneo (Almagro-Basch et al., 1974; Fernandez-Jurado et al., 1997). El impacto de esos procesos culturales del Mediterráneo oriental en Huelva tiene su impacto más relevante en la gran necrópolis Tartésica, *La Joya*, que data desde la segunda mitad del siglo VIII a 550 a.C. (Torres-Ortiz, 1999).

La importancia de *Tartessos* está íntimamente ligada a la minería de estaño, plata, cobre y oro y al comercio a gran distancia asociado con estos metales (Calderón et al., 2006). *Tartessos* floreció desde el primer milenio a. C. hasta su destrucción, hacia el año 500 a. C., con lo que Cartago se adueñó del Mediterráneo occidental.

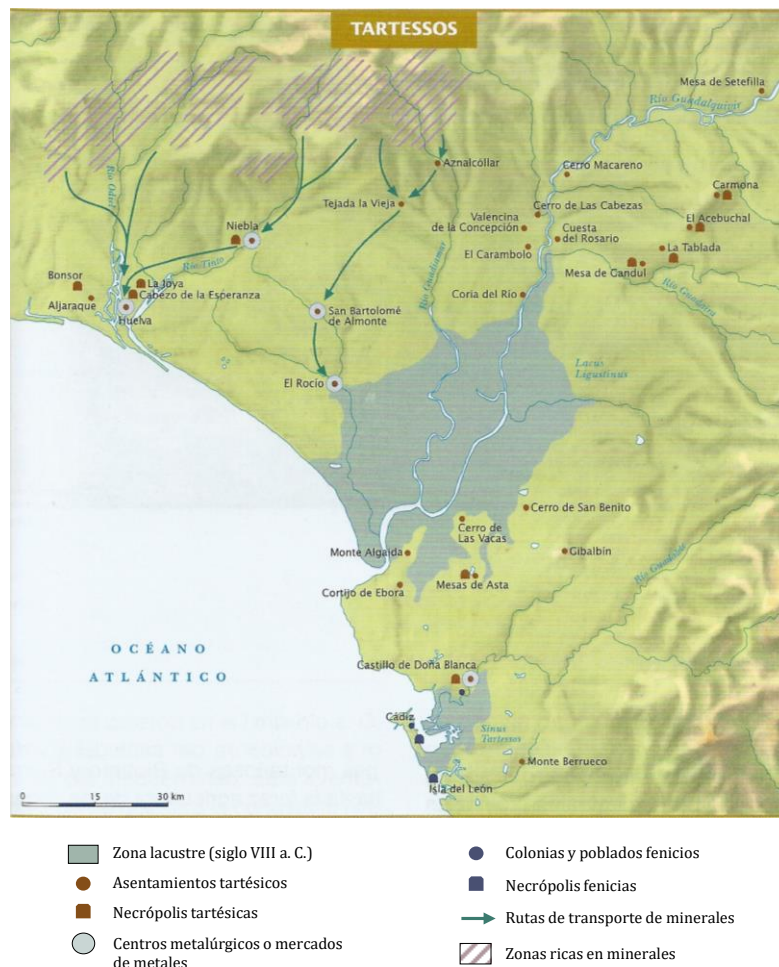


Figura II-7. Zona de influencia de la Cultura Tartésica en el sudoeste de España. Fuente: García de Cortázar, 2012.

El siglo tercero a.C. representa otro punto de inflexión en la historia de Iberia ya que representa los comienzos del dominio romano en el Mediterráneo occidental tras la segunda guerra púnica (218-204 a.C.) (Tuñón de Lara et al., 1980). La conquista y anexión al Imperio supuso un nuevo marco geográfico, dividiéndose la Península en dos unidades, *Hispania Ulterior* y *Citerior*. El sur peninsular se integró en la primera. Posteriormente, *H. Ulterior* fue dividida en la *Bética*, que abarcó gran parte de Andalucía (Martínez & Marín, 2000) (ver **Figura II-8**). Las implicaciones de este nuevo régimen afectaron a muchos campos (económicos, sociales, políticos, administrativos e ideológicos). La romanización provocó el desarrollo de los núcleos urbanos derivados de antiguos asentamientos y la intensificación de las explotaciones anteriores. Además, las exigencias militares y comerciales condujeron a la construcción de numerosos puertos marítimos y fluviales, así como una importante red de caminos de comunicación terrestre, las *'calzadas'*. Entre esas vías terrestres está la llamada *vía de la Plata*, que une la región occidental de España de norte a sur (León-Sevilla). Entre *Hispania* y otras provincias del mundo romano, el comercio de mayor volumen fue el *'comercio exterior'*, que se llevaba a cabo casi exclusivamente por vía marítima. Las exportaciones e importaciones no solo

fueron efectuadas entre Hispania y el mundo Mediterráneo sino también con Britania y las provincias del Rin y del Danubio mediante complejas redes comerciales (Tuñón de Lara et al., 1980).

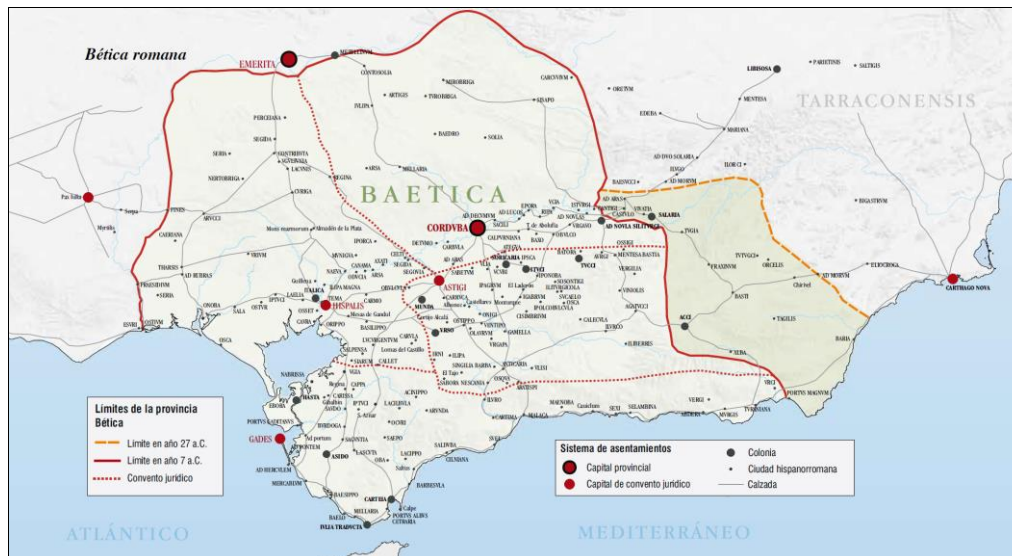


Figura II-8. Provincia Bética romana. Fuente: *Atlas de la Historia del Territorio de Andalucía* (IEA, 2008).

Tras la caída del Imperio Romano y las posteriores invasiones germanas en la Península (Sayas & García, 1982), podemos destacar uno de los episodios más importantes para la historia de Andalucía, y por ende, de Iberia: la **dominación musulmana** (711-1492). La primera mención del nombre *al-Andalus* (que etimológicamente se refiere a los invasores vándalos que atravesaron la Península y se asentaron en el norte de África) pertenece a una moneda acuñada en el año 716, probablemente en tierras andaluzas. Este término hace referencia al territorio peninsular bajo administración musulmana durante toda la Edad Media, con una geografía cambiante desde la conquista en el siglo VIII hasta la constitución del *Reino Nazarí de Granada* en el XV (Valencia, 2000) (**Figura II-9**).

El proceso de islamización introdujo notables cambios en la cultura, política y sociedad de Iberia (Chalmeta, 1989). Se debe tener en cuenta que la expansión musulmana tuvo en sus orígenes un carácter predominantemente militar. El primer contingente militar (año 711) se encontraba compuesto por árabes y bereberes. Posteriormente (año 741), un nuevo ejército se introdujo en la Península, formado en este caso por contingentes de regiones islamizadas del Oriente Próximo y Egipto. Estos componentes poblacionales se asentaron diferencialmente en Iberia. El contingente de Damasco se estableció en Granada, el de Jordán en la provincia de Málaga y el proveniente de Egipto se instaló en Beja (Portugal), cerca de Huelva. Por otro lado, los árabes tendieron a asentarse en las ciudades, mientras que los bereberes habitaban preferentemente regiones montañosas, alcanzando el centro y el noroeste de la Península –Galicia, norte de

Portugal y León (Ajbar Machmuâ, 1867). Otros dos ejércitos bereberes entraron posteriormente en Iberia: los almorávides (1090-1147) y los almohades (1146-1229).

Entre la población que llegó a la Península con las sucesivas olas de inmigrantes, el elemento preponderante parece haber sido el bereber (se ha estimado que representase el 1,5-2% de la población peninsular); el componente árabe, sin embargo, no fue numéricamente tan importante (~0,5%) (ver detalles en Ambrosio et al., 2010b). La contribución de otros componentes –subsaharianos y *esclavones*– fue menor, aunque supuso el tercer elemento extranjero en la sociedad hispano-musulmana (Arié, 1983). La población negra tiene en parte su origen en las redes de comercio trans-sahariano (que comenzaron alrededor del siglo V) y que, además de esclavos, transportaban sal y oro. Estas caravanas permitieron la mezcla poblacional entre el África mediterránea y *Bilad-el-Sudan* (África subsahariana) (Braudel, 1966; von Grunebaum, 1984). Por otro lado, los *esclavones* eran cautivos hechos en la Europa continental –desde Germania hasta las tierras eslavas– que eran vendidos en el mundo musulmán y muchos formaron parte de la guardia del Califa de Córdoba. Aunque se piensa que no se mezclaron mucho biológicamente, a la caída del Califato se agruparon en la parte oriental de *al-Andalus* y en la taifa de Valencia y acabaron integrándose con otros componentes de la población. La base primordial de la población hispanomusulmana fue el antiguo componente de origen hispanorromano y godo, quienes convivieron con los conquistadores (Arié, 1983).

El *Reino Nazarí de Granada* (1013-1492) fue la última de las ‘*taifas andalusíes*’ surgidas de la descomposición del Califato de Córdoba y prolongó el dominio musulmán allí, más de dos siglos después que en el resto de la Península (ver **Figura II-9**). Desde un punto de vista territorial, abarcaba bastante aproximadamente las actuales provincias de Málaga, Granada y Almería, con una extensión geográfica de unos 30.000 km². En el siglo XIV, la ciudad de Granada era considerada una de las ciudades más pobladas de Europa (Ladero Quesada, 1988). Entre los siglos XIV y XV tuvieron lugar importantes migraciones de mudéjares (musulmanes que vivían en territorios bajo gobierno cristiano) procedentes de tierras de la Corona de Aragón buscando refugio por la presión cristiana (Arié, 1983; de Santiago, 2000).

La Reconquista cristiana provocó complejos procesos de colonización y asimilación poblacional. Andalucía fue repoblada por ‘cristianos viejos’ de Castilla, Aragón y otras áreas más distantes (Galicia o Portugal). Tras la conquista del Reino de Granada (1492), numerosos *moriscos* permanecieron en la región hasta que en 1609 fueron expulsados de la Península. El destino de muchos de estos individuos expulsados fue el Magreb (Caro Baroja, 1991).



Figura II-9. Fases de la Reconquista cristiana.
Fuente: *Atlas de la Historia del Territorio de Andalucía* (IEA, 2008).

Los datos presentados en esta sección revelan que Andalucía ha experimentado numerosos movimientos poblacionales desde el Mediterráneo oriental, el norte de África y también del resto de Europa. Los procesos prehistóricos, neolíticos, protohistóricos e históricos de flujo génico han jugado un papel decisivo en la evolución de las poblaciones humanas modernas en Iberia y, por extensión, en la región mediterránea. El impacto genético de estos movimientos migratorios se encuentra fuertemente afectado por factores demográficos y de dinámica poblacional, teniendo en cuenta el considerable tamaño poblacional de la Península Ibérica con respecto a otros territorios continentales vecinos.

2. El proceso de muestreo poblacional

La relevancia y las implicaciones de un adecuado proceso de muestreo en los estudios de Genética de Poblaciones Humanas son indiscutibles. Un correcto diseño del mismo, teniendo en cuenta factores clave como el tamaño muestral, el número de localidades de muestreo y sus características histórico-demográficas, permite fortalecer y sustentar los resultados que se obtienen y las inferencias que se postulan en este tipo de estudios.

El informe *Evaluating Human Genetic Diversity*, elaborado por el *National Research Council (US) Committee on Human Genome Diversity* en 1997 expone estos puntos con claridad y propone, además, diversas estrategias de muestreo. En nuestro caso, hemos llevado a cabo la estrategia III (ver más abajo) gracias a la cual no se registra simplemente el origen geográfico de las personas que componen la muestra, sino también el sexo, la edad y los lugares de nacimiento de los padres. Siguiendo esta estrategia, la muestra obtenida permite: (1) probar

hipótesis relacionadas con patrones de variación genómicos y espaciales, (2) ampliar el marco para el análisis de hipótesis sobre patrones de migración y flujo génico y, (3) evaluar subdivisiones poblacionales y, más ampliamente, contrastar inferencias con otras fuentes de información complementarias pero de alto interés en la interpretación genética de los datos, como la arqueología, la historia o la lingüística. Ese conocimiento integrado puede permitirnos interpretar de una forma más refinada patrones de variación interpoblacional.

La estrategia seguida tiene el valor añadido de ofrecer el balance más adecuado entre la evaluación de hipótesis, los costes del proceso y las implicaciones éticas. De manera literal, se indica lo siguiente:

“A coordinated global sampling effort to develop a common resource for research on human genome variation should use a population-based sampling design in which the geographic location of the sample and self-reported ethnicity, primary language, sex, age, and parental birthplaces are recorded”.

En nuestras investigaciones sobre diversidad genética humana en el sur de la Península Ibérica (Andalucía), la elección de las localidades de muestreo se basó en las características demográficas e históricas de los municipios de Huelva y Granada. La información acerca de la evolución de la población (tamaños demográficos, N) en los municipios de estas dos provincias andaluzas fue tomada del Instituto Nacional de Estadística (INE, www.ine.es). El tamaño poblacional ($\approx 3.000-5.000$ habitantes), la estabilidad demográfica y aspectos relevantes de su pasado histórico fueron los criterios en los que se fundamentó la selección de las localidades para el muestreo. Las áreas costeras y metropolitanas de ambas provincias fueron evitadas por contar con mayores tasas de inmigración y de mezcla, y por tanto, menor probabilidad de autoctonía. El muestreo fue una tarea delicada, prolongada en el tiempo y que representó costes importantes. La colección de las muestras se efectuó entre los años 2004 y 2008 (6 campañas en Huelva y 6 en Granada). Este trabajo de campo fue llevado a cabo personalmente por la Prof. R. Calderón, como responsable de nuestro grupo de investigación, acompañada siempre por personal sanitario (médicos y enfermeras) que formaban parte de las unidades móviles de donación de sangre, gestionadas por el Servicio Andaluz de Salud (SAS). Esta planificación personal y detallada permite tener un control total sobre el origen de las muestras.

Las localidades de muestreo, repartidas por los territorios de Huelva y Granada, se muestran en la **Figura II-10**. En la provincia de Huelva, se recogieron muestras en El Repilado (n=29), Aracena (n=95), El Cerro del Andévalo (n=33), la Puebla del Guzmán (n=102), Valverde del Camino (n=88), Villablanca (n=28) y Niebla (n=72). En la provincia de Granada se seleccionaron las localidades de Huéscar (n=123), Baza (n=91), Montefrío (n=128), Loja (n=78), Alhama de Granada (n=66) y Órgiva (n=98). El total de muestras recogidas fue, por tanto, de 447 en Huelva (191 mujeres y 256 varones) y de 584 en Granada (295 mujeres y 289 varones). La muestra total representa a 1.031 individuos andaluces.

Las muestras biológicas consistieron en sangre periférica. Este procedimiento resulta más complejo, puesto que se requiere la presencia obligatoria de personal sanitario. Sin embargo, las ventajas de disponer de sangre total frente a otro tipo de muestras biológicas (p.e. saliva) son claras. Esas ventajas se centran en la posterior extracción de ADN, y su mejor rendimiento, referido a una mayor cantidad y concentración de ADN. Estas particularidades, sin duda, facilitan los subsiguientes análisis de genotipado (Philibert et al., 2008).

Durante el proceso de muestreo, se obtuvo asimismo el consentimiento informado de cada participante junto con la muestra biológica. Los donantes proporcionaron amablemente datos acerca de sus orígenes familiares (hasta la tercera generación), así como de su parentesco biológico con otros posibles participantes. Esta es una información esencial para el posterior filtrado de las muestras y los consiguientes análisis genéticos y estadísticos de los datos. El proceso de muestreo y recogida de las muestras biológicas así como los objetivos de nuestra investigación en Andalucía fueron evaluados positivamente por la Comisión de Bioética de la Universidad Complutense de Madrid (UCM), cumpliendo con la legislación vigente (Ley 14/2007, de 3 de julio, de Investigación Biomédica).

A cada muestra se le asignó una codificación individual, lo que garantiza el anonimato del donante y la protección de datos. Para fundamentar la representatividad de los individuos muestreados con respecto a sus provincias, realizamos un análisis del origen geográfico de los mismos. En la **Figura II-11** se puede observar la distribución de los lugares de nacimiento (por municipios) de los donantes de la muestra sanguínea del presente estudio. En total, estarían representados 48 municipios de los 79 de Huelva (~60%) y 57 de los 169 municipios de los que está formada la provincia de Granada (~34%). Factores como las diferencias en extensión media por localidad dentro de cada una de las provincias (Huelva: 128,5 km², Granada: 74,8 km²) o una mayor movilidad dentro de la provincia de Huelva podrían explicar las diferencias en representatividad.

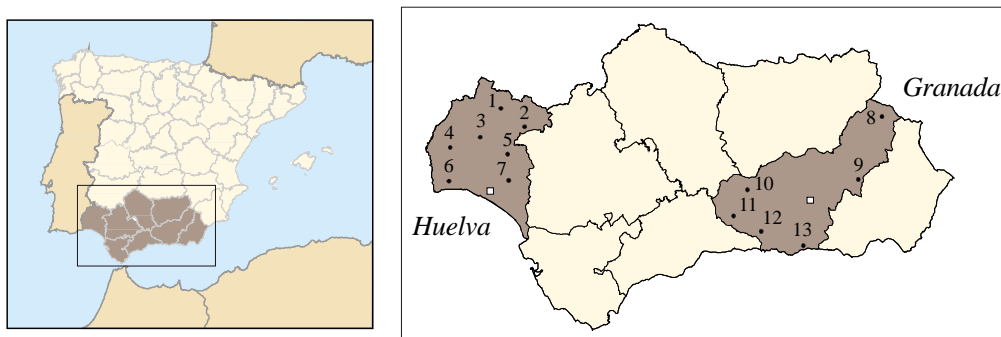
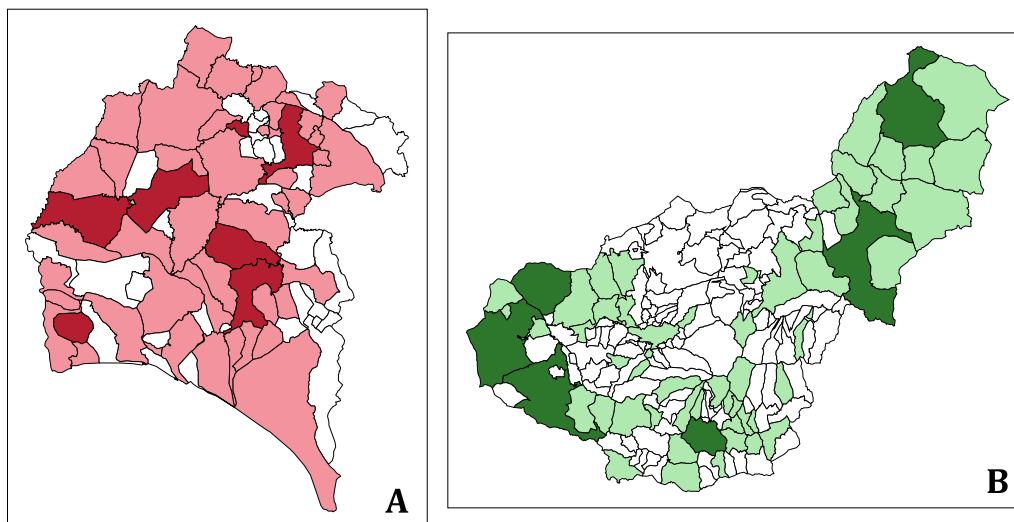


Figura II-10. Posición de las localidades de muestreo en Huelva y Granada. Las capitales se indican con cuadrados blancos (1: El Repilado, 2: Aracena, 3: El Cerro del Andévalo, 4: La Puebla de Guzmán, 5: Valverde del Camino, 6: Villablanca, 7: Niebla, 8: Huéscar, 9: Baza, 10: Montefrío, 11: Loja, 12: Alhama de Granada, 13: Órgiva).

Figura II-11. Municipios representados en el conjunto de muestras tomadas para el presente estudio. Huelva (A) y Granada (B). En ambos casos se indica en color más oscuro los municipios de muestreo.



3. Obtención de material genético

3.1. Aislamiento y extracción del ADN genómico

1. Transferir el volumen total a tubos Falcon 50 ml. Completar con solución NaCl al 0,9%.
2. Centrifugar (18 min., 3.300 r.p.m., 4°C) y eliminar el sobrenadante.
3. Añadir agua destilada hasta alcanzar 25 ml y Tampón de Lisis I hasta 50 ml.
4. Incubar en hielo con agitación (20-30 min.).
5. Centrifugar (15 min., 3.800 r.p.m., 4°C) y eliminar el sobrenadante.
6. Resuspender el *pellet* con Tampón de Lisis I (hasta los 25 ml).
7. Incubar en hielo con agitación (10 min.).
8. Centrifugar (15 min., 3.800 r.p.m., 4°C) y eliminar el sobrenadante.
9. Resuspender del *pellet* (añadiendo 5 ml de Tampón de Lisis II, 0,3 ml de SDS al 10% y 0,75 ml de solución de proteinasa K)
10. Agitar en rotor orbital (18 horas a 37°C).
11. Añadir 1,5 ml de NaClO₄ 5M.
12. Agitar en rotor orbital (2 horas a 37°C).
13. Añadir 1,5 ml de NaCl 6M. Agitar y obtener una emulsión completa.
14. Centrifugar (20 min., 3400 r.p.m., temperatura ambiente).
15. Transferir el sobrenadante a otro tubo estéril.
16. Centrifugar (20 min., 3400 r.p.m., temperatura ambiente).
17. Transferir el sobrenadante a otro tubo estéril.
18. Añadir 5 ml de isopropanol frío al 100 % y mezclar suavemente hasta que aparezca el precipitado de ADN.
19. Lavar el ADN en etanol 70%.
20. Eliminar el etanol y resuspender el ADN en 500 µl de agua ultrapura.
21. Mantener en placa calefactora (2-3 horas a 55°C).
22. Almacenamiento (24 horas a 4°C).

Tampón Lisis I [Sacarosa (0,3M), Tris-HCl (pH= 7,5; 10mM), MgCl₂ (5mM), Tritón X-100 (1%)]. Tampón Lisis II [NaCl (0,075M), Na₂-EDTA (0,024M)].

Las muestras de sangre (~5-7 ml por individuo) fueron depositadas en tubos EDTA en el momento de su recolección. Posteriormente, se trasladaron al Laboratorio de Antropología Molecular (Facultad de Biología, UCM, ver **Figura II-12**). El primer paso a abordar con las muestras, una vez que llegaron a nuestro Laboratorio, fue la centrifugación para la separación del plasma, el cual fue almacenado por separado para permitir el estudio de otros marcadores genéticos (p.e. alotipos de las inmunoglobulinas). Tras esta etapa inicial, se procedió a la extracción del ADN genómico, siguiendo el protocolo que se indica arriba.



Figura II-12. Laboratorio de Antropología Molecular (Dpto. de Zoología y Antropología Física, Facultad de Biología, UCM).

3.2. ADN. Su cuantificación, control de calidad y estandarización de las muestras

Tras la extracción, se procedió a la cuantificación del ADN y a la valoración de su calidad por medio de la espectrometría óptica (BioPhotometer; Eppendorf, Westbury, NY). Mediante este procedimiento se estandariza la concentración de ADN, lo que es fundamental para una aplicación más eficiente de las técnicas moleculares, esencialmente durante la fase de amplificación.

Para todas las muestras se obtuvieron valores de concentración y de pureza de ADN. La concentración se presenta en valores de ng/ μ l, mientras que la calidad del ADN viene dada por la relación entre las lecturas de las absorbancias a 260 y 280 nm. Se considera que, preparaciones puras de ADN ofrecen relaciones A260/A280 alrededor de 1,8. Los valores medios de concentración fueron: 458,17 \pm 290,97 ng/ μ l para las muestras de Huelva y 706,56 \pm 264,21 ng/ μ l para las de Granada; la ratio de pureza fue muy similar, 1,60 \pm 0,12 y 1,61 \pm 0,05, respectivamente.

Con el dato de concentración de cada una de las muestras se pudieron preparar diluciones estandarizadas de ADN (100 ng/ μ l).

3.3. Almacenamiento de las muestras

Los tubos derivados de la extracción, con su concentración original, se encuentran almacenados en nuestro Laboratorio de Antropología Molecular a -80°C. Las diluciones se congelaron a -20°C para facilitar su uso.

4. Criterios de filtrado y selección de muestras

Las muestras de ADN de las dos poblaciones andaluzas han sido y están siendo analizadas para una amplia batería de marcadores genéticos, entre los que se incluyen polimorfismos autosómicos (Calderón et al., 2006; Coudray et al., 2007; Fortes-Lima et al., 2014; Reales et al., 2014) y del cromosoma Y (Ambrosio et al., 2010a, 2010b, 2012; Calderón et al., 2015). Debido a los diversos patrones de herencia de estos marcadores (uniparental o autosómica) la selección de las muestras autóctonas es diferencial.

En el presente estudio nos centramos en marcadores genómicos que obedecen a una herencia materna (matrilínea). En consecuencia, el primer filtrado que realizamos en nuestra base de datos fue la selección de todos aquellos donantes cuyas abuelas maternas hubieran nacido en alguno de los dos territorios seleccionados (Huelva/Granada). Este procedimiento es clave en la medida en que el grado de autoctonía de la muestra está asegurado más allá del lugar de nacimiento del individuo objeto de estudio. Gracias a la información proporcionada por los donantes, también fueron eliminados los parientes

biológicos, seleccionando al azar una persona entre el grupo de parientes (p.e. hermano/hermana, madre/hijo, tío(a)-sobrina(o), etc.). Con todas estas precauciones, el número de muestras 'útiles' (autóctonas y no-emparentadas entre sí) para nuestro estudio sobre la diversidad mitocondrial en Andalucía fue de 280 para Huelva y de 470 en Granada, lo que representa un total de 750 muestras. Teniendo en cuenta que las muestras sanguíneas recogidas fueron $n=1.031$, el 72,74% de las mismas cumplieron los criterios de autoctonía para la línea materna. Este resultado confiere una especial solidez al proceso de muestreo y selección de las localidades.

A partir de esas 750 muestras, se tomaron aleatoriamente 279 (158 de Huelva, 121 de Granada) para los análisis descritos en el **Capítulo III**, es decir, la caracterización global del ADNmt en el sur de Iberia. Las 158 muestras pertenecientes a Huelva formaron parte de la Tesis Doctoral de un miembro de nuestro equipo (Ambrosio, 2011). Los resultados derivados de este trabajo se han enriquecido con un re-análisis de las citadas muestras a mayor resolución, y nos llevaron a estudiar también, con un objetivo mucho más prospectivo, la diversidad mitocondrial en Andalucía, pero abordando esta vez su extremo geográfico oriental (Granada). La investigación tuvo otra interesante proyección al centrarnos en la herencia africana en el sur peninsular (**Capítulo IV**) y en el haplogrupo materno más representado de Europa (clado H) (**Capítulo V**). Dentro del segundo y tercer eje de investigación, realizamos una sensible ampliación del número de muestras andaluzas a genotipar de aquellas disponibles en nuestro Laboratorio, es decir, 471 muestras, 122 de Huelva y 349 de Granada. Por tanto, el amplio conjunto de muestras autóctonas que hemos analizado ($n=750$) otorga robustez a los resultados obtenidos, ya que podemos asumir que estaríamos considerando toda la diversidad mitocondrial presente en Andalucía.

5. Técnicas de análisis de marcadores moleculares mitocondriales

5.1. Secuenciación de la región control del ADNmt

La secuenciación de la región control se aplicó al estudio de la diversidad mitocondrial en un marco amplio general (muestras de Granada, **Capítulo III**) y del haplogrupo H (muestras de Huelva y Granada, **Capítulo V**). Se amplificó la región hipervariable I (HVS-I) y parte de la HVS-II, representando un total de 820 pares de bases. Las condiciones se detallan en la **Tabla II-1**. La amplificación del ADN se llevó a cabo en un termociclador GeneAmp 9700 (Life Technologies, Foster City, CA, USA). Tras la reacción PCR (*Polymerase Chain Reaction*) se comprobó la presencia de bandas mediante un procedimiento de electroforesis estándar (con un gel de agarosa al 2%, Panreac Química SAU). Los productos se purificaron con SpeedTools PCR Clean-Up kit (Biotools B&M Labs. S.A., Madrid, España). Por último, las reacciones de secuenciación y lectura de marcadores se llevaron a cabo en la Unidad de Genómica (UCM-Parque Científico de Madrid) y en SECUGEN (Madrid). Para todas las muestras se secuenciaron ambas cadenas

de ADN usando los *primers* F15973 y R296. La presencia de un fragmento homopolimérico de citosinas (conocido como *stretch*) entre las posiciones 16148-16193 (respecto a la rCRS) produce un deslizamiento de la ADN polimerasa y, por tanto, una interrupción de la lectura. En esos casos se secuenciaron los fragmentos 4 veces, 2 en sentido F (*Forward*) y dos en R (*Reverse*) con el objetivo de obtener confirmación de los polimorfismos.

Tabla II-1. Secuencias de los *primers* (Coudray et al., 2009) y condiciones para la amplificación de la región control (HVS-I) del ADNmt. Todos los reactivos son de Life Technologies, excepto los *primers* (Metabion International, Martinsried, Alemania).

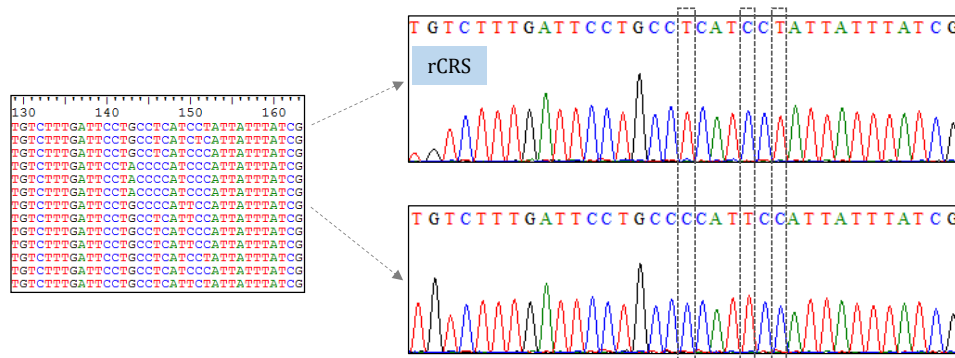
F15973	5'- AACTCCACCATTAGCACCCA-3'	
R296	5'-GGAAATTTTTTGTATGATGTCT-3'	
H ₂ O mQ		34,5
Buffer (10x)		5,0
MgCl ₂ (25mM)		6,0
dNTPs (10mM)		1,0
F15973 (10μM)		1,0
R296 (10μM)		1,0
AmpliTaq Gold (5U/μl)		0,5
ADN (100ng/μl)		1,0

94°C	10 min.	35 ciclos
94°C	1 min.	
58°C	1 min.	
72°C	1 min. 30 seg.	
72°C	5 min.	

Vol./individuo = 50 μl

La visualización de las secuencias se realizó gracias al programa Sequence Scanner v.1 (Life Technologies) y todas ellas fueron revisadas y corregidas manualmente. A continuación, se realizó el alineamiento y la obtención de la secuencia consenso. La **Figura II-13** nos muestra un alineamiento de la secuencia de referencia (rCRS, Andrews et al., 1999) y unas secuencias problema, donde se observa la detección de varios polimorfismos. El programa mtDNA GeneSyn (Pereira et al., 2009) permite la extracción automática de dichas mutaciones de nuestras muestras con respecto a la secuencia rCRS.

Figura II-13. Alineamiento de secuencias control mitocondriales con la rCRS (situada en la primera fila). Los electroferogramas de la derecha muestran la rCRS (superior) y una muestra problema (inferior) con 3 transiciones respecto a la rCRS.



5.2. PCR-RFLPs

La secuenciación de la región control conducente a la definición de haplotipos permite aproximarse a la asignación de cada muestra a los haplogrupos y sub-haplogrupos mitocondriales correspondientes. Sin embargo, en muchas ocasiones y, fundamentalmente, cuando nos enfrentamos a linajes muy cercanos a la secuencia de referencia rCRS (perteneciente al sub-haplogrupo H2a2a), las regiones hipervariables no son demasiado informativas por lo que se requiere la presencia de otros polimorfismos para poder clasificar a una muestra en un determinado clado. Para ello, la estrategia metodológica que hemos seguido para las muestras de Granada analizadas en el **Capítulo III** fue la técnica combinada de PCR y RFLPs (*Restriction Fragment Length Polymorphisms*), la cual conlleva los siguientes pasos:

A. Diseño experimental y amplificación del fragmento que incluye la mutación. En algunos casos, el protocolo experimental estaba establecido previamente (Torrioni et al., 1994, 1996; Macaulay et al., 1999), en otros, sin embargo, se tuvo que diseñar *de novo*. Para ello, primero se debe localizar el polimorfismo (marcador genético) que se sospeche pueda poseer el individuo (dato que se obtiene del análisis previo de la secuencia de la región control), consultando la filogenia mitocondrial humana (van Oven & Kayser, 2009). Tras esta fase, se diseñan *primers* (usando el programa Primer Express v. 3.0, Life Technologies) que engloben la mutación, y, por otro lado, se introduce la secuencia en cuestión en herramientas informáticas (como por ejemplo Restriction Mapper v. 3, www.restrictionmapper.org o NEBcutter v. 2.0, Vincze et al., 2003) que permiten la detección de dianas de restricción (*targets*) de una batería de enzimas en dicha secuencia. A continuación, se elige la enzima cuyo sitio de corte se vea afectado por el polimorfismo a detectar (ver **Figura II-14**).

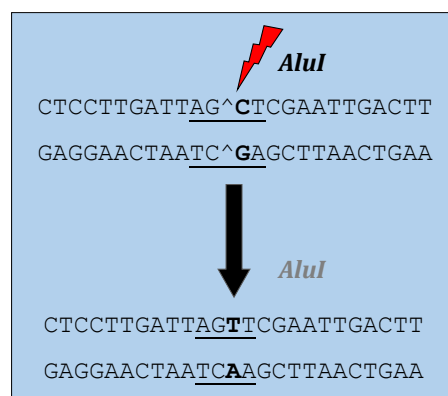


Figura II-14. Inactivación de la diana de restricción de la enzima *AluI* por una transición C>T. En la secuencia superior, el fragmento en cuestión posee una diana para la enzima. Sin embargo, la ocurrencia de una mutación (ver secuencia inferior) impide la reacción enzimática.

La **Tabla II-2** muestra las condiciones para la amplificación. Los *primers* son de Metabion International (Martinsried, Alemania) y el resto de componentes, de

Life Technologies. Tras la amplificación, se realiza una comprobación de la reacción en un gel de agarosa al 2%.

Tabla II-2. Condiciones para la amplificación de fragmentos que son sometidos posteriormente a la digestión enzimática.

H ₂ O mQ	16,75	95°C	5 min.	35 ciclos
Buffer (10x)	2,50	95°C	1 min.	
MgCl ₂ (25mM)	3,00	56°C	1 min.	
dNTPs (10mM)	0,50	72°C	1 min. 10 seg.	
Primer F (10μM)	0,50	72°C	10 min.	
Primer R (10μM)	0,50			
AmpliTaq Gold (5U/μl)	0,25			
ADN (100ng/μl)	1,00			

Vol./individuo = 25 μl

B. Reacción de digestión enzimática. Este paso requiere el contacto del fragmento amplificado con 5 unidades de la enzima en cuestión. La mezcla se mantiene en incubación a 37°C en un baño termostático (Raypa) alrededor de 12 horas, tiempo necesario para lograr la digestión total de los fragmentos amplificados de PCR. La lectura de la digestión se realiza también mediante un gel de agarosa estándar al 2% y agarosa NuSieve (Lonza Rockland, Inc., Rockland, ME, USA) al 1%. Este último compuesto es una agarosa de bajo punto de fusión y se emplea para obtener un gel con una trama más estrecha, con el fin de lograr una correcta separación de los fragmentos obtenidos tras la digestión.

Un ejemplo de estas lecturas por electroforesis lo tenemos en la **Figura II-15**. Podemos observar la diferencia entre los patrones de bandas obtenidos en función de la posesión de una diana de restricción o no. La primera mutación (3348, Hg U6) es una transición A>G que produce la ganancia de una diana de restricción para la enzima *MboI*. Esto significa que, si el individuo es U6 (como es el caso de las muestras problema 2, 3 y 4), la secuencia será cortada por la enzima y se generarán 2 fragmentos (de 311 y 263 pb). En el caso contrario, si se posee una adenina (A) en la posición 3348, no existirá diana para la enzima *MboI*, por lo que solo se podrá observar un fragmento de 574 pb (ver carril 5, control negativo). En el caso de la transición A14179G (Hg U6a) se produce la misma situación: la pertenencia al haplogrupo conlleva la ganancia de una diana de restricción, en este caso, para la enzima *AccI* y, por tanto, se generan 2 fragmentos, de 224 y de 77 pb (éste último no se observa por su reducido tamaño). Para el individuo no-U6a (carril 8, control negativo) observaremos el fragmento sin digerir (301 pb).

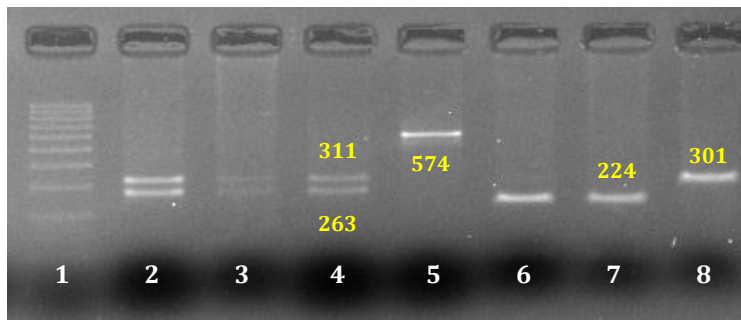


Figura II-15. Patrón de restricción para el análisis de las posiciones 3348 (carriles 2-4, determinación del Hg U6) y 14179 (carriles 6-8, Hg U6a). Los carriles 5 y 8 corresponden a controles negativos para dichas posiciones. En color amarillo se indica el tamaño de los fragmentos en pares de bases.

En ciertas ocasiones no fue posible encontrar una diana de restricción para alguna enzima que se viera afectada por el polimorfismo que queríamos analizar. En ese caso, se procedió al diseño de un par de *primers* (con las herramientas informáticas mencionadas anteriormente) y se realizó una secuenciación directa de dicha mutación para la confirmación de su *status*. La evaluación de las mutaciones se llevó a cabo de manera jerárquica, es decir, comenzando por las mutaciones que caracterizaban a los macro-haplogrupos para, posteriormente, ir descendiendo de acuerdo al árbol genealógico mitocondrial. La **Tabla II-3** muestra un resumen de las variantes de la región codificante analizadas y de la técnica empleada en cada caso.

Además de lo reflejado en el **Capítulo III**, la técnica de PCR-RFLPs fue empleada también para la caracterización del sub-haplogrupo H1b1 (A3796G) en el estudio del haplogrupo H (**Capítulo V**).

5.3. Ensayos de PCR a tiempo real

Otro procedimiento para la evaluación de SNPs es la PCR a tiempo real (*Real Time PCR, PCR-RT*). Ello representa un salto metodológico significativo con respecto a la técnica PCR-RFLP. La razón está en que permite un genotipado automatizado de las variantes nucleotídicas. Estos ensayos de genotipado se denominan *TaqMan Genotyping Assays* (Life Technologies) y el desarrollo experimental se muestra en la **Figura II-16**. Cada ensayo consta de un tubo que incluye dos *primers* que amplifican la secuencia de interés y dos sondas (*probes*) para distinguir entre los dos alelos. Las sondas están compuestas por un marcador fluorescente (*reporter dye*) en el extremo 5', con dos opciones: VIC (alelo 1, marcaje en verde) y 6-FAM (alelo 2, azul) las cuales se alinearán con la secuencia problema, dependiendo del *status* del SNP de interés. El alineamiento provoca una fluorescencia determinada. Además, las sondas poseen otras moléculas asociadas (*minor groove binder, MGB* y *non-fluorescent quencher, NFQ*) en el extremo 3' que permiten un incremento en la discriminación alélica y evitan la aparición de falsos positivos.

Tabla II-3. Polimorfismos analizados de la región codificante mitocondrial. Se señala la mutación que caracteriza a cada haplogrupo, la técnica empleada y la enzima usada en el caso de RFLPs. Se indica también la diana de restricción, con acento circunflejo en la posición de corte en la secuencia nucleotídica (N=A, G, C, ó T; K=G ó T; M=A ó C; R=A ó G; W=A ó T; Y=C ó T).

Haplogrupo	Posición	SNP	Técnica de análisis	Enzima
H	7028	C-T	PCR-RFLP	<i>Alu I</i> (AG [^] CT)
H1, J1	3010	G-A	PCR-RFLP	<i>Taq I</i> (T [^] CGA)
H1b1	3796	A-G	PCR-RFLP	<i>Aci I</i> (C [^] CGC)
H2	1438	A-G	PCR-Secuenciación	-
H2a	4769	A-G	PCR-RFLP	<i>Alu I</i> (AG [^] CT)
H3	6776	T-C	PCR-RFLP	<i>CviA II</i> (C [^] ATG)
HV	14766	C-T	PCR-RFLP	<i>Mse I</i> (T [^] TAA)
HV0a	15904	C-T	PCR-RFLP	<i>Mse I</i> (T [^] TAA)
HV1	15218	A-G	PCR-Secuenciación	-
I	10034	T-C	PCR-RFLP	<i>Alu I</i> (AG [^] CT)
I3	2849	G-A	PCR-RFLP	<i>TatI</i> (W [^] GTACW)
JT	4216	T-C	PCR-Secuenciación	-
J	13708	G-A	PCR-RFLP	<i>Bst NI</i> (CC [^] WGG)
J1b	13879	T-C	PCR-Secuenciación	-
J1c	14798	T-C	PCR-RFLP	<i>Pac I</i> (TTAAT [^] TAA)
J2	7476	C-T	PCR-RFLP	<i>Alu I</i> (AG [^] CT)
UK	12308	A-G	PCR-RFLP	<i>Hinf I</i> (G [^] ANTC)
K	9055	G-A	PCR-RFLP	<i>Hae II</i> (RGC [^] GY)
K1	1189	T-C	PCR-RFLP	<i>Rsa I</i> (GT [^] AC)
L3-M	10873	T-C	PCR-RFLP	<i>Mnl I</i> (CCTC(N) ₇ [^])
L0, L1, L2, L5, L3	3594	C-T	PCR-RFLP	<i>Hpa I</i> (GTT [^] AAC)
N1	10238	T-C	PCR-RFLP	<i>HphI</i> (GGTGA(N) ₃ [^])
R0	11719	G-A	PCR-RFLP	<i>Hae II</i> (RGC [^] GY)
T	13368	G-A	PCR-RFLP	<i>Tsp509 I</i> ([^] AATT)
T1	12633	C-A	PCR-RFLP	<i>Tsp509 I</i> ([^] AATT)
T2	14233	A-G	PCR-RFLP	<i>Tsp509 I</i> ([^] AATT)
U2, U3, U4, U7, U8, U9	1811	A-G	PCR-RFLP	<i>Psi I</i> (TTA [^] TAA)
U3	14139	A-G	PCR-RFLP	<i>Mbo I</i> ([^] GATC)
U3a	9266	G-A	PCR-RFLP	<i>Hae III</i> (GG [^] CC)
U4	4646	T-C	PCR-RFLP	<i>Rsa I</i> (GT [^] AC)
U5	13617	T-C	PCR-RFLP	<i>Tsp509 I</i> ([^] AATT)
U5a	14793	A-G	PCR-RFLP	<i>Bsr BI</i> (CCG [^] CTC)
U5b	14182	T-C	PCR-Secuenciación	-
U6	3348	A-G	PCR-RFLP	<i>Mbo I</i> ([^] GATC)
U6a	14179	A-G	PCR-RFLP	<i>Acc I</i> (GT [^] MKAC)
U6b	9438	G-A	PCR-RFLP	<i>Hae III</i> (GG [^] CC)
W	8994	G-A	PCR-RFLP	<i>Hae III</i> (GG [^] CC)
W3	13263	A-G	PCR-RFLP	<i>Alu I</i> (AG [^] CT)
X	14470	T-C	PCR-RFLP	<i>Acc I</i> (GT [^] WRAC)
X2	1719	G-A	PCR-RFLP	<i>Dde I</i> (C [^] TNAG)
X2b	8393	C-T	PCR-RFLP	<i>Hae III</i> (GG [^] CC)
X2d	8503	T-C	PCR-RFLP	<i>Psi I</i> (TTA [^] TAA)

Esta técnica permite detectar homocigotos y heterocigotos de una manera gráfica en un *plot* bidimensional. En nuestro caso, como nos encontramos con un marcador haploide, solo aparecen dos categorías de homocigotos para cada uno de los alelos.

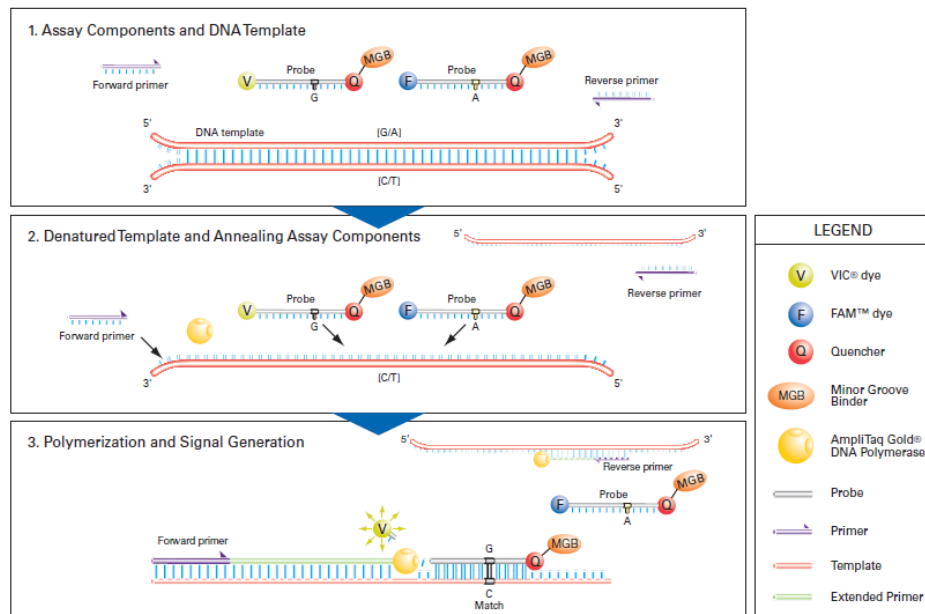


Figura II-16. Fundamentos químicos de los ensayos TaqMan de genotipado de SNPs. Fuente: *TaqMan SNP Genotyping Assays User Guide*.

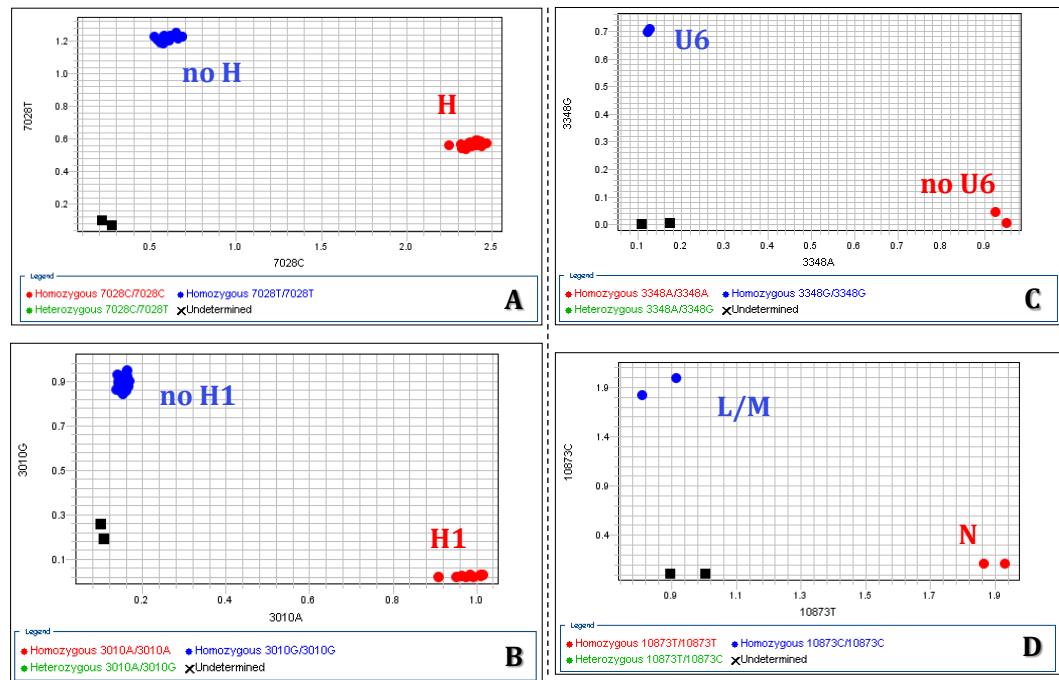
Esta tecnología ha sido utilizada en este trabajo de Tesis Doctoral para alcanzar diversos objetivos en nuestra investigación. La técnica PCR-RT conduce a una rápida clasificación de los individuos en los linajes mitocondriales de interés. En el **Capítulo IV** empleamos dos ensayos *TaqMan* para la determinación de los haplogrupos africanos U6 y L/M (posiciones A3348G y T10873C, respectivamente). Del mismo modo, en el **Capítulo V** diseñamos otros dos ensayos para la clasificación de muestras como pertenecientes a los linajes europeos H o H1 (C7028T y A3010G, respectivamente). Los ensayos se diseñaron *de novo* en nuestro Laboratorio, empleando el programa Primer Express v. 3.0 (Life Technologies). Los detalles del diseño (*primers*, sondas y condiciones de la reacción) se exponen en los Capítulos correspondientes. El desarrollo de la técnica se llevó a cabo en un equipo *StepOne* (Life Technologies) y la amplificación mediante *TaqMan Genotyping Master Mix* (Life Technologies). En las 4 reacciones de genotipado se empleó el mismo programa de amplificación, que se detalla en la **Tabla II-4**.

Tabla II-4. Programa de PCR para la amplificación a tiempo real.

60°C	30 seg	50 ciclos
95°C	10 min.	
92°C	15 seg.	
60°C	1 min.	
60°C	30 seg.	

Los resultados obtenidos se muestran en unos gráficos (*allelic discrimination plots*) en los que se observa el genotipado de las muestras problema dependiendo de la posición en la que se sitúe cada una (ver **Figura II-17**). Estos *plots* se visualizan y analizan en el programa *StepOne Software* (Life Technologies).

Figura II-17. Genotipado de polimorfismos mediante sondas TaqMan y PCR a tiempo real para la clasificación de muestras en H (A), H1 (B), U6 (C) y L/M (D). Los cuadrados negros indican los blancos empleados en cada reacción.



5.4. Secuenciación completa del genoma mitocondrial

Con el fin de lograr una disección profunda de la arquitectura molecular de los linajes maternos africanos en el sur de Iberia (objetivo del **Capítulo IV**) planteamos alcanzar el máximo nivel de resolución posible a través de la secuenciación completa del genoma mitocondrial.

Para ello, seleccionamos en primer lugar las muestras candidatas para ser completamente secuenciadas. Razones económicas han hecho que no sea abordable el desarrollo de esta técnica utilizando un tamaño muestral elevado. Los detalles acerca del proceso de selección de muestras se exponen en el **Capítulo IV**. En concreto, se eligieron 32 individuos andaluces pertenecientes a los haplogrupos U6, M1 y L. Del mismo modo, y con el fin de enriquecer el estudio, fueron seleccionadas otras 30 muestras de bereberes de Marruecos pertenecientes a los mismos linajes mitocondriales. Dichas muestras norteafricanas fueron cedidas por el Dr. Jean M. Dugoujon, miembro de nuestro equipo de investigación.

La secuenciación completa del genoma mitocondrial se llevó a cabo amplificando sus 16.569 pb mediante 32 fragmentos solapantes gracias a los *primers* propuestos por Maca-Meyer et al., 2001. Dichos *primers* incluyen las posiciones detalladas en la **Tabla II-5**.

El proceso de secuenciación de las 62 muestras andaluzas y marroquíes se llevó a cabo enteramente en los laboratorios del *Instituto de Patología e Imunología Molecular da Universidade do Porto* (IPATIMUP) bajo la supervisión

de la Dra. L. Pereira. Para la amplificación de los 32 fragmentos, se emplearon las condiciones descritas en la **Tabla II-6**.

Fragmento	Rango	Fragmento	Rango
1	16318-429	17	8280-8882
2	362-964	18	8779-9416
3	902-1508	19	9342-9950
4	1445-2073	20	9865-10481
5	2004-2612	21	10383-10994
6	2538-3128	22	10930-11546
7	3051-3690	23	11467-12095
8	3625-4247	24	12008-12623
9	4189-4813	25	12553-13143
10	4729-5327	26	13068-13685
11	5259-5851	27	13593-14206
12	5762-6387	28	14104-14705
13	6318-6918	29	14629-15232
14	6850-7427	30	15143-15739
15	7358-7937	31	15657-16180
16	7861-8366	32	15975-16420

Tabla II-5. Fragmentos y rangos de la secuencia de referencia rCRS para la secuenciación completa del genoma mitocondrial (Maca-Meyer et al., 2001).

H ₂ O mQ	11,8	94°C	2 min.	35 ciclos
Buffer (10x)	5	94°C	30 seg.	
MgCl ₂ (25mM)	1,5	60°C	30 seg.	
dNTPs (10mM)	0,5	72°C	30 seg.	
Primer F (10µM)	2,5	72°C	10 min.	
Primer R (10µM)	2,5			
AmpliTaq Gold (5U/µl)	0,2			
DNA (100ng/µl)	1			

Vol./individuo = 25 µl

Tabla II-6. Condiciones para la amplificación inicial de los fragmentos empleados para la caracterización del genoma mitocondrial.

La amplificación del ADN se comprobó en un gel de poliacrilamida (acrilamida al 25%). La reacción de secuenciación se llevó a cabo empleando BigDye v3.1 Terminator Cycle Sequencing Kit y un equipo 3100 DNA Analyzer (todo de Life Technologies). Para todos los fragmentos, empleamos los *primers* F en la reacción de secuenciación. Las condiciones de la misma se detallan en la **Tabla II-7**.

Buffer	1	96°C	4 min.	35 ciclos
Kit secuenciación	1	96°C	15 seg.	
Primer F (10µM)	0,5	50°C	9 seg.	
Producto amplificado	2,5	60°C	2 min.	
		60°C	10 min.	

Vol./individuo = 5 µl

Tabla II-7. Condiciones para la secuenciación de los 32 fragmentos solapantes.

Las secuencias obtenidas fueron visualizadas y corregidas manualmente tal y como se expone en el apartado 5.1. Se debe tener en cuenta que algunos de los fragmentos pueden generar problemas durante su secuenciación debido a la presencia de ciertos polimorfismos (generación de fragmentos homopoliméricos o *stretches*). En los casos necesarios, se emplearon los *primers* R para solucionar estas incidencias. Finalmente, los 32 fragmentos solapantes fueron alineados y combinados en una secuencia consenso.

5.5. Genotipado múltiple por SNaPshot

Otro procedimiento para una refinada definición de sub-clados dentro de haplogrupos mitocondriales es el genotipado por SNaPshot (Life Technologies). Se trata de un *kit* que permite analizar en una misma reacción un número considerable de polimorfismos (<40, Salas et al., 2005a). Las ventajas de esta metodología son evidentes por representar un ahorro de tiempo frente a otras estrategias de genotipado, por ejemplo, la secuenciación directa o las reacciones PCR-RFLPs. Del mismo modo, resulta muy indicado en aquellos casos en que se quiere llevar a cabo un estudio a alta resolución y no se plantea realizar el análisis del genoma mitocondrial completo. Esta última estrategia la hemos empleado en el **Capítulo V** para el estudio de la diversidad interna del clado H en Andalucía. Dado el elevado volumen de individuos pertenecientes al linaje H en nuestras muestras poblacionales (~50%), la secuenciación completa habría sido inviable, simplemente por razones de presupuesto.

La **Figura II-18** detalla las bases de esta metodología. En primer lugar, se diseñan *primers* específicos para la amplificación de un fragmento que incluya la mutación de interés. Tras esta reacción PCR, se ponen en contacto los fragmentos amplificados con unas 'sondas' (fragmentos más cortos que finalizan en el nucleótido anterior al SNP a analizar) y con los 4 dideoxi-nucleótidos, cada uno de ellos marcados con un fluorocromo (ddATP: dR6G; ddCTP: dTAMRA™; ddGTP: dR110; ddTTP: dROX™). Esta reacción de minisequenciación está seguida de la lectura de los fragmentos obtenidos. El tamaño de las sondas es el que determina el orden de lectura de los picos, los cuales reflejan con su color el nucleótido incorporado. En el diseño de este tipo de reacciones, hay que tener en cuenta que debe dejarse un margen mínimo de ~4 pb para que los picos sean distinguibles entre sí.

El diseño de los *primers* de la amplificación inicial y las sondas fue adaptado a partir del trabajo de Álvarez-Iglesias et al., 2009. Se seleccionaron 21 SNPs que caracterizan a los sub-haplogrupos H más frecuentes en Iberia y se organizaron en 3 reacciones *multiplex* (ver detalles en el **Capítulo V**). Las reacciones se llevaron a cabo como se indica a continuación:

A. Amplificación inicial. En primer lugar, se amplificaron en la misma reacción PCR fragmentos de ADN que contienen los SNPs de interés, empleando las condiciones descritas en la **Tabla II-8**. Posteriormente, los productos PCR obtenidos fueron purificados con ExoSAP-IT (Amersham Biosciences, Uppsala,

Tabla II-8. Programa para la amplificación inicial y la reacción de minisequenciación en el genotipado por tecnología SNaPshot.

Amplificación inicial		30 ciclos
95°C	15 min.	
94°C	30 seg.	
53°C	1 min. 30 seg.	
72°C	1 min. 30 seg.	
72°C	15 min.	
Minisequenciación		25 ciclos
96°C	10 seg.	
50°C	5 seg.	
60°C	30 seg.	

Suecia) mediante una incubación 15 minutos a 37°C e inactivación enzimática 15 minutos a 85°C.

B. Reacción de minisequenciación. Los fragmentos purificados se pusieron en contacto con el *kit* de SNaPshot (ver condiciones en **Tabla II-8**). De nuevo, se llevó a cabo una purificación con SAP (Amersham Biosciences) 60 minutos a 37°C e inactivación 15 minutos a 85°C.

Los fragmentos se llevaron para su análisis al centro de genotipado SECUGEN (Madrid). Empleamos el programa Peak Scanner v1.0 (Life Technologies) para la lectura de los resultados. En la **Figura II-19** se presentan algunos patrones de las 3 reacciones *multiplex* generadas.

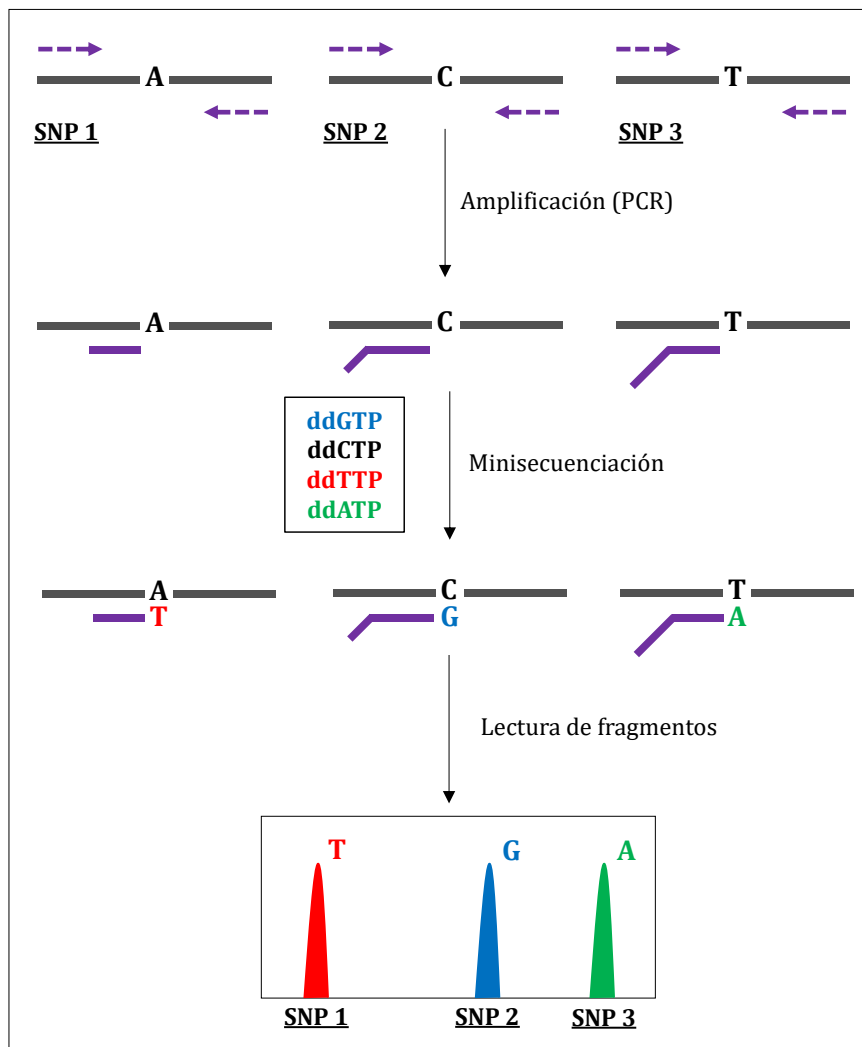


Figura II-18. Fundamentos químicos del proceso de genotipado mediante el kit SNaPshot.

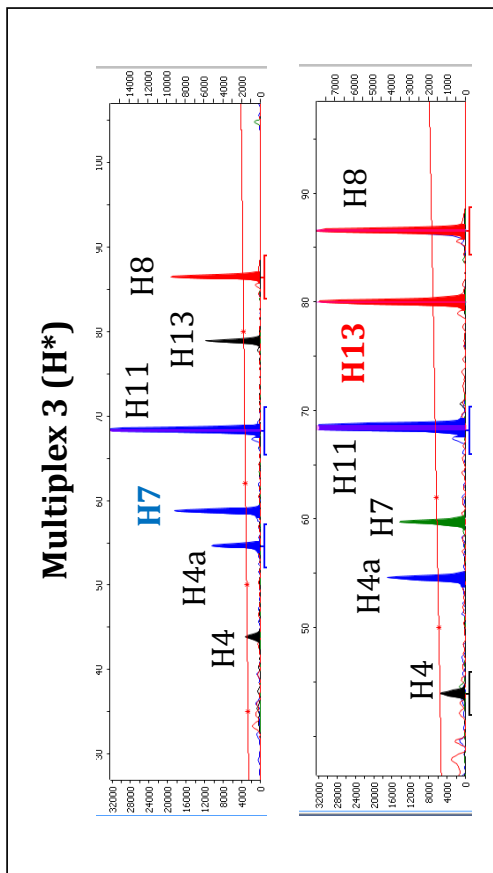
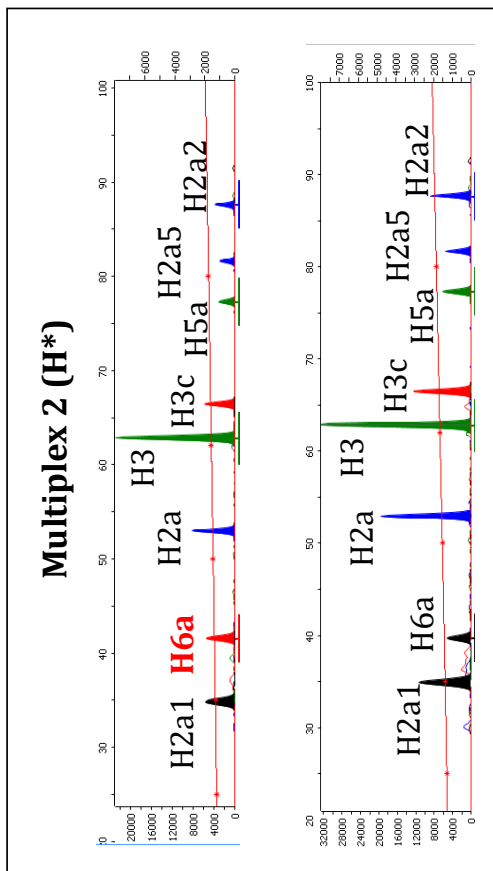
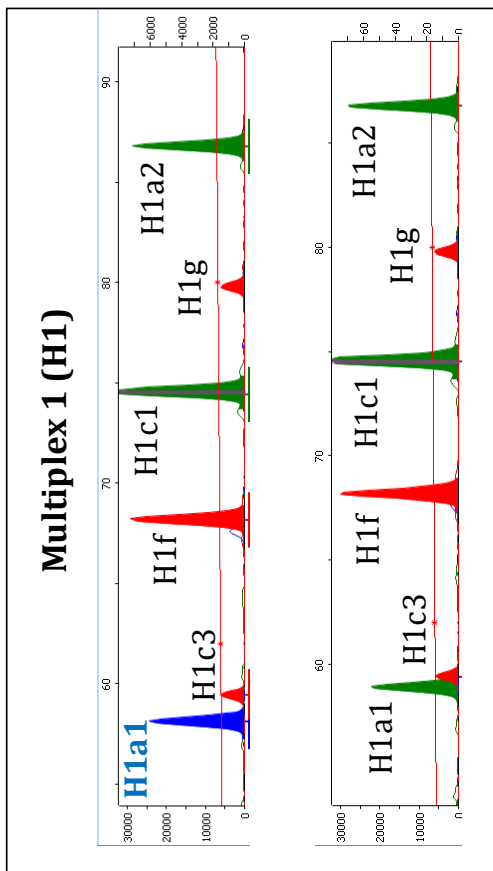


Figura II-19. Diseño de las reacciones *multiplex* en la determinación de los principales sub-haplogrupos del clado H. La reacción 1 permite el genotipado de 6 SNPs; la 2 de 8 y, la 3 de 6. El color de los picos corresponde al nucleótido presente en cada posición (A: verde, C: negro, G: azul, T: rojo). En cada reacción se muestran 2 ejemplos, resaltándose en color la confirmación de pertenencia a un sub-haplogrupo determinado.

6. El tratamiento de los datos a nivel de secuencias de ADN, estadístico y filogenético

6.1. Análisis de secuencias de ADN

Todas las secuencias de ADN obtenidas en el presente estudio fueron, en primer lugar, analizadas y corregidas manualmente. Para ello, se utilizó el programa BioEdit v. 7.2.5 (Hall, 1999). Se examinaron y verificaron los polimorfismos individualmente en los electroferogramas, y en el caso de la detección de mutaciones no descritas previamente en la literatura o en las bases de datos se volvió a secuenciar el fragmento en cuestión. El alineamiento de las secuencias entre sí para hallar la secuencia consenso se llevó a cabo con la herramienta ClustalW, incluida dentro de BioEdit. Mediante este proceso se alinearon 2 secuencias (cuando analizamos la región control) o 32 fragmentos (para la secuenciación del genoma mitocondrial completo).

El alineamiento con la secuencia de referencia rCRS se realizó también en BioEdit, empleando, en concreto, la secuencia que tiene por código NC_012920 (*GenBank Accession Number*).

En este punto se debe hacer una observación acerca de la selección de la secuencia de referencia para los estudios de ADNmt. En 2012, se publicó un artículo que condujo a un cambio de paradigma: “A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root” (Behar et al., 2012b). Sus autores proponían que la rCRS debía dejar de emplearse como referencia universal en favor de una secuencia ancestral (*Reconstructed Sapiens Reference Sequence*, RSRS) simulada a partir de la información mitocondrial que se conoce en la actualidad, incorporando también genomas Neandertales. La RSRS supondría, por tanto, un punto de referencia desde el punto de vista filogenético, situándose en el origen del árbol mitocondrial humano, y de este modo, permitiendo incorporar a la definición de las mutaciones el carácter ancestral o derivado de las mismas. Desde entonces, varias publicaciones han tenido en cuenta esta nueva secuencia para definir los polimorfismos de sus muestras (p.e. Pennarun et al., 2012; Barbieri et al., 2013). Sin embargo, el grado de aceptación no es muy alto, y las razones para ello se exponen en el también reciente artículo: “The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies” (Bandelt et al., 2014).

Teniendo en cuenta los factores expuestos y fundamentados en este último trabajo, entre los que se encuentran la posible incorporación de errores de notación, la problemática en el campo forense y de medicina legal así como la ausencia de ventajas que justifiquen el reemplazamiento de la secuencia de referencia, en la presente trabajo de Tesis Doctoral se considera siempre la rCRS para la definición de los polimorfismos mitocondriales.

Posteriormente al procesamiento de las secuencias, empleamos el programa mtDNA-GeneSyn (Pereira et al., 2009) para obtener las mutaciones (haplotipos) de cada individuo. Este *software* resulta también interesante en cuanto a que

permite generar archivos de entrada (*input*) para el programa Network (ver más abajo).

Para abordar el control de calidad de las secuencias, se siguieron los protocolos de prevención de errores en el análisis del ADNmt (Salas et al., 2005b). Se revisaron aquellas mutaciones que se encuentran fuera de lo detectado comúnmente en las bases de datos para poblaciones geográficamente próximas a la nuestra. También se tiene un especial cuidado con las denominadas *phantom mutations*, variantes no-comunes que aparecen simultáneamente en diferentes linajes mitocondriales de nuestra serie de datos. Asimismo, una sobrerrepresentación de transversiones o de inserciones-delecciones podría ser un indicativo de errores en la amplificación o secuenciación.

Enmarcar las mutaciones obtenidas en un contexto filogenético es fundamental para resolver ambigüedades. Por ejemplo, en el **Capítulo V**, sobre el estudio del clado H, nos hemos encontrado con dos problemas concretos. En primer lugar, 6 muestras poseían el polimorfismo 3915, el cual fue analizado para detectar la pertenencia al sub-haplogrupo H6a. Sin embargo, estos individuos carecían de las mutaciones diagnósticas que son típicas de H6 en la región control (16362, 16482 y 239). Una búsqueda en bases de datos nos reveló que la variante 3915 también define el sub-linaje H17. En otras 6 muestras, el polimorfismo 4769 (H2a) no se encontró junto con la variante que define la rama derivada H2a5 (13708). De este modo, se adscribieron dichas secuencias al haplogrupo H18 (también definido por la posición 13708).

Para la asignación de haplogrupos y sub-haplogrupos y la búsqueda de posiciones diagnósticas empleamos la versión más actualizada del árbol filogenético humano mitocondrial (*PhyloTree*, ver sección 7). Empleamos la información de las secuencias, combinada con la obtenida por el análisis de polimorfismos de la región codificante.

Para hacer más accesible la consulta de *PhyloTree* y el establecimiento de sub-clados concretos, se ha desarrollado la herramienta *online* HaploGrep (Kloss-Brandstätter et al., 2011), que permite una clasificación automática de los perfiles de ADNmt en sub-clados muy específicos, usando la información filogenética recogida en *PhyloTree*.

En la **Figura II-20** se muestra el esquema generado por la aplicación cuando se le introduce un haplotipo. En círculos verdes se van indicando jerárquicamente los sub-clados a los que pertenece la muestra, dependiendo de los polimorfismos que posee la misma. Esta herramienta es especialmente útil en la construcción de filogenias de haplogrupos específicos usando secuencias completas de ADNmt (ver **Capítulo IV**), dado el gran volumen de información que se maneja.

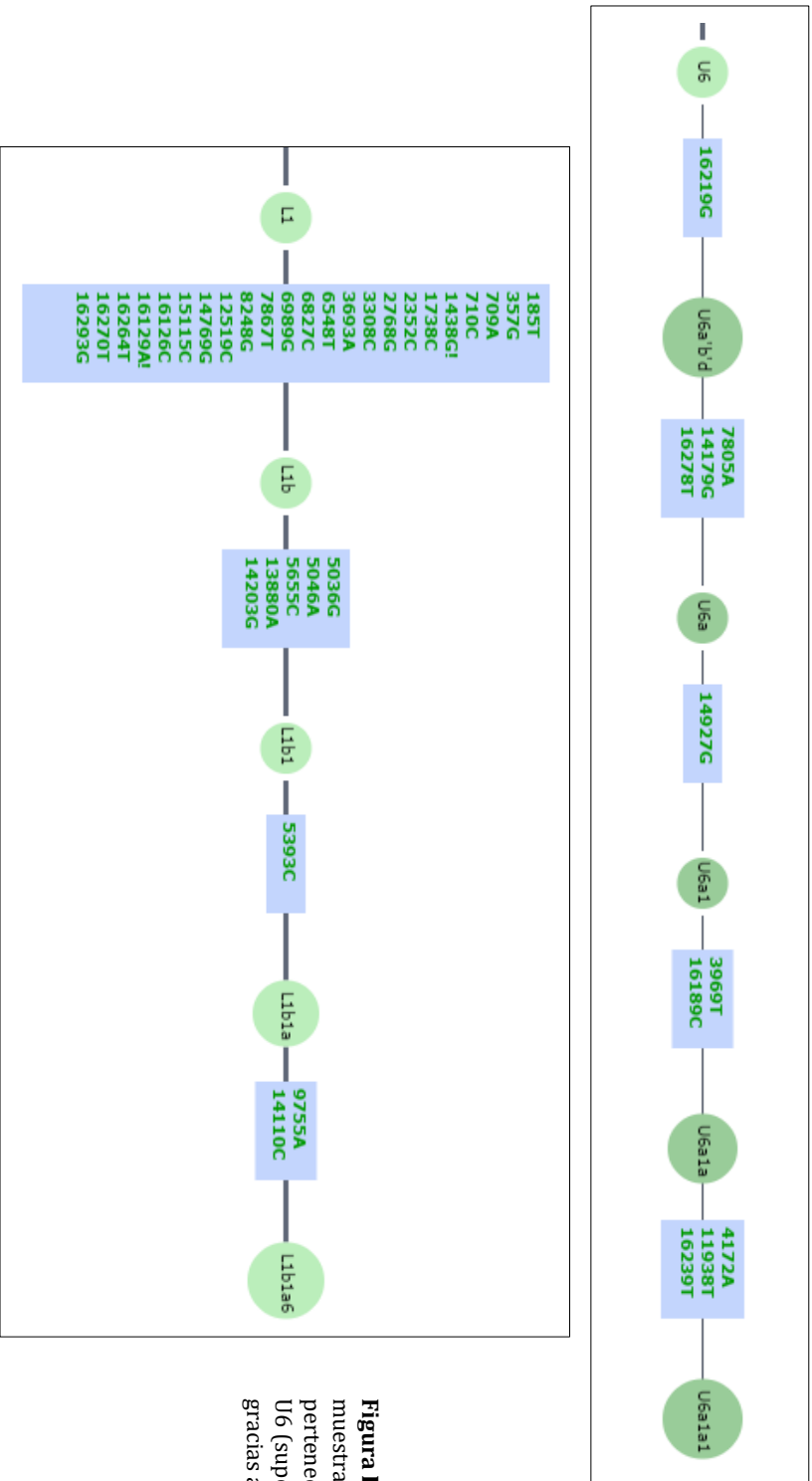


Figura II-20. Clasificación de dos muestras andaluzas pertenecientes a los haplogrupos U6 (superior) y L1b (inferior) gracias a HaploGrep.

Las secuencias mitocondriales son directamente empleadas para los análisis que se detallan a continuación.

A. Parámetros de diversidad y de historia demográfica de la población. El programa Arlequin 3.5 (Excoffier & Lischer, 2010) nos permite calcular una serie de índices que reflejan la magnitud de variación de las secuencias mitocondriales.

- Diversidad génica (*Gene Diversity*) (Nei, 1987). Se define como la probabilidad de que dos haplotipos tomados al azar de la muestra sean diferentes. Equivaldría a la heterozigosidad esperada para marcadores diploides (autosómicos).

$$\hat{H} = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$

Donde n es el número de copias del gen en la muestra, k es el número de haplotipos y p_i es la frecuencia de cada haplotipo.

- Diversidad nucleotídica (*Nucleotide Diversity*) (Nei, 1987). Es la probabilidad de que dos sitios nucleotídicos homólogos escogidos al azar sean diferentes.

$$\hat{\pi}_n = \frac{\sum_{i=1}^k \sum_{j < i} p_i p_j \hat{d}_{ij}}{L}$$

Siendo L el número de *loci*, k el número de secuencias diferentes, p_i y p_j las frecuencias de las secuencias de tipo i y j ; d_{ij} la proporción de nucleótidos diferentes entre ellas.

- Número medio de diferencias por parejas (*Mean Number of Pairwise Differences*) (Tajima, 1983). Es el número medio de diferencias que existen entre pares de haplotipos de la muestra.

$$\hat{\pi} = \frac{n}{n-1} \sum_{i=1}^k \sum_{j=1}^k p_i p_j \hat{d}_{ij}$$

Donde d_{ij} es una estima del número de mutaciones que han ocurrido desde la divergencia de los haplotipos i y j ; k es el número de haplotipos, p_i es la frecuencia del haplotipo i , n es el tamaño muestral.

El análisis de las secuencias mitocondriales nos revela también información acerca de la historia demográfica de la población bajo estudio. La distribución *mismatch* (distribución del número observado de diferencias nucleotídicas entre pares de haplotipos) dependerá de procesos de dinámica poblacional. La **Figura II-21** nos muestra la equivalencia entre las curvas de distribución *mismatch* de las secuencias de poblaciones que han tenido un tamaño demográfico constante

(en azul) o que han experimentado expansiones poblacionales a diferentes profundidades temporales (en naranja).

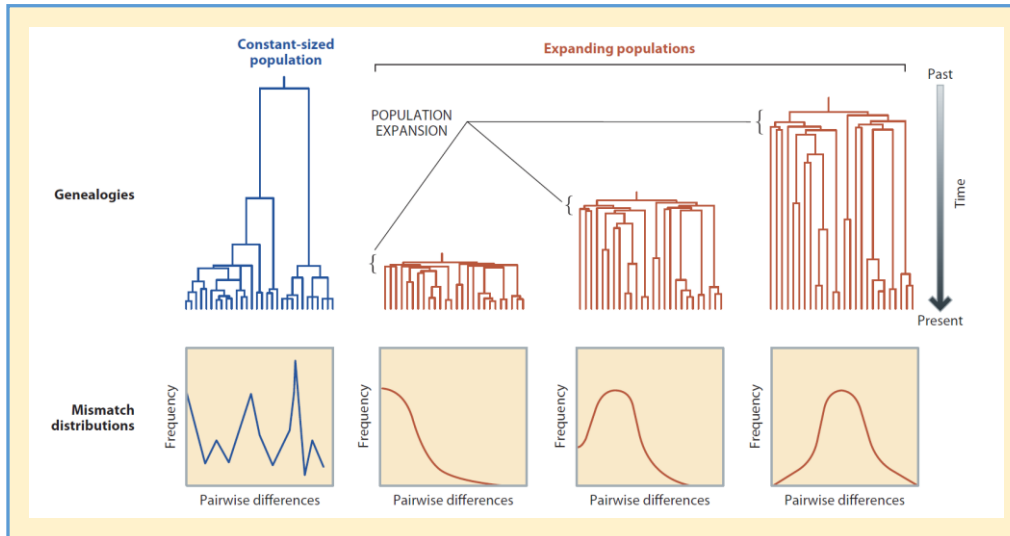


Figura II-21. Diferencias en genealogías y distribuciones *mismatch* para una población de tamaño constante (azul) y 3 poblaciones que han experimentado expansiones poblacionales a diferentes tiempos (naranja). Fuente: Jobling et al., 2013.

Para evaluar el ajuste a la normalidad de las curvas, se pueden computar dos parámetros. En primer lugar, se emplea la suma de las desviaciones al cuadrado entre el número de diferencias observadas y esperado (SSD, Schneider & Excoffier, 1999):

$$SSD = \sum_{i=0}^n (F_{i_obs} - F_{i_exp})^2$$

De manera análoga, se calcula el índice de Harpending (*Raggedness Index*, Harpending, 1994), definido como:

$$r = \sum_{i=1}^{d+1} (x_i - x_{i-1})^2$$

Donde d es el número máximo de diferencias observadas entre haplotipos y x_i es la frecuencia relativa observada del número de diferencias.

En ambos casos se trata de evaluar la coincidencia entre la distribución observada y la esperada por azar. De manera general, las distribuciones *mismatch* multimodales se suelen encontrar en poblaciones estacionarias, y las unimodales, en poblaciones en expansión.

Además, en los mismos términos se computan dos índices llamados de 'neutralidad selectiva'. El de Tajima, D (Tajima, 1989) compara dos estimas del parámetro poblacional θ , una basada en el número de sitios segregantes y la otra

en el número medio de diferencias por parejas entre haplotipos. Bajo el modelo de alelos infinitos, ambas estimas deben ser las mismas.

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_s}{\sqrt{\text{Var}(\hat{\theta}_\pi - \hat{\theta}_s)}}$$

El test de neutralidad, F_s , de Fu (Fu, 1997) sigue el mismo planteamiento que el anterior y, de nuevo, nos va a permitir evaluar si la población ha sufrido una expansión demográfica reciente.

$$F_s = \ln\left(\frac{S'}{1 - S'}\right)$$

Valores significativos de D y de F_s se interpretan normalmente como reflejo de fenómenos de expansión poblacionales.

B. Construcción de árboles filogenéticos. Una estrategia para evaluar las relaciones de parentesco entre las secuencias mitocondriales de una población dada, o de varias poblaciones entre sí, es la construcción de redes de haplotipos (*networks*). Con estas genealogías se pretende reconstruir la evolución de las secuencias de la manera más parsimoniosa posible, es decir, involucrando el menor número de pasos mutacionales.

El genoma mitocondrial está caracterizado por regiones consideradas como *hotspots*, y por tanto, hay polimorfismos que son recurrentes en el árbol mitocondrial humano independientemente del linaje. Por tanto, en el análisis de secuencias de ADNmt, principalmente cuando se utilizan mitogenomas, se eliminan las posiciones más frecuentes (inserciones en la posición 309 y 315, *indels* entre las posiciones 515-522 y las mutaciones 16182C, 16183C y 16519). Además de estas variantes, existen otras que, pese a que no se descartan de los análisis filogenéticos, se deben evaluar conforme a su ocurrencia en la filogenia. En consecuencia, a la hora de generar las redes de haplotipos con el programa Network 4.612 (<http://www.fluxus-engineering.com/>) se ponderan las mutaciones (*weighting*) según su frecuencia en la filogenia (Richards et al., 1998). La publicación de Soares y colaboradores en 2009 (Soares et al., 2009) nos proporciona un criterio para la confirmación manual de la topología de las *networks* y la resolución de posibles reticulaciones. En el mencionado trabajo se incluye un listado con los polimorfismos más frecuentemente detectados en la filogenia mitocondrial humana.

La construcción de *networks* puede llevarse a cabo empleando diversos algoritmos. En la presente Tesis Doctoral se ha empleado el método *Median-Joining* (MJ), Bandelt et al., 1999), el cual es específico para datos genéticos haploides (marcadores del ADNmt o del cromosoma Y) y emplea conjuntamente características de dos algoritmos (de Kruskal y de Farris).

Las genealogías se pueden construir para un conjunto de secuencias de una misma población o para linajes concretos empleando varias poblaciones. Además, estas metodologías se aplican a la evaluación de relaciones de secuencias control (**Capítulo III** y **Capítulo V**) o de genomas completos (**Capítulo IV**). En la **Figura II-22** se observa la reconstrucción de la filogenia de una rama derivada del haplogrupo mitocondrial K1 (Costa et al., 2013).

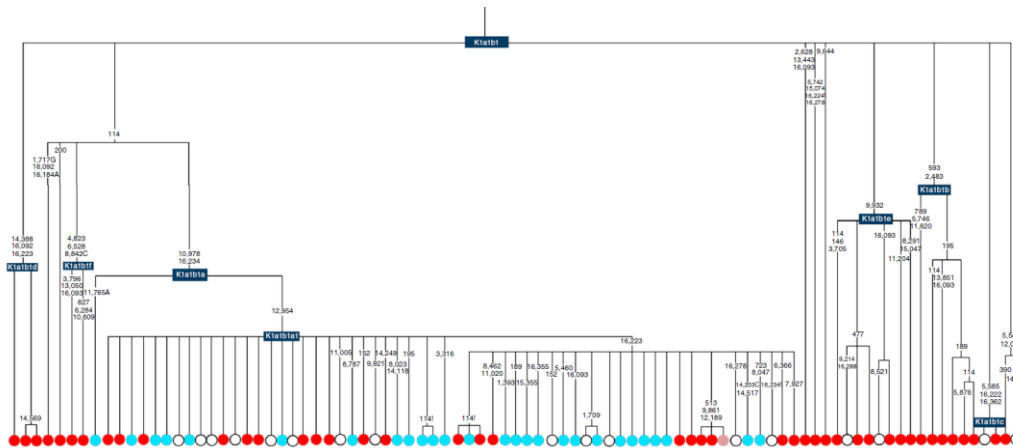


Figura II-22. Árbol filogenético del haplogrupo K1a1b1. En las ramas se indican las mutaciones, y los nodos corresponden a individuos, codificados por colores según su origen geográfico (rojo: Europa, azul: judíos, blanco: desconocido). Fuente: Costa et al., 2013.

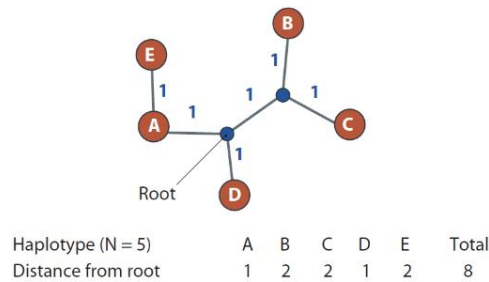
6.2. La estimación de edades de coalescencia

El cálculo de los TMRCA para los haplogrupos y sub-haplogrupos mitocondriales se ha basado en dos metodologías. En primer lugar, realizamos un cálculo directo a partir de las genealogías del estadístico ρ (*rho*) (Forster et al., 1996) definido como la divergencia media de la secuencia desde el haplotipo ancestral inferido del clado en cuestión.

La **Figura II-23** indica el procedimiento de cálculo del estadístico ρ a partir de una genealogía sencilla. Se trata de contabilizar los pasos mutacionales desde la 'raíz' o el haplotipo origen de la *network* hasta los nodos terminales. En el ejemplo gráfico, tendríamos una distancia media desde la raíz de 8 cambios mutacionales –entre la raíz y los 5 haplotipos observados– es decir, $\rho=8/5=1,6$. El valor de ρ debe ser transformado en años mediante el uso de una tasa de mutación (μ) para la obtención del TMRCA (ρ/μ).

En nuestro caso, para efectuar esta transformación, tuvimos en cuenta la tasa de mutación, μ propuesta por Soares et al. (2009) para cada partición del genoma mitocondrial. Para el cálculo de los valores de ρ , basados en secuencias de la región control exclusivamente (**Capítulos III** y **V**), hemos empleado el programa Network 4.612, y las estimas a partir de genomas mitocondriales completos (**Capítulo IV**) se han realizado gracias a un software desarrollado por el equipo de la Dra. L. Pereira (IPATIMUP, Portugal).

Figura II-23. Bases metodológicas del cálculo del estadístico ρ para estimar edades de coalescencia. Se indica la distancia desde la raíz de cada uno de los 5 nodos o haplotipos. Fuente: Jobling et al., 2013.



Las estimas de edad evolutiva basadas en ρ no incorporan ninguna asunción ni modelo evolutivo. En el **Capítulo IV** (donde analizamos genomas completos pertenecientes a linajes africanos) hemos aplicado una metodología para estimas de edades evolutivas, TMRCA, que utiliza un modelo de sustitución nucleotídica. La Máxima Verosimilitud (*Maximum Likelihood*, ML) se basa en funciones de verosimilitud (*likelihood*), las cuales se definen como la probabilidad de la ocurrencia de los datos observados, sobre la base de los parámetros del análisis. Este procedimiento tiene como ventaja que dispone de un amplio repertorio de modelos evolutivos (Yang & Rannala, 2012). En nuestro caso, utilizamos el modelo HKY85 (que permite que los 4 nucleótidos estén presentes en diferentes frecuencias y que las transiciones y transversiones ocurran a tasas también diferentes) con una distribución gamma (Γ) de la tasa de mutación por sitios, que se aproxima a una distribución discreta con 32 categorías. Para la estima de la longitud de las ramas de la filogenia se usó el programa PAML (Yang, 1997).

6.3. Análisis fundador (*Founder Analysis*, FA)

Desde el análisis fundador (Richards et al., 2000) se trata de identificar y datar los linajes que presumiblemente han surgido autóctonamente dentro de una población dada o región geográfica (*sink population*) como consecuencia de una inmigración desde una fuente externa (*source population*). Para alcanzar este objetivo, se debe distinguir la diversidad que se ha acumulado desde el momento del asentamiento respecto a la que existía antes del mismo. En la práctica, esto significa identificar secuencias 'fundadoras' (*source*) que han migrado desde la fuente hacia la población 'sumidero' (*sink*), y usar cualquier variación que ha ocurrido para datar cada una de ellas. Las secuencias coincidentes entre las poblaciones fuente/sumidero serían candidatas a ser consideradas como fundadoras. Sin embargo, los supuestos que se incorporan en esta metodología pueden ser equívocos, ya que las migraciones en el sentido opuesto (recíproco) al considerado también generarían identidad de secuencia (Macauley & Richards, 2013).

La **Figura II-24** muestra de manera gráfica los principios del análisis fundador, FA. La población sumidero proviene de 3 individuos que migraron desde la población fuente (A). La genealogía de la población original es profunda,

mientras que la de la población derivada consiste en 3 *clusters* de individuos emparentados muy próximos entre sí. El análisis pretende detectar las secuencias que migraron (marcadas con el símbolo +) y emplear las mutaciones acumuladas, a partir de estos tipos fundadores, como un límite inferior para datar el evento de flujo génico (B). De esta manera, no se tienen en cuenta las mutaciones que ocurren previamente al episodio migratorio (Figura II-24A, líneas punteadas rojas).

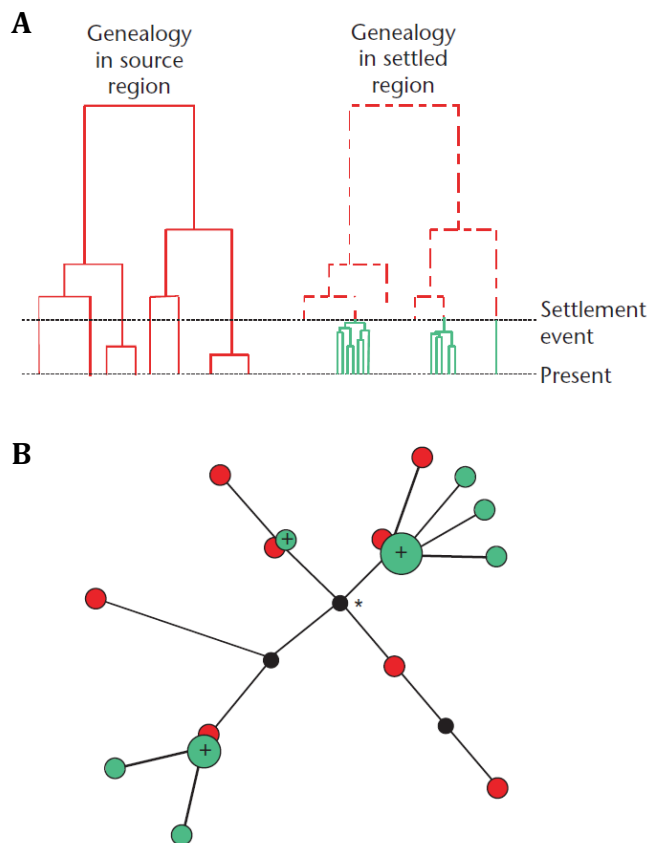


Figura II-24. Representación gráfica de las bases del análisis fundador (FA). **A:** Genealogías en la región fuente y región sumidero de las secuencias. **B:** Red de haplotipos entre las secuencias de ambas regiones. Las secuencias de la población fuente se muestran en rojo, y las de la población sumidero, en verde. El nodo ancestral común se indica con una estrella y los tipos fundadores candidatos con el símbolo +. Fuente: Macaulay & Richards, 2013.

Para identificar una secuencia fundadora se requiere que ésta tenga cierto grado de diversificación en la población fuente que indique una profundidad temporal mínima de residencia. Así, se establecen dos criterios, $f1$ si la secuencia presenta al menos una rama derivada o $f2$ si muestra al menos dos. El procedimiento metodológico comienza con la construcción de una *network* para la región control (rango 16051-16400). Posteriormente, se identifican los tipos fundadores y descendientes, utilizando una herramienta desarrollada por el equipo de L. Pereira (Alves et al., 2012). Por último, se estima la edad de migración de cada fundador usando de nuevo el estadístico ρ y transformando sus valores en “años” usando la tasa de mutación previamente indicada (para este rango nucleotídico, de 1 mutación cada 16.677 años, Soares et al., 2009). En

el **Capítulo IV** se exponen más detalles de esta metodología y su aplicación concreta a nuestros análisis.

6.4. La estimación de tamaños efectivos (N_e) poblacionales

La Inferencia Bayesiana (BI) es un método general de deducción estadística en la que los parámetros del modelo se consideran variables aleatorias con distribuciones estadísticas. Para otras metodologías, sin embargo, como la ML, los parámetros son constantes fijas desconocidas. Previamente al análisis de los datos, se establecen distribuciones '*prior*' para los parámetros, los cuales se combinan con los datos para generar distribuciones posteriores (teorema de Bayes: la probabilidad Bayesiana posterior es proporcional a los *priors*, más la información de los datos). La probabilidad posterior de los árboles no es calculada directamente sino que se emplean los algoritmos MCMC para generar este parámetro (Yang & Rannala, 2012).

Los *Bayesian Skyline Plots* (BSPs) se emplean para modelizar el tamaño poblacional frente a un amplio rango de escenarios demográficos. Como procedimientos de muestreo, emplea métodos de cadenas de Markov Monte Carlo (*Markov Chain Monte Carlo*, MCMC) para estimar una distribución posterior de tamaño efectivo poblacional (N_e) a través del tiempo, y directamente, a partir de una muestra de secuencias. La ventaja de esta metodología radica en la inclusión de intervalos de credibilidad para este N_e en cualquier punto de la escala temporal hasta el MRCA de las secuencias empleadas (Drummond et al., 2005).

Para la implementación de esta metodología se utilizó el programa BEAST 1.4.6 (Drummond & Rambaut, 2007). En concreto, se efectuaron 5×10^7 iteraciones, con muestras dibujadas cada 1.000 pasos MCMC tras un *burn-in* descartado de 5 millones de pasos. Se empleó un tiempo generacional de 25 años para re-escalar el eje vertical del BSP.

6.5. Análisis de estructura poblacional

La evaluación de estructuras poblacionales se llevó a cabo usando el Análisis Molecular de la Varianza (*Analysis of Molecular Variance*, AMOVA) (Excoffier et al., 1992).

Para implementar este método se asocian las poblaciones en grupos geográficos entre los que se pretende probar una posible estructura genética. Posteriormente, se realiza un análisis jerárquico de las particiones de la varianza total (σ^2) en componentes de covarianza debidos a diferencias entre haplotipos dentro de una población (σ_c^2), entre haplotipos de diferentes poblaciones dentro de un grupo geográfico (σ_b^2) y entre los grupos geográficos (σ_a^2). Estos componentes se emplean para calcular los conocidos índices de fijación (Wright, 1965; Weir & Cockerham, 1984): (1) dentro de una población (F_{ST}), (2) entre poblaciones (F_{SC}) y (3) entre grupos de poblaciones (F_{CT}).

Por otro lado, las comparaciones entre poblaciones por pares se calcularon empleando el índice *pairwise* (por parejas) F_{ST} . Estas comparaciones se pueden realizar a nivel de las secuencias de ADNmt o bien a partir de la sub-estructuración de la población en diferentes haplogrupos.

$$F_{ST} = \frac{V_p}{p(1-p)}$$

Donde p y V_p son la media y la varianza de las frecuencias génicas entre las dos poblaciones, respectivamente. A la hora de interpretar el índice F_{ST} , se considera que refleja una diferenciación genética baja si es $<0,05$; moderada entre $0,05$ y $0,15$; alta entre $0,15$ y $0,25$ y muy alta si es $>0,25$.

Estos análisis de estructura génica se realizaron con el programa Arlequin 3.5 (Excoffier & Lischer, 2010).

Las diferencias entre los perfiles de linajes mitocondriales entre pares de poblaciones pueden ser estimadas también gracias a un test χ^2 y al estudio de los residuos corregidos tipificados. Los valores de estos residuos nos indican cuales son los haplogrupos que estadísticamente determinan las diferencias entre un par de poblaciones. Por tanto, se construyen tablas de contingencia y se evalúan aquellos sub-conjuntos cuyas proporciones de columna difieran significativamente entre sí. Este test estadístico se llevó a cabo en el programa SPSS 19 (IBM Corp., Armonk, NY).

6.6. Análisis multivariantes

Una manera gráfica de observar las diferencias de perfiles genéticos asociados a diversas poblaciones, es la de representar a las mismas en un espacio bidimensional mediante un 'análisis multivariante', un método estadístico que extrae información a partir de datos multi-dimensionales (Jobling et al., 2013).

Esta información puede ser introducida en el análisis mediante distancias genéticas (por ejemplo, para la construcción de un *Multi-dimensional scaling*, MDS) o directamente a partir de los datos de frecuencias alélicas de las poblaciones. En la presente Tesis Doctoral hemos seleccionado el último procedimiento, utilizando las frecuencias de haplogrupos de cada población, es decir, la estructura de linajes maternos.

Para la construcción de mapas genéticos hemos empleado el Análisis Jerárquico de Cluster (*Hierarchical Cluster Analysis*, HCA) mediante el programa SPAD (*Système Portable pour l'Analyse des Données*, Lebart et al., 1984). Esta metodología se emplea habitualmente para evaluar relaciones entre poblaciones debido a que permite un elevado nivel de resolución estadística. Los *clusters* se generan de manera jerárquica y progresiva en función de la distancia entre las poblaciones (matriz de distancias Euclídeas). Para calcular la distancia entre

clusters, minimizando la suma de los cuadrados dentro de cada grupo, se emplea el algoritmo de Ward.

La información de salida (*output*) del programa es muy completa e informativa. Se obtiene: (1) el *plot* propiamente dicho, con las poblaciones indicadas como puntos en el espacio bidimensional (y asociadas estadísticamente a un *cluster*), (2) los haplogrupos expresados como vectores que condicionan la topología de las poblaciones y, (3) los porcentajes de absorción de la varianza de los datos de cada uno de los dos factores representados. Además, podemos analizar parámetros complejos que nos ayudan a obtener la máxima información de los datos de entrada. El análisis nos indica, también, cuales son los linajes que estadísticamente definen la distribución del *plot* en general y de cada uno de los *clusters* determinados. Por último, se indica qué población o poblaciones son las más características de cada *cluster* (es decir, las que menor distancia al centro del *cluster* registran). Todo ello hace que el HCA sea una buena elección a la hora de representar datos genéticos y de estructura de poblaciones humanas.

Para esta metodología, como para otras técnicas, es crucial seleccionar la escala de definición de los haplogrupos. Habitualmente, el grado de resolución filogenética usado en nuestros análisis no es equiparable con el alcanzado en la mayoría de las poblaciones utilizadas con fines comparativos. Por ello, en un alto número de casos hay que combinar diferentes haplogrupos en clados parentales que engloben más diversidad y que permitan la comparación interpoblacional.

6.7. La evaluación de la correlación entre geografía y variantes genéticas

Para analizar la filogeografía de los linajes mitocondriales se empleó en primer lugar la representación espacial de los mismos en mapas conocidos como de contornos, de superficies o de interpolación.

Los procedimientos de interpolación buscan generar valores para los puntos intermedios entre dos observaciones o sitios en los que el valor real del parámetro es conocido (**Figura II-25**). El análisis parte de la definición mediante puntos geográficos (XY, de coordenadas) de las localidades o poblaciones de las cuales tenemos información genética. El dato de frecuencia alélica de los puntos es empleado para estimar o predecir las frecuencias en los sitios para los que se carece de dicha información. Las líneas de contorno unen puntos con la misma frecuencia (en nuestro caso, del linaje mitocondrial) y se genera un gradiente de intensidad de color que muestra de manera gráfica los picos de frecuencias y las tendencias generales de la variable dentro de un área geográfica. Existen numerosos métodos de interpolación, en nuestro caso, hemos empleado *Inverse Distance Weighted* (IDW), que asume el hecho de que la influencia de la variable que se quiere interpretar decrece con la distancia desde los puntos geográficos representados. Los mapas se construyeron con la extensión *Spatial Analyst* del programa ArcGIS 10.1.

Para la representación de los puntos geográficos se utilizaron las coordenadas o localización específica de las poblaciones en el caso de que el estudio incluyese

esa información. Si no fuese así, se emplearon las capitales administrativas del estado o región. En ocasiones, fue necesario combinar la información de varias poblaciones con diferentes localizaciones para crear un punto único, lo que conseguimos empleando la opción *Mean Center* de la plataforma ArcGIS.

Se debe prestar atención aquí a una cuidada selección de las poblaciones a utilizar, evitando, por ejemplo, poblaciones muy aisladas o de muy bajo tamaño muestral, cuya inclusión provocará probablemente patrones confusos en los que se podría magnificar el efecto de la deriva genética, u otros factores de diferente naturaleza.

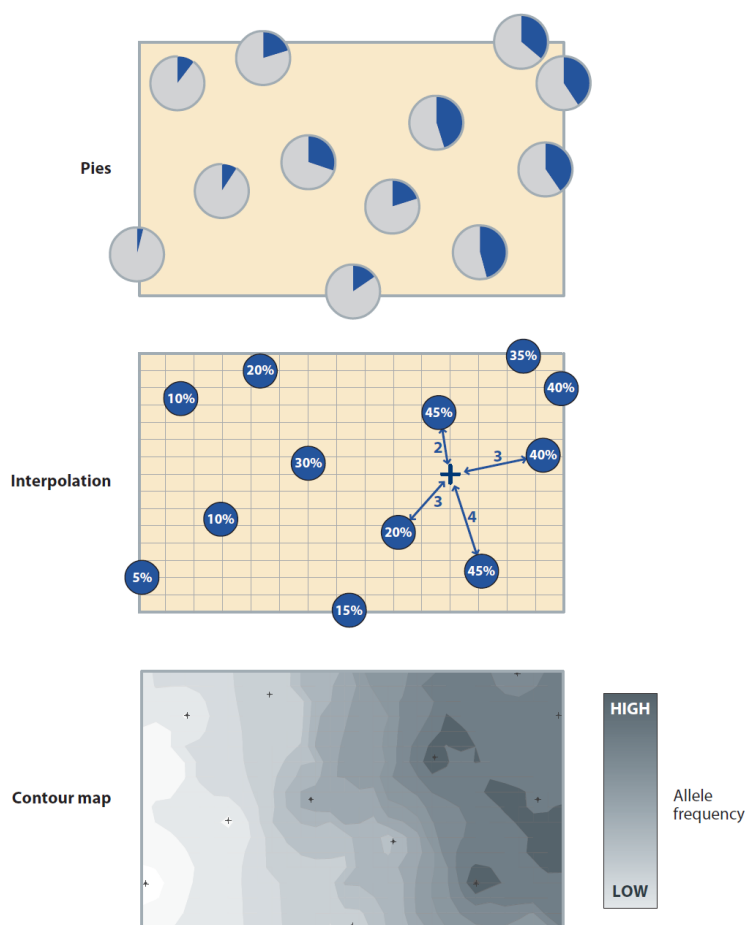
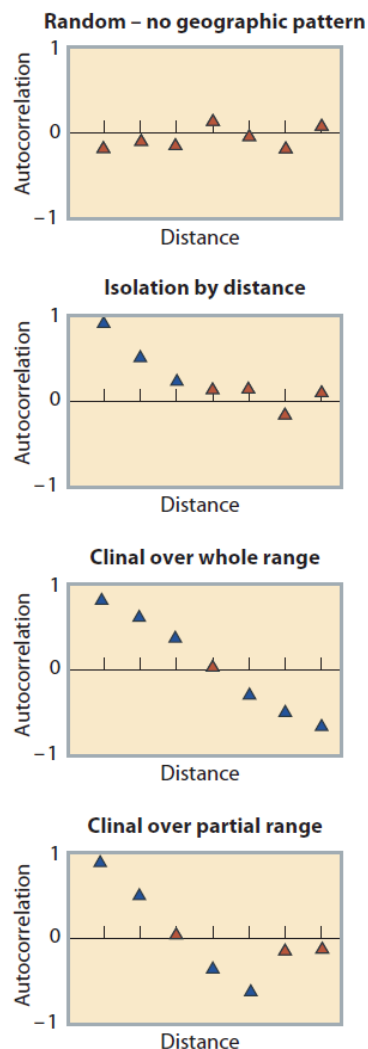


Figura II-25. Representación del proceso de interpolación a partir de frecuencias alélicas. Fuente: Jobling et al., 2013.

La interpretación de estos mapas de frecuencias génicas debe hacerse de manera prudente, ya que se trata de un procedimiento principalmente descriptivo. Además, el proceso mismo de la interpolación puede crear fenómenos como variaciones clinales o gradientes artificiales (Jobling et al., 2013). En este punto, se puede ir un paso más lejos y cuantificar la correlación entre genes y geografía. Una metodología ampliamente utilizada es la

representación del nivel de autocorrelación espacial respecto a la distancia geográfica. En primer lugar, se muestrean pares de localidades en diferentes clases de distancias y para cada una se calcula el nivel de autocorrelación usando una medida de similitud genética que varía entre 1 (correlación fuerte positiva) y -1 (correlación fuerte negativa). Esta correlación se estima estadísticamente para cada clase de distancia. La representación de esos valores frente a la distancia genera un '*correlograma*' y la forma de la curva resultante nos permite realizar inferencias acerca de estructura geográfica (Legendre & Fortin, 1989). La **Figura II-26** muestra los perfiles de los correlogramas esperados al azar, bajo aislamiento por distancia o variación clinal (total o parcial).

Figura II-26. Correlogramas de autocorrelación espacial bajo diversos supuestos. Los triángulos en azul indican un *P*-valor significativo para esa clase de distancia, mientras que los triángulos naranjas no revelan significación. Fuente: Jobling et al., 2013.



En la presente Tesis Doctoral, hemos empleado los análisis de autocorrelación espacial en el estudio del linaje mitocondrial H como un todo y de sus sub-clados,

algunos de los cuales son paradigmas de la variación clinal en Europa (**Capítulo V**). Para ello, hemos empleado el programa PASSaGE v. 2 (Rosenberg & Anderson, 2011). Como medida de correlación se usó el índice I de Moran en 10 clases de distancia geográfica.

7. Uso de bases de datos y selección de poblaciones para análisis comparativos

Finalmente, queremos destacar el papel central que tienen las bases de datos internacionales en la estandarización de la notación científica de las variantes genéticas humanas, así como en la difusión y accesibilidad de esta información.

Para la presente Tesis Doctoral, además de la consulta de publicaciones científicas, se han empleado regularmente páginas *web* oficiales y repositorios de secuencias mitocondriales. Entre estas plataformas, destaca *PhyloTree* (van Oven & Kayser, 2009) (<http://www.phylotree.org/>) como herramienta fundamental para abordar el estudio del ADNmt. Se trata de la visión más refinada del árbol mitocondrial humano, construido a partir de los mitogenomas publicados hasta la fecha. Se actualiza periódicamente para incluir nuevas variantes derivadas de las publicaciones, y en su última versión (19 de febrero de 2014) incluye información derivada de 20.666 genomas mitocondriales completos. La notación de haplogrupos y sub-haplogrupos y sus polimorfismos diagnósticos se encuentra, por tanto, actualizada, lo que evita problemas de ambigüedad en la nomenclatura en estudios de ADNmt humano.

Por otro lado, la *web* MITOMAP (<http://www.mitomap.org>) es un compendio de las mutaciones del ADNmt humano, tanto de la región control como de la región codificante. Esta base de datos de referencia contiene también información interesante acerca de polimorfismos mitocondriales asociados a diversas patologías.

Las secuencias mitocondriales se extraen en su mayor parte de GenBank (*National Center for Biotechnology Information*, NCBI). Para contribuir a la máxima difusión y al enriquecimiento acerca del conocimiento filogenético humano, toda la información de secuencias generada en la presente Tesis (502 secuencias de la región control y 62 mitogenomas) ha sido o será depositada en dicha base de datos.

Para finalizar, queremos hacer una reflexión acerca de la selección de poblaciones con propósitos comparativos. No solo resulta crucial el proceso de muestreo poblacional que se realiza en estudios de diversidad genética humana, sino que la contextualización de los datos obtenidos con respecto a otras poblaciones humanas va a permitir de nuevo que las inferencias evolutivas tengan un soporte sólido y robusto. Se trata de un esfuerzo notable porque supone la revisión y filtrado de numerosa información. En los 3 Capítulos de Resultados de esta Tesis (**Capítulo III**, **Capítulo IV** y **Capítulo V**) se

seleccionaron cuidadosamente estas poblaciones teniendo en cuenta una serie de criterios. En primer lugar, el tamaño muestral de dichas poblaciones debía superar los 45 individuos. En el proceso de filtrado es importante localizar a la población geográficamente de la manera más detallada posible, lo que es útil en representaciones espaciales de mapas, como se ha comentado anteriormente. Asimismo, una dificultad que puede surgir se refiere al nivel de resolución de los estudios lo que ocasiona que se deban descartar algunas poblaciones por falta de detalles filogenéticos.

Capítulo III

Herencia materna en Andalucía: estructura y diversidad

Este capítulo está basado en la siguiente publicación: **Hernández C.L.**, Reales G., Dugoujon J.M., Novelletto A., Rodríguez J.N., Cuesta P., Calderón R. (2014) Human maternal heritage in Andalusia (Spain): its composition reveals high internal complexity and distinctive influences of mtDNA haplogroups U6 and L in the western and eastern side of region. *BMC Genet* 15:11. doi: 10.1186/1471-2156-15-11.

La arqueología e historia del Mediterráneo muestra que este mar ha sido un obstáculo permeable para las migraciones humanas. Este fenómeno ha tenido como consecuencia múltiples intercambios culturales a lo largo y ancho del Mediterráneo y subsecuentes procesos de mezcla poblacional. Un importante territorio para esas migraciones es la Península Ibérica. En el presente estudio se ha llevado a cabo un análisis exhaustivo del *pool* génico materno –por medio de secuenciación de la región control mitocondrial y genotipado mediante PCR-RFLPs– de individuos autóctonos andaluces y con origen en las provincias costeras de Huelva y Granada localizadas, respectivamente, en el oeste y en el este de Andalucía.

La composición de haplogrupos de estas dos poblaciones de la España meridional revela un amplio espectro de linajes con orígenes en diversas áreas continentales. Las frecuencias de marcadores eurasiáticos, junto con una alta incidencia y diversificación de linajes maternos africanos, que alcanzan el 15% de la variabilidad mitocondrial total en Huelva, en comparación con lo detectado en la población oriental de Granada y en otras regiones de Iberia, representan resultados inéditos acerca de las características del ADN mitocondrial en Andalucía y testifican, además, una subestructura poblacional desde el punto de vista materno. Por ello, Andalucía no puede ser considerada como un territorio genéticamente homogéneo.

El legado materno de los andaluces refleja historias locales distintivas, apuntando a la región más occidental de la España peninsular como un relevante receptor de diversas y múltiples migraciones humanas. Los resultados obtenidos refuerzan la necesidad de investigaciones más detalladas acerca de las relaciones genéticas a ambos lados del oeste del Mediterráneo, usando muestras autóctonas. Muchos estudios se han centrado en el flujo génico reciente desde el Norte de África a Iberia, por lo que la atención debe ser ahora dirigida al estudio en profundidad de la introducción de genes europeos en el noroeste africano a través del mar, con el fin de determinar su magnitud y escala temporal para compararlos con aquellos movimientos terrestres desde el este de África y el sudoeste asiático.

The archeology and history of the ancient Mediterranean have shown that this sea has been a permeable obstacle to human migration. Multiple cultural exchanges around the Mediterranean have taken place with presumably population admixtures. A gravitational territory of those migrations has been the Iberian Peninsula. Here we present a comprehensive analysis of the maternal gene pool, by means of control region sequencing and PCR-RFLP typing, of autochthonous Andalusians originating from the coastal provinces of Huelva and Granada, located respectively in the west and the east of the region.

The mtDNA haplogroup composition of these two southern Spanish populations has revealed a wide spectrum of haplogroups from different geographical origins.

Resumen

Abstract

The registered frequencies of Eurasian markers, together with the high incidence and diversification of African maternal lineages (15% of the total mitochondrial variability) among Huelva Andalusians when compared to its eastwards relatives of Granada and other Iberian populations, constitute relevant findings unknown up-to-date on the characteristics of mtDNA within Andalusia that testifies a female population substructure. Therefore, Andalusia must not be considered a single, unique population.

The maternal legacy among Andalusians reflects distinctive local histories, pointing out the role of the westernmost territory of Peninsular Spain as a noticeable recipient of multiple and diverse human migrations. The obtained results underline the necessity of further research on genetic relationships in both sides of the western Mediterranean, using carefully collected samples from autochthonous individuals. Many studies have focused on recent North African gene flow towards Iberia, yet scientific attention should be now directed to thoroughly study the introduction of European genes in northwest Africa across the sea, in order to determine its magnitude, timescale and methods, and to compare them to those terrestrial movements from eastern Africa and southwestern Asia.

1. Introducción

La arqueología e historia del Mediterráneo muestra que este mar ha sido siempre un obstáculo permeable para las migraciones humanas. Desde tiempos prehistóricos, el hombre ha viajado a través del mar Mediterráneo y se ha movido en numerosas direcciones. Algunas evidencias de antiguos movimientos humanos en esta área geográfica, tanto por mar como por tierra, se encuentran en el comercio de obsidiana de la isla de Milo con Grecia continental (~8.000 BC) o en la ocupación de las islas de Chipre, Córcega, Cerdeña o Mallorca (~8.500 BC) (Knapp, 2010; Demand, 2011). Por tanto, múltiples contactos culturales se han sucedido alrededor del Mediterráneo con presumibles mezclas poblacionales. Una región crítica para estas migraciones ha sido la Península Ibérica. Iberia posee una situación geográfica estratégica, en el extremo sudoccidental de Europa, hallándose en contacto con dos mares, el Océano Atlántico y el mar Mediterráneo y, actuando como el nexo de unión más cercano entre África y Europa, a través del Estrecho de Gibraltar (14 km). Además, su extenso territorio de orografía compleja, con gran variación marcada entre sus regiones (Cohen & Small, 1998), su intrincada historia, riqueza arqueológica y sus diversos y persistentes patrones socioculturales, son factores que justifican el creciente interés científico en conocer la diversidad genética de las poblaciones contemporáneas ibéricas y el impacto de las migraciones en su *pool* génico (Alonso et al., 2005; Pereira et al., 2005b; Calderón et al., 2006; Álvarez et al., 2010; Ambrosio et al., 2010a, 2010b; Cerezo et al., 2012). La franja cantábrica de Iberia, junto con el este de Andalucía y el Levante habrían tenido un papel relevante como áreas refugio durante el último máximo glacial (*Last Glacial*

Maximum, LGM, hace ~20.000 años) además de haber actuado como fuentes para el posterior repoblamiento de Europa cuando las condiciones climáticas mejoraron (Gamble et al., 2004; Bocquet-Appel et al., 2005).

El genoma mitocondrial (ADNmt) es una de las herramientas moleculares más ampliamente utilizadas en el estudio de la diversidad genética de Iberia, y comúnmente mediante el uso conjunto de la región control y de ciertas variantes de la región codificante. En los últimos años se está alcanzando el máximo nivel de resolución filogenética en el análisis de la herencia materna de las poblaciones ibéricas, gracias a la secuenciación completa del ADNmt. Estos estudios de mitogenomas, habitualmente, se centran en linajes maternos concretos con raíces ancestrales en diversas áreas continentales (Maca-Meyer et al., 2003b; Achilli et al., 2004; Pereira et al., 2010a; Cerezo et al., 2012).

Dentro de la España peninsular, la región más ampliamente analizada desde el punto de vista del ADNmt ha sido, sin duda, la Cornisa Cantábrica –en particular Galicia y el área vasca (Salas et al., 1998; Álvarez-Iglesias et al., 2009; García et al., 2011; Behar et al., 2012a). Ambas poblaciones comparten valores elevados del linaje mitocondrial H y, específicamente, de H1 y H3, alcanzando estos últimos picos de frecuencias (valores modales) en el norte ibérico.

Dada la abundancia de estudios acerca de la variación genética de la población de Iberia peninsular y de sus islas (sobre todo las Canarias), resulta llamativa la ausencia de un análisis monográfico y exhaustivo acerca de los perfiles mitocondriales de poblaciones andaluzas definidas. La atractiva historia y prehistoria de Andalucía –la región más amplia y poblada de España– convierte a su población autóctona actual en un objetivo preeminente para la detección de escenarios de subestructura poblacional así como para la evaluación del esperado impacto de poblaciones africanas y mediterráneas en el *pool* génico de Iberia. Dada la proximidad geográfica entre el sur de España y África, estudios previos han mostrado evidencias de linajes mitocondriales asociados con el continente africano entre la población andaluza y altos niveles de diversidad tanto en el análisis de muestras regionales generales (Côrte-Real et al., 1996; Larruga et al., 2001; Plaza et al., 2003) como en territorios concretos de la región (Casas et al., 2006).

En el presente trabajo se proporcionan datos novedosos acerca de la composición de haplogrupos del ADNmt en individuos autóctonos de las provincias de Huelva y Granada, localizadas respectivamente en el oeste y en el este de la región de Andalucía (ver **Figura II-10**). Más detalles acerca de las características demográficas e históricas de ambas provincias se pueden encontrar en el **Capítulo II**, secciones **1.2** y **1.3**.

Algunos estudios recientes centrados en el análisis en alta resolución filogenética de los linajes paternos E y J desarrollados por nuestro equipo de investigación en las mismas muestras andaluzas, han revelado que tanto Huelva como Granada han integrado múltiples migraciones, y que las fuentes de flujo génico parecen ser más intensas en el extremo occidental de Andalucía. La presencia del E-M81, comúnmente conocido como el “marcador bereber”, la

ocurrencia del E-M34, prevalente entre población judía y la apreciable representación de otros linajes paternos, tales como el E-V13, J1-M267 y J2b-M12, apoyarían el papel de Andalucía como una puerta abierta a los movimientos humanos poblacionales, sobre todo, a través del Mediterráneo y con una especial intensidad desde el Neolítico hasta tiempos históricos (Ambrosio et al., 2010a, 2010b).

Basándonos en estos datos previamente publicados acerca de polimorfismos de ADN haploides en poblaciones ibéricas actuales, el principal objetivo del presente trabajo es la evaluación de la composición de linajes mitocondriales de la población autóctona andaluza. Hasta la fecha no se han llevado a cabo estudios centrados en estos territorios o poblaciones. Los datos presentados aquí proporcionan fuertes evidencias de una influencia africana más visible en el oeste que en el este de Andalucía. Además, se analizan *networks* filogenéticos basados en los haplogrupos del ADNmt U6 y L en muestras andaluzas, ibéricas y del noroeste africano y se interpretan otras notables características presentes en la composición materna de los actuales habitantes del sur de Iberia. Todos estos nuevos datos nos permiten desvelar nuevas claves acerca de los procesos de poblamiento ocurridos en la Península Ibérica desde tiempos prehistóricos hasta la actualidad.

2. Material y Métodos

2.1. Poblaciones y muestras

Hemos analizado 279 secuencias de ADNmt de individuos sanos, sin emparentar y de ambos sexos con orígenes geográficos en las provincias de Huelva (n= 158) y Granada (n= 121). El proceso de muestreo poblacional se encuentra detallado en la sección II-2.

2.2. Análisis moleculares de ADNmt

Para cada muestra se amplificó la región hipervariable I del ADNmt (HVS-I) y parte de la HVS-II (~820 pb) (ver sección II-5.1). Además de la secuenciación de la región control, se analizó un conjunto de SNPs pertenecientes a la región codificante con el fin de realizar una asignación inequívoca a haplogrupos mitocondriales (ver **Tabla II-3**).

Dos individuos de Granada no pudieron ser clasificados en ningún linaje concreto con la información genética obtenida. Por tanto, fueron categorizados como NR*, ya que solo se pudo confirmar la presencia del polimorfismo 10873.

Las 279 secuencias obtenidas en el presente estudio se depositaron en la base de datos GenBank (*Accession Numbers* KJ169731 - KJ170009).

2.3. Análisis estadísticos y filogenéticos

Las secuencias control mitocondriales se procesaron siguiendo los pasos descritos en la sección II-6.1. Se calcularon parámetros de diversidad genética intrapoblacional, índices de neutralidad selectiva (*Tajima* y *Fu*) y se evaluó la distribución normal de las diferencias por pares de secuencias (*mismatch distributions*) (índice de Harpending y SSD). Se calculó el índice F_{ST} entre las dos poblaciones andaluzas (usando tanto información de secuencias como aquella derivada de haplogrupos) y su significación fue evaluada con un test no paramétrico de permutación. Todos estos análisis se realizaron con el programa Arlequin 3.5 (Excoffier & Lischer, 2010). Además, se realizó un test χ^2 y se analizaron sus residuos tipificados corregidos para valorar las diferencias entre los perfiles de haplogrupos y su peso en el análisis (IBM SPSS 19).

Se construyeron genealogías maternas, utilizando el método de *Median-Joining Networks (MJN)*, tanto para el total de los haplotipos encontrados en Huelva y Granada como para los linajes africanos U6 y L. Se calcularon los valores de diversidad interna para el haplogrupo africano U6 empleando fuentes de información obtenidas de la literatura (Cherni et al., 2005; Coudray et al., 2009). Además, se estimaron las edades de coalescencia de algunos clados africanos seleccionados mediante el estadístico ρ (ver detalles en la sección II-6.2).

Para llevar a cabo el análisis de variación molecular, AMOVA (sección II-6.5) se construyó una base de datos de 53 poblaciones, seleccionadas de la literatura (ver **Anexo I, Tabla S1**). La misma base de datos se empleó para evaluar las relaciones de parentesco genéticas entre poblaciones mediante un Análisis Jerárquico de Cluster (HCA, sección II-6.6).

Finalmente, se representó la variación geográfica del linaje U6 en el espacio mediterráneo en un mapa de interpolación de superficies (ver sección II-6.7) con ArcGIS 10.1. Se usó un valor de 2 en la variable 'power'.

3. Resultados

3.1. Diversidad de las secuencias de ADNmt en Andalucía

Entre los 279 individuos analizados se encontró un total de 197 haplotipos diferentes, usando conjuntamente las secuencias de la región control y los marcadores –SNPs– de la región codificante (los haplotipos de los 121 individuos pertenecientes a Granada se presentan en la **Tabla S2**). Una fracción reducida de estos haplotipos (11/279) fue compartida en las dos poblaciones andaluzas. Teniendo en cuenta solamente la información de la región control (HVS), se definieron 104 secuencias diferentes en las 158 muestras de Huelva (65,8%) y 95 en los 121 andaluces de Granada (78,5%). El número de haplotipos *singleton* (únicos) fueron 77, 80 y 147 en Huelva, Granada y ambas provincias, respectivamente. En los tres casos, la distribución de haplotipos repetidos

(abundancia-ocurrencia) se ajusta a una ley potencial con exponentes que varían entre -2,12 y -3,03 y coeficientes de determinación (R^2) de 0,86 y 0,91. Los parámetros derivados del análisis de secuencias de ADNmt se presentan en la **Tabla III-1**.

Tabla III-1. Parámetros estadísticos de las secuencias de ADNmt en Andalucía.

	Andalucía occidental (Huelva)	Andalucía oriental (Granada)
n	158	121
K (%)	104 (65,82%)	95 (78,51%)
P	114	114
H (SD)	0,9856 ($\pm 0,0046$)	0,9917 ($\pm 0,0035$)
D (SD)	0,0113 ($\pm 0,0058$)	0,0090 ($\pm 0,0047$)
M (SD)	9,1776 ($\pm 4,2431$)	7,3161 ($\pm 3,4479$)
SSD	0,0017 (ns)	0,0013 (ns)
r	0,0028 (ns)	0,0033 (ns)
D	-1,7122*	-2,1126**
F _s	-24,3904***	-24,7983***

K: número de secuencias diferentes (%); P: número de sitios polimórficos; H: diversidad haplotípica (desviación estándar); D: diversidad nucleotídica; M: número medio de diferencias por parejas; SSD: suma de las desviaciones al cuadrado entre distribución *mismatch* observada y esperada; r: índice de Harpending; D y F_s son, respectivamente, los test de Tajima y Fu de neutralidad selectiva. ns: no-significativo; *p < 0,05; **p < 0,01; ***p < 0,001

Los valores estimados de diversidad génica (H) son similares en Huelva y en Granada, pero ligeramente superiores a lo encontrado en otras poblaciones peninsulares de Iberia (0,95-0,96) (Casas et al., 2006; Álvarez et al., 2010). Los test de neutralidad de Tajima (D) y Fu (F_s) muestran, en ambos casos, valores negativos y significativos. Estos resultados, junto con el patrón unimodal observado para las distribuciones *mismatch* (ver **Figura III-1**) nos conduce a inferir presumibles episodios de expansiones poblacionales recientes. El valor no-significativo del índice de Harpending (r) y de la suma de desviaciones al cuadrado entre la distribución *mismatch* observada y esperada apoyarían también este resultado.

La composición de haplogrupos mitocondriales en Andalucía (**Tabla III-2**) revela un perfil típicamente del oeste europeo, alcanzando los haplogrupos eurasiáticos (R0, HV, H, J, T, U sin incluir U6, K, N1, N2 y X) un 85,4% en el oeste (Huelva) y un 96,7% en el este (Granada). Sin embargo, se detectó una significativa diferenciación genética en la comparación de las secuencias de la región control (F_{ST}= 0,011; P= 0,000) así como en la composición de haplogrupos global (F_{ST}= 0,025; P= 0,000). El valor del test χ^2 apoya estas observaciones ($\chi^2= 41,54$; g.l.= 13; P= 0,000). El análisis de los residuos tipificados corregidos reveló

que los haplogrupos que significativamente contribuyen a las diferencias en los perfiles mitocondriales en Andalucía fueron los linajes K, H, U6, N1 y L.

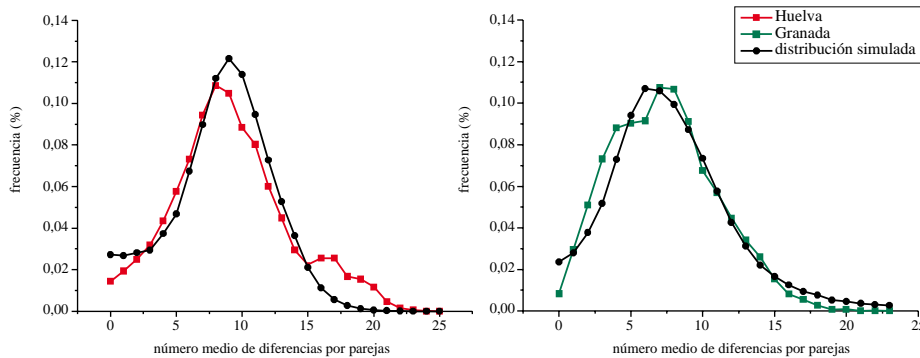


Figura III-1. Distribuciones *mismatch* observadas y esperadas (en color negro) de las secuencias de la región control mitocondrial de Huelva y Granada.

3.2. El influjo de los linajes mitocondriales en Andalucía

Haplogrupos Eurasiáticos

Casi la mitad del acervo génico matrilineal de Europa occidental está compuesto por linajes y sub-linajes que surgen del clado parental R0. Uno de sus clados descendientes, el haplogrupo H, representa el 40-50% de la variación total en Europa, con un gradiente negativo hacia el este del continente (Achilli et al., 2004; Pereira et al., 2005a; Roostalu et al., 2007). En el presente estudio, la población de Huelva registró una frecuencia marcadamente menor de H (32,9% de la diversidad total mitocondrial) en comparación con sus vecinos de Granada (50,4%). Este linaje alcanza sus mayores frecuencias en el norte de España (Larruga et al., 2001; González et al., 2003).

Los sub-haplogrupos H1 y H3 muestran picos de frecuencia en la Península Ibérica, escenario que ha sido tradicionalmente interpretado como una señal del ADNmt de una expansión poblacional masiva (~15 ky) desde el refugio franco-cantábrico hacia el noreste de Europa (Achilli et al., 2004; Pereira et al., 2005a). Estos linajes se han detectado también con frecuencias altas en Andalucía (Huelva, H1: 17,7% y H3: 2,5%; y Granada, H1: 16,5% y H3: 5,8%) y en la costa mediterránea de África (Achilli et al., 2004; Cherni et al., 2009; Ennafaa et al., 2009; Ottoni et al., 2010). Sin embargo, en Oriente Medio, estos niveles son más moderados, estando incluso el sub-haplogrupo H3 ausente de algunas poblaciones (Roostalu et al., 2007).

Tabla III-2. Linajes mitocondriales detectados en las dos poblaciones contemporáneas andaluzas.

Haplogrupo	Andalucía occidental (Huelva)		Andalucía oriental (Granada)		Haplogrupo	Andalucía occidental (Huelva)		Andalucía oriental (Granada)	
	n	%	n	%		n	%	n	%
R0:	2	1,27	0	0,00	K:	18	11,39	2	1,65
R0a	2	1,27	0	0,00	K1	16	10,13	2	1,65
HV:	7	4,43	10	8,26	K2a	2	1,27	0	0,00
HV0	5	3,16	4	3,31	NR*:	0	0,00	2	1,65
HV0a	2	1,27	6	4,96	NR*	0	0,00	2	1,65
H:	52	32,91	61	50,41	N1:	0	0,00	4	3,31
H*	15	9,49	24	19,83	N1b	0	0,00	2	1,65
H1*	27	17,09	19	15,70	I*	0	0,00	2	1,65
H1a	1	0,63	0	0,00	N2:	6	3,80	2	1,65
H1b1	0	0,00	1	0,83	W*	6	3,80	2	1,65
H2a	1	0,63	1	0,83	X:	5	3,16	1	0,83
H3	4	2,53	7	5,79	X2*	1	0,63	1	0,83
H5	2	1,27	8	6,61	X2b	4	2,53	0	0,00
H6	2	1,27	1	0,83	U6:	14	8,86	2	1,65
J:	10	6,33	9	7,44	U6a	12	7,59	2	1,65
J1*	8	5,06	7	5,79	U6b	1	0,63	0	0,00
J1b1	2	1,27	0	0,00	U6c	1	0,63	0	0,00
J2	0	0,00	2	1,65	M:	0	0,00	1	0,83
T:	10	6,33	11	9,09	M1	0	0,00	1	0,83
T*	0	0,00	3	2,48	L:	9	5,70	1	0,83
T1	1	0,63	0	0,00	L1b	4	2,53	0	0,00
T2	9	5,70	8	6,61	L2a	2	1,27	0	0,00
U(xU6):	25	15,82	15	12,40	L2b	3	1,90	0	0,00
U2e	2	1,27	0	0,00	L3	0	0,00	1	0,83
U3*	0	0,00	1	0,83	<i>Total linajes eurasiáticos [R0, HV, H, J, T, U(xU6), K, NR*, N1, N2, X]</i>				
U3a	7	4,43	0	0,00	135 85,44 117 96,69				
U4	3	1,90	0	0,00	<i>Total linajes norteafricanos [U6, M1]</i>				
U5*	0	0,00	1	0,83	14 8,86 3 2,48				
U5a	2	1,27	2	1,65	<i>Total linajes subsaharianos [L1, L2, L3]</i>				
U5b	11	6,96	10	8,26	9 5,70 1 0,83				
U7	0	0,00	1	0,83					

El sub-clado H5 muestra una frecuencia elevada en Granada (6,6%) comparado con la observada en Huelva (1,3%) y en otras poblaciones ibéricas (2-4%, Álvarez-Iglesias et al., 2009). La proporción de muestras H sin clasificar (H*, es decir, solo confirmado el polimorfismo 7028) en Granada es también mayor (20%) que lo encontrado en otros estudios de poblaciones ibéricas y norteafricanas (Pereira et al., 2005b; Coudray et al., 2009; Ennafaa et al., 2009; Bekada et al., 2013). Cabe destacar que H* es el *cluster* más representado dentro de H en Oriente Próximo, hecho que podría indicar que este grupo parafilético surgiese en esa región y se dispersara más tarde por Europa (Pereira et al., 2005a).

La incidencia de los clados hermanos J y T, ambos derivados del gran clado R, fue similar en nuestras dos poblaciones andaluzas, aunque se encontraron interesantes diversificaciones cuando analizamos las filogenias internas de estos dos haplogrupos. El sub-clado T2 agrupa ~80% de los linajes T en Europa

occidental (Pala et al., 2012), lo que se ajusta a lo detectado en nuestro estudio (en Huelva, el 90,0% de las muestras T fueron T2, mientras que encontramos un 72,7% en Granada). Escenarios comparables se han observado en otras poblaciones costeras del Mediterráneo (por ejemplo, Italia peninsular, Ottoni et al., 2009) con la excepción de la población bereber de Marruecos (Coudray et al., 2009) y en población egipcia (Rowold et al., 2007), entre los cuales T1 es el sublinaje más frecuente dentro de T.

En nuestras poblaciones andaluzas detectamos también los sub-clados J1b1a (np 242), J2a1a (nps 215, 16145, 16231) y J2b1a (np 16278). Los dos primeros (J1b1a y J2a1a) se han considerado como señales de expansiones masivas ocurridas desde Oriente Próximo hacia Europa en el periodo glacial final (~16-12 ky), mientras que el último (J2b1a) se cree que sea un marcador exclusivo europeo (Pala et al., 2012).

Nuestros datos muestran las frecuencias más altas del *cluster* U/K en Huelva (36,1% del total mitocondrial; 27,2% excluyendo el haplogrupo U6) que en Granada (15,7% U/K; 14,1% sin el U6). El rango de variación del *cluster* U/K en otras muestras ibéricas oscila entre 10-20%. El haplogrupo U es un clado complejo y antiguo, el cual presenta una distribución heterogénea de sus sublinajes a través de Eurasia (Kivisild et al., 1999). En Andalucía, U(xU6) alcanza valores alrededor del 14%, con una presencia diferencial de U2, U3, U4, U5 y U7 en la región occidental y oriental del territorio. U5 es uno de los linajes más frecuentes dentro de U entre la población contemporánea europea (7% como media, Malyarchuk et al., 2010). De acuerdo con este dato, U5 (específicamente U5b) registra en Andalucía la mayoría de la variación interna (7,0% del 15,8% de U en Huelva y un 8,3% de 12,4% en Granada). Es preciso enfatizar aquí que U5b no alcanza más del 10% entre la población mediterránea europea y está prácticamente ausente del norte de África y del Oriente Medio (Torroni et al., 2006).

Algunos miembros de las ramas más tempranas del macro-haplogrupo N (N1 y N2), y otros clados filogenéticamente relacionados como X (X2) han sido encontrados en la muestra andaluza con frecuencias <5%. El linaje N1 (N1b1), el cual es raro en europeos pero muy frecuente entre población judía (~10%, Behar et al., 2004), ha sido también detectado en la muestra de andaluces de Granada.

La presencia del linaje K en Huelva (11,4%) constituye un carácter distintivo de su *pool* mitocondrial, recordando a los valores observados en población bereber de Túnez y Egipto (11-15%, Fadhlaoui-Zid et al., 2004; Coudray et al., 2009). Sin embargo, las frecuencias de K en otras poblaciones ibéricas (5-7%) se aproximan más a lo encontrado en la mayoría de Europa (González et al., 2003; Pereira et al., 2004; Álvarez et al., 2010). Las frecuencias más elevadas de K se detectan entre judíos Askenazis (30%, Behar et al., 2004). Sin embargo, estas poblaciones judías son raras y de presencia muy reciente en España.

Haplogrupos africanos

Los linajes U6 y M1 se consideran marcadores maternos de la población contemporánea norteafricana, principalmente, de aquella que puebla el Magreb y áreas más orientales (Maca-Meyer et al., 2003b; Olivieri et al., 2006; González et al., 2007; Pennarun et al., 2012). En un marco europeo, estos haplogrupos han mostrado distribuciones irregulares, junto con el haplogrupo L (linaje mitocondrial ligado al África subsahariana) aunque los valores más altos de frecuencias se han observado en la Península Ibérica (Pereira et al., 2005b; Cerezo et al., 2012). La población andaluza occidental (Huelva) alcanza las frecuencias más elevadas (14,6%) de linajes africanos registrada hasta la fecha en Iberia y el resto de Europa. La proporción encontrada en los andaluces orientales (Granada) ha sido mucho menor (3,3%).

La presencia de U6 en la Península oscila entre el 0,5-5% del *pool* mitocondrial total (Plaza et al., 2003; Pereira et al., 2004). En el noroeste de África, este linaje es relevante tanto en frecuencia como en diversidad (en bereberes de Marruecos alcanza un 11%, con una diversidad $H = 1,000 \pm 0,096$, Coudray et al., 2009; en árabes de Túnez registra el 8%, $H = 0,833 \pm 0,222$, Cherni et al., 2005). Las secuencias U6 en Huelva están presentes por encima del rango ibérico (~9%) y también están caracterizadas por un elevado valor de diversidad ($H = 0,890 \pm 0,060$). De las ramas que emergen de U6, U6a es la más prevalente, ampliamente distribuida y diversa, y alcanza un 7,6% en Huelva. Además, las radiaciones del Hg U6 en Andalucía occidental (presente estudio) también se extienden a los sub-linajes U6b y U6c, los cuales muestran una filogeografía más limitada y fragmentada en comparación con el sub-haplogrupo U6a (Maca-Meyer et al., 2003b; Pereira et al., 2010a). El Hg M1 se encontró en un único individuo de Granada (0,83%), un patrón muy similar al detectado en el resto de la Península, donde este Hg está ausente o registra frecuencias muy bajas (<1%).

El Hg subsahariano L se encuentra en Andalucía occidental (Huelva: 5,7%, n=9) a través de sus clados derivados L1b, L2a y L2b. Comparativamente, solo se ha detectado una muestra L (L3h1b) en la provincia de Granada. El Hg L2a es el linaje mitocondrial más común y ampliamente distribuido en el continente africano (Salas et al., 2002), mientras que L1b se concentra en el África occidental y central (Harich et al., 2010; Rosa & Brehm, 2011).

La **Figura III-2** muestra una visión gráfica actualizada de la variación geográfica de los perfiles del ADNmt en Iberia.

3.3. Reconstruyendo árboles de secuencias mitocondriales

La **Figura S1** muestra el árbol genético construido con los diferentes haplotipos encontrados entre la población andaluza occidental y oriental. Como era de esperar, un gran número de haplotipos componen el *cluster* H (incluyendo H* y linajes derivados) que presenta una filogenia '*star-like*' (similar a una estrella). Aunque el resto de Hgs no muestran una estructura aparentemente definida en la *network*, las secuencias pertenecientes a L1b, L2b y U6a se caracterizan por un

gran número de pasos mutacionales desde la rCRS (en la filogenia son las ramas más profundas y largas), con la excepción de la longitud de rama asociada con L3h y U5b. Estos resultados serían coherentes con las edades de coalescencia tempranas calculadas para estos Hgs, y por tanto, relacionados con su historia genética profunda (Salas et al., 2002; Olivieri et al., 2006; Behar et al., 2008; Soares et al., 2009).

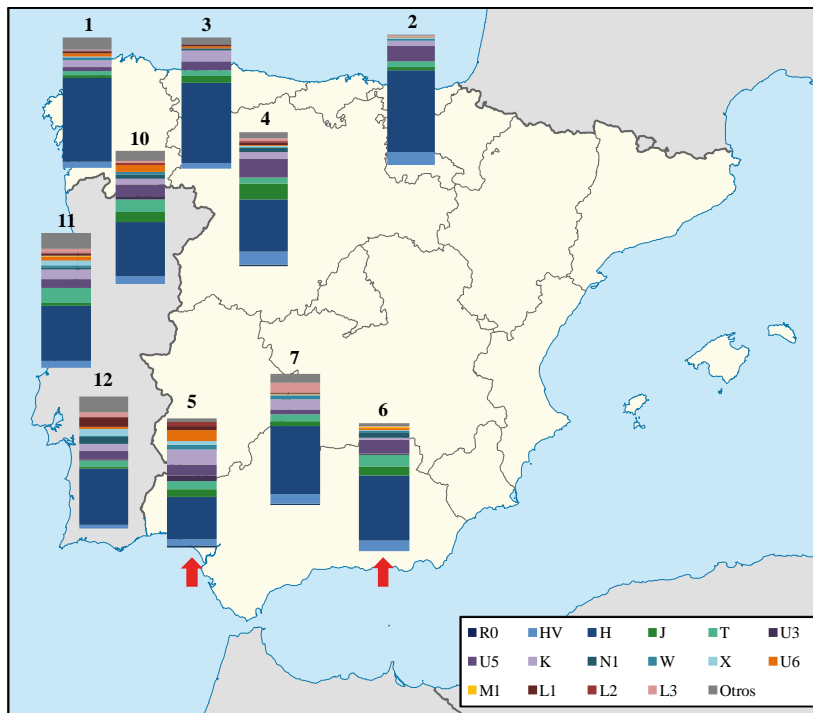


Figura III-2. Perfiles de los haplogrupos del ADNmt en algunas poblaciones de la Península Ibérica. Con flechas se resaltan las provincias de Huelva y Granada. Ver códigos de las poblaciones en la **Tabla S1**.

La **Figura III-3** representa las relaciones genealógicas de los haplotipos HVS-I de los Hgs U6 y L observados en un conjunto de poblaciones ibéricas, canarias y norteafricanas. En la filogenia de U6 (**Figura III-3A**), el sub-clado U6a (49,5% de las muestras U6) parece estar estructurado en dos sub-*clusters* cada uno mostrando una topología *star-like*. La estima TMRCA para U6a ($\rho = 1,28$; $\sigma = 0,63$) revela una edad de $21,29 \pm 10,50$ ky. El haplotipo modal (en el lado derecho), con 10 ocurrencias, se distingue por las transiciones 16172-16219-16278 y comprende 5 secuencias mitocondriales de Andalucía (Huelva y Granada) junto con muestras portuguesas (Pereira et al., 2004), canarias (Rando et al., 1999) y bereberes de Marruecos (Coudray et al., 2009). El segundo nodo de U6a (16172-16183C-16189-16219-16278) posee una contribución muy destacada de los andaluces de Huelva junto con muestras únicas de árabes de Túnez, bereberes de Marruecos y población autóctona canaria. En menor medida, otros haplotipos derivados son comunes en dos o más de las poblaciones utilizadas en el análisis. Al margen del U6a, U6 también comprende el sub-Hg U6bd, estando

bereberes (de Marruecos y Túnez). La edad evolutiva calculada para este linaje es de $18,13 \pm 5,49$ ky ($\rho = 1,09$; $\sigma = 0,33$). El sub-clado L2a muestra una situación diferente, ya que presenta altas dosis de complejidad y diversificación. El sub-haplogrupo L2b está constituido solo por una rama central ocupada por haplotipos de origen castellano y bereber y dos *tips* (ramas terminales) pertenecientes a dos muestras de Huelva y del centro de Portugal. El Hg minoritario L3h1b constituye una rama periférica de la *network*.

3.4. Estructura genética poblacional

Se ha calculado un AMOVA basado en el conjunto de poblaciones que se indica en la **Tabla S1**. Los resultados (ver **Tabla III-3**) revelan que las diferencias mitocondriales entre las poblaciones ibéricas y el resto de grupos geográficos (marcadas por los índices de fijación F_{CT}) fueron significativas, excepto en el caso de la comparación Iberia/Europa. Es interesante destacar que el alcance de la significación de F_{CT} entre Andalucía y el norte de África es menos intenso ($F_{CT} = 0,029$, $P < 0,050$) que cuando se compara la Península Ibérica global con el norte de África ($F_{CT} = 0,041$, $P < 0,001$).

Tabla III-3. Análisis molecular de la varianza (AMOVA) teniendo en cuenta diferentes escenarios geográficos. El grupo "Andalucía" se compone de las poblaciones 5, 6 y 7 (ver **Tabla S1**).

	Europa	Oriente Medio/Próximo	Sudeste de Asia	Norte de África
Península Ibérica	$F_{CT} = 0,0008$ (ns)	0,08	$F_{CT} = 0,0618^{***}$ 6,18	$F_{CT} = 0,0781^{***}$ 7,18
	$F_{SC} = 0,0310^{***}$	3,09	$F_{SC} = 0,0233^{***}$ 2,18	$F_{SC} = 0,0330^{***}$ 3,04
	$F_{ST} = 0,0318^{***}$	96,83	$F_{ST} = 0,0837^{***}$ 91,63	$F_{ST} = 0,1085^{***}$ 89,15
Andalucía	$F_{CT} = -0,0048$ (ns)	-0,48	$F_{CT} = 0,0487$ (ns) 4,87	$F_{CT} = 0,0697^*$ 6,97
	$F_{SC} = 0,0359^{***}$	3,61	$F_{SC} = 0,0376^{***}$ 3,57	$F_{SC} = 0,0414^{***}$ 3,85
	$F_{ST} = 0,0312^{***}$	96,88	$F_{ST} = 0,0844^{***}$ 91,56	$F_{ST} = 0,1082^{***}$ 89,18

ns: no-significativo; * $P < 0,05$; *** $P < 0,001$

La **Figura III-4** muestra un *Análisis Jerárquico de Cluster* (HCA) basado en el mismo conjunto de poblaciones incluidas en la **Tabla S1**. Los factores 1 y 2 representan el 62,2% de la varianza total, y cuando se consideran los 4 primeros factores, la varianza absorbida es del 81,1%. El análisis multivariante reveló 6 *clusters* (grupos) con los Hg H, U5, V, L3 y 'Otros' definiendo significativamente la topología de las poblaciones. Cuando se computa la descomposición de la inercia de los dos primeros factores, el cociente 'inercia interclusters/inercia total' es igual a 0,843. Este resultado indica que un alto porcentaje de la variación de los datos (84,3%) se encuentra explicado por esos 6 *clusters*.

El *cluster* 1 (C1) está fuertemente influenciado por las altas frecuencias del Hg H (media del *cluster* = 0,571; $P = 0,000$). Las muestras andaluzas de Granada (#6) y de Córdoba (#7) se encuentran en este agrupamiento. Las poblaciones del norte

de España como las de Galicia (#1), País Vasco (#2) y de Castilla-León (#3) están posicionadas en el extremo derecho del *espacio multidimensional*, debido a la preeminencia del Hg H en esas regiones. El C2 está también significativamente definido por el Hg H (media= 0,386; $P= 0,002$) pero en menor medida que en el C1. Este *cluster* agrupa un gran número ($n= 23$) de poblaciones: 7 de Iberia [incluyendo a Huelva (#5) y las 3 poblaciones portuguesas (#10, #11, #12)] así como otras 9 poblaciones de Europa mediterránea, Oriente Medio, Europa continental y el norte de África. La posición genética de la población de Huelva en el extremo inferior derecho y su proximidad con las Islas Canarias (#8) se interpreta sobre la base del relevante peso que tiene el Hg U6 en su *pool* mitocondrial. Además, se detecta una afinidad entre la población castellana de Zamora (#4) y el norte de Portugal (uno de sus vecinos geográficos más próximos, #10). Este resultado podría indicar antiguas mezclas poblacionales previas a la emergencia del reino de Portugal en el siglo XII.

El *cluster* C3 está significativamente caracterizado por frecuencias altas de los clados del ADNmt típicamente africanos: L3 (media= 0,172; $P= 0,000$), L2 (media= 0,089; $P= 0,000$), L1 (media= 0,051; $P= 0,000$), U6 (media= 0,045; $P= 0,002$) y M (M1) (media= 0,052; $P= 0,013$) y, por tanto, coherentemente agrupa 7 poblaciones del norte de África. Finalmente, el C5 –definido por los Hgs R0, M, HV*, N y K– concentra a poblaciones del sudoeste asiático, mientras que el C6 está compuesto fundamentalmente por poblaciones del Oriente Medio y otras cercanas como Siria (#32), Irak (#34), Yemen (#41) y Egipto (#53). Los Hgs englobados en la categoría ‘Otros’ (linajes no encontrados frecuentemente en Europa) determinan este último *cluster*. También desarrollamos un HCA adicional (no mostrado aquí), excluyendo a las poblaciones pertenecientes a C4, C5 y C6. La topología encontrada mostró el mismo patrón que el HCA original, no alcanzándose en consecuencia una mejora de la interpretación visual de los datos.

4. Discusión

El *pool* genético mitocondrial de las poblaciones andaluzas de Huelva y Granada ha revelado un amplio espectro de haplogrupos con diferentes orígenes continentales. Las bajas frecuencias de marcadores euroasiáticos, junto con una alta incidencia y diversificación de los linajes maternos africanos entre los andaluces occidentales (Huelva) –en comparación con sus vecinos de Granada y otras áreas de Iberia– constituyen resultados relevantes no conocidos previamente acerca de los patrones mitocondriales dentro de Andalucía y la Península Ibérica, indicando una sub-estructura genética femenina. Por tanto, Andalucía no puede ser considerada como una población única y homogénea.

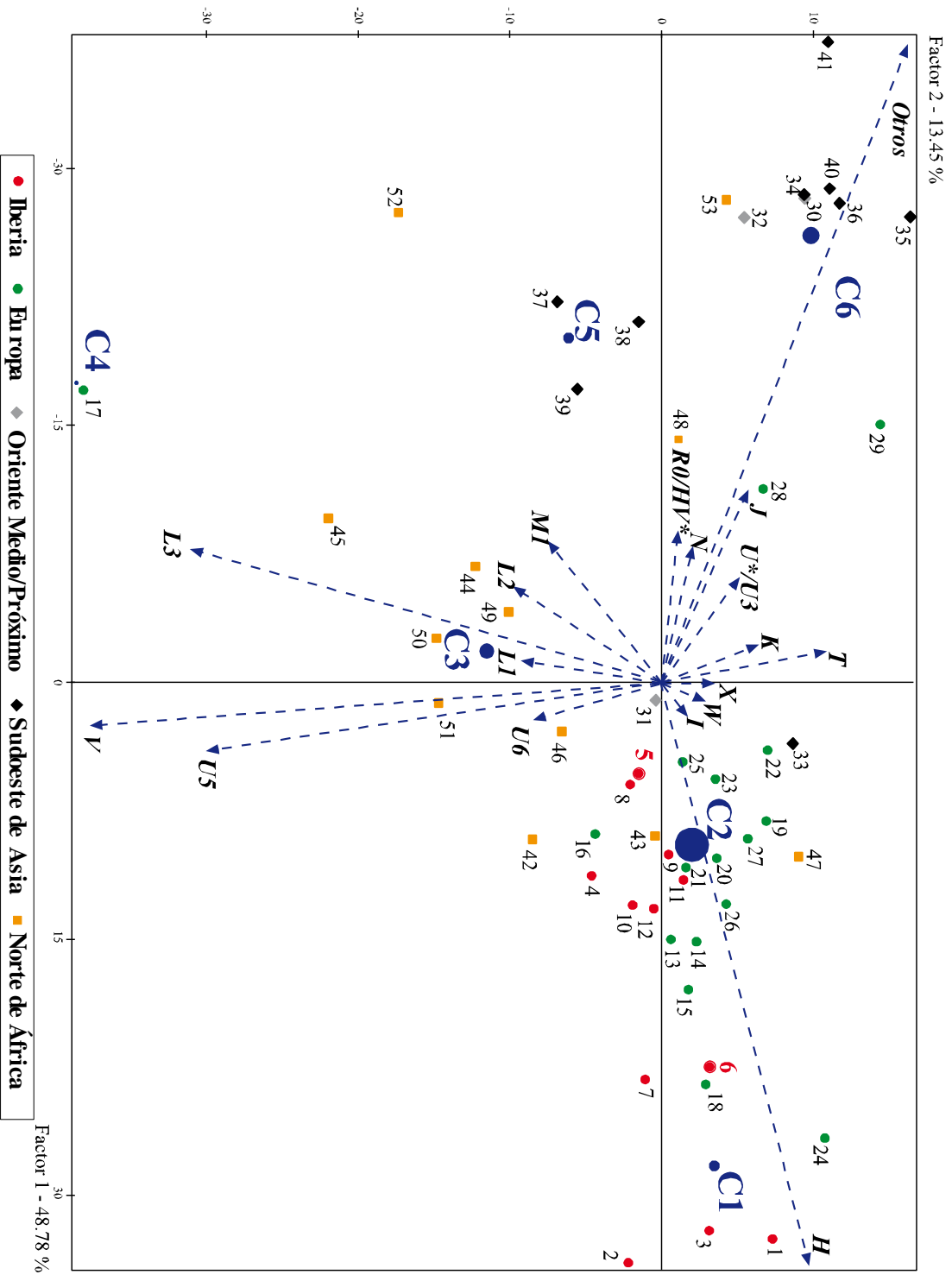


Figura III-4. Análisis jerárquico de Cluster (HCA) de 53 poblaciones basado en su diversidad mitocondrial. Los haplogrupos usados se marcan con flechas (vectores). Las poblaciones se indican con sus códigos (ver **Tabla S1**).

En el HCA de la **Figura III-4** se observa que las dos poblaciones andaluzas se sitúan en diferentes *clusters*. Cabe destacar que Huelva está localizada en una posición central del espacio bidimensional, sugiriendo una mezcla variada que implica a poblaciones del Mediterráneo oriental y central, así como del Magreb. Además, cuando se analiza la topología del árbol dibujada a partir del HCA (datos no mostrados), la primera separación se produce entre las poblaciones más orientales (árabes, egipcios, poblaciones del Cáucaso y de Irán), del resto de poblaciones europeas y del noroeste africano. Esta escisión sugiere que el alcance de la mezcla poblacional de estos grupos del Este fuera de su región geográfica haya sido baja, y por tanto, que la prolongada costa del Mediterráneo oriental entre Egipto y Túnez habría actuado como una barrera suficientemente impermeable al flujo de genes maternos.

Las migraciones deben ser definidas por su dirección, siendo determinadas tanto por las poblaciones fuente y sumidero como por el tiempo en el cual ese movimiento tuvo lugar. La relación entre los tamaños demográficos de estas dos poblaciones y la distribución de sus frecuencias génicas son factores importantes para que la migración deje una señal genéticamente distinguible. El número de individuos que comprende cada migración debe alcanzar un tamaño moderado, y aunque los linajes puedan extenderse o dispersarse en más de una migración, el número de esos movimientos debe ser limitado, incluso en territorios en los cuales se han sucedido migraciones desde diversos orígenes, como es el caso de la Península Ibérica. En el Magreb, este número podría ser menor que en Iberia, debido a su posición geográfica. Además, para que las migraciones entre regiones espacialmente distantes puedan ser genéticamente efectivas, deben haber ocurrido con un número de etapas intermedias. El marco temporal plausible para esos movimientos humanos debe ser posterior al estimado por el TMRCA para cada linaje. Este hecho proporcionaría un tiempo suficiente de maduración del linaje para que alcance una frecuencia considerable en la población fuente. Así, un número moderado de emigrantes podrán asegurar una reproducción efectiva de ese marcador en un nuevo asentamiento. Para los linajes maternos, esos movimientos deben implicar a mujeres, lo que limita aún más el tamaño efectivo migratorio final.

Botigué et al. (2013), basándose en un análisis de barrido genómico masivo, han inferido una elevada ancestría africana común en Iberia, que puede ser trazada a poblaciones del Magreb, postulando que la mayor diversidad en el sudoeste de Europa es el resultado de una contribución sustancial de migraciones recientes desde el norte de África. Sin embargo, ellos también muestran (ver Figuras 1 y 2) una influencia significativa del Oriente Próximo en Iberia. Nuestros resultados, basados en el ADNmt, revelan linajes compartidos entre poblaciones de Andalucía, noroeste de África y otras regiones del espacio Mediterráneo.

Los linajes maternos más adecuados para estudiar migraciones entre la Península Ibérica y el norte de África son el U6 y el L. En la actualidad, existe un interesante debate acerca de los orígenes y dispersión del Hg U6. Algunos autores (Olivieri et al., 2006) han propuesto que U6 pudo estar involucrado en una

dispersión temprana (40-45 ky) desde el sudoeste de Asia hacia el norte de África. Recientemente, otros investigadores (Pennarun et al., 2012) han postulado que algunas de las ramas topológicamente más tempranas del U6, hubieran estado ligadas a la cultura Iberomaurisiense (20-9 ky), la cual se distribuía principalmente por el noroeste de África.

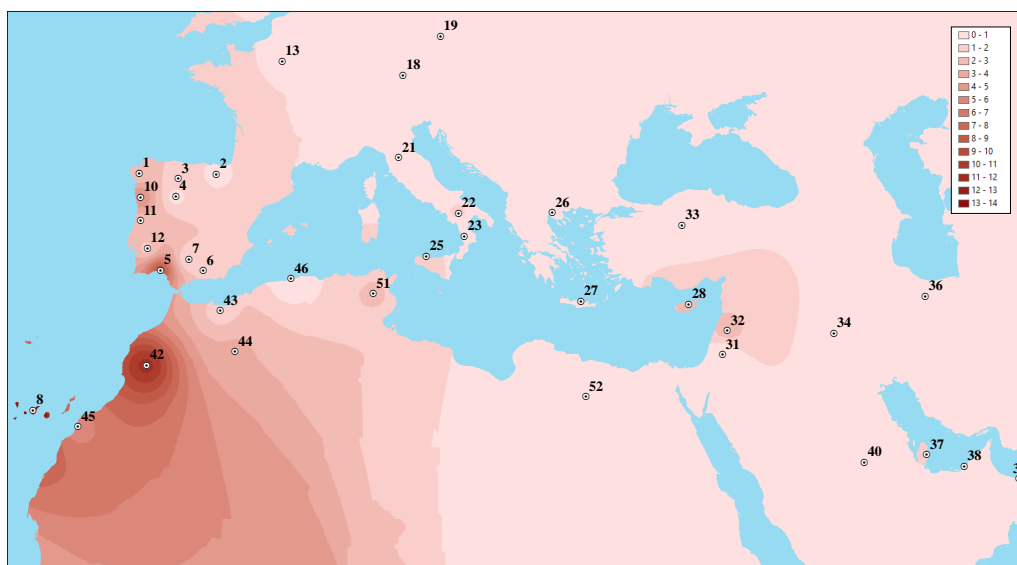
La **Figura III-5** muestra un mapa de interpolación con las variaciones geográficas del Hg U6 alrededor del Mediterráneo. Mientras este tipo de análisis solo permite visualizar ampliamente la filogeografía de U6 (y no inferir tiempos de migración) esta representación destaca: (1) cómo la fachada atlántica de Marruecos –incluyendo a las vecinas islas Canarias– concentra altas frecuencias de ese linaje, y (2) cómo su patrón espacial exhibe una suave y continua disminución desde Marruecos en dirección norte, alcanzando el cinturón occidental de la Península Ibérica, con las frecuencias más perceptibles en el territorio de Huelva (presente estudio).

Los árboles *median-joining* aportan luz sobre las relaciones entre linajes compartidos (haplotipos). En la *network* del clado U6 (ver **Figura III-3A**) las muestras canarias (excepto en dos de ellas) pertenecen al Hg U6b1. La alta frecuencia de este marcador, específico del archipiélago, no habría influido en las migraciones entre Europa y África. Ello puede ser, debido a su pequeño tamaño poblacional, el tiempo tardío de colonización de las islas (2,5 ky) y la relativamente reciente edad evolutiva de U6b (~13 ky) (Pennarun et al., 2012). El sub-linaje U6c, con una edad de coalescencia de 11 ky (Soares et al., 2009; Pennarun et al., 2012) comprende solo 2 individuos andaluces, sin traza de influencia africana. La mayoría de las muestras no canarias pertenecen al U6a, el cual muestra dos importantes, grandes nodos, ambos conectados por dos mutaciones donde los linajes ibéricos (n= 11) predominan sobre los norteafricanos. De hecho, las ramas terminales de U6a no muestran derivaciones a través de individuos norteafricanos.

La expansión musulmana ha sido habitualmente citada como la principal causa para el flujo génico entre el norte de África y la Península Ibérica, aunque ese escenario podría ser solo un episodio de varias interacciones debidas a múltiples movimientos de costa a costa. El mayor tamaño poblacional del lado ibérico habría dejado probablemente mayores y apreciables huellas genéticas en la población norteafricana que el movimiento recíproco de genes. El tamaño poblacional del sur de Iberia –información basada en sitios arqueológicos– era ya relevante desde el Paleolítico (ver Figura 5 en Bocquet-Appel et al., 2005) y, aunque no se tienen datos del tamaño poblacional del norte de África, debería ser presumiblemente menor utilizando metodología similar. Considerando esta información demográfica junto con la estructura de la *network* y la abundancia de U6a en el sudoeste de Iberia, algunos clados de este linaje podrían haber tenido su origen en la Península Ibérica, desde donde se habrían re-expandido hacia el norte de África. La presencia de U6 en el Oriente Próximo –con menores frecuencias que en África– y también en Etiopía, sería el reflejo de contactos a través del mar en el pasado. La hipótesis de una vía terrestre a lo largo de la franja costera norteafricana parece menos plausible, dada la gran distancia y la

ausencia de sitios de asentamiento dotados de suficientes recursos de abastecimiento entre el estuario del Nilo y la región de Trípoli. Estas consideraciones indican que la existencia de repetidos contactos entre poblaciones de esas localizaciones geográficas opuestas en el pasado es poco probable.

Figura III-5. Mapa de interpolación de frecuencias (%) del linaje U6 a través del espacio Mediterráneo. Ver códigos y referencias en la **Tabla S1**.



Algunos investigadores (Zilhão, 2001; Guilaine, 2003, entre otros) han propuesto que una ruta marítima sea la vía más probable de llegada del Neolítico a Iberia. Este proceso fue por fases, usando navegación y barcos suficientemente grandes para transportar hombres, mujeres y el ‘paquete Neolítico’, todo ello en un movimiento probablemente originado en el golfo de Génova. Métodos similares podrían haber sido usados entre Andalucía y Marruecos. Todos estos acontecimientos sugieren que las interacciones entre las poblaciones marroquíes y andaluzas han sido antiguas, continuas, en ambas direcciones y con diferentes orígenes.

Posteriormente, durante la expansión colonial al principio del primer milenio BC, hubo algunos contactos marítimos entre el este y el oeste del Mediterráneo. Estos contactos continuaron durante la hegemonía cartaginesa y el dominio del Imperio Romano. Como resultado de la reorganización administrativa de Diocleciano (a finales del siglo III AD), la provincia *Mauretania Tingitania* (norte de Marruecos) fue agrupada junto a las provincias peninsulares para formar la *Diocesis Hispaniarum*, de las cuales la región de Tánger (Marruecos) y Andalucía fueron los primeros territorios en ser cristianizados a comienzos del siglo IV. En consecuencia, parece poco probable que los linajes U6a localizados en el nodo central de la *network* fueran el resultado de una expansión posterior del Islam en Iberia, a pesar de las múltiples invasiones que ocurrieron desde el Magreb hacia

la Península durante ese período. Cabe destacar que otros datos genéticos apoyan esta hipótesis. La presencia remarcable de haplogrupos específicos europeos en el *pool* materno de las poblaciones contemporáneas del noroeste de África puede ser explicada por la ocurrencia de migraciones desde Europa, siendo la Península Ibérica una fuente importante de este flujo de genes (Cherni et al., 2009; Ottoni et al., 2010; Bekada et al., 2013). Por ello, el Magreb habría experimentado flujo génico materno desde el continente europeo a través del mar desde tiempos antiguos.

En lo que concierne al linaje subsahariano L (**Figura III-3B**), el L1b está caracterizado por un elevado número de individuos en su nodo central. De acuerdo con Soares et al. (2009) su edad estimada de coalescencia es 9,7 ky, por lo que su expansión fuera de África habría tenido lugar durante o después del Neolítico. Su topología *star-like* indica una expansión poblacional probablemente modelada en tiempos muy antiguos, por lo que la expansión musulmana o migraciones más recientes no podrían ser la razón causal principal. Además, este nodo central está compuesto por bereberes provenientes de una amplia área geográfica –cubriendo desde Egipto a Marruecos– y por población española de las provincias occidentales, desde Huelva (en el extremo sur) hasta León (en el noroeste). Ambos territorios españoles estaban conectados por la calzada romana denominada ‘*Vía de la Plata*’ que adquirió una importancia militar y comercial relevante no solo durante el Imperio Romano, sino también antes y después, debido a que conectaba importantes ciudades y yacimientos mineros. Esta calzada habría permitido los movimientos poblacionales desde el sur y noroeste de España durante largos períodos. Los linajes maternos compartidos entre las poblaciones españolas occidentales podrían corresponder a mujeres que apoyaran los contingentes militares o comerciales. El hecho de que las poblaciones portuguesas estén ausentes justo del nodo central de L1b –todas las secuencias portuguesas que pertenecen a este clado son derivadas– revela que las historias genéticas de España y Portugal son diferentes para este sub-haplogrupo mitocondrial.

El linaje L2a, el segundo más representado en la *network* (edad de coalescencia ~48 ky, Soares et al., 2009) no muestra una topología *star-like*. Este haplogrupo se compone de una mayoría de bereberes y algunos miembros ibéricos, y solo unos pocos haplotipos son compartidos. En Iberia, estos linajes se podrían asociar a la expansión islámica, que penetró hasta el norte de Portugal, por lo que relacionar la presencia del L2a con el reciente tráfico de esclavos es poco probable. Sin embargo, no se puede descartar la influencia de la esclavitud durante el Imperio Romano o la dominación musulmana.

Algunos autores (Pereira et al., 2005b) han apuntado a la trata de esclavos atlántica (siglos XV-XIX) como la fuente de los linajes L en Portugal –el destino principal de los esclavos negros en Europa. De este modo, se ha descartado un movimiento de introducción similar para los linajes U6 y L en la Península Ibérica, exponiendo que ambos clados mitocondriales muestran distribuciones geográficas distintivas y diferentes frecuencias en población española y portuguesa. La introducción en Iberia de genes subsaharianos provenientes de

Iberoamérica podría ser también un fenómeno a considerar. La presencia y las relativas altas frecuencias del alotipo africano de las inmunoglobulinas *GM*1,17 23' 5** en poblaciones occidentales de la Península Ibérica ha sido interpretado en esa misma dirección (Calderón et al., 2006, 2007).

Las migraciones entre el Magreb y la Península Ibérica han tenido lugar a través del mar desde el Paleolítico. En este contexto, el transporte de mujeres implica el uso de naves más grandes que las que están solo tripuladas por hombres. La navegación marítima podría haber sido una práctica común como técnica de pesca, sin perder visión de la tierra por seguridad. La menor distancia entre Europa y África se localiza en el Estrecho de Gibraltar, y esa distancia se incrementa marcadamente hacia el este, alcanzando una separación de unos 150 km. Las corrientes del viento y océanos son muy variables y peligrosas alrededor del Estrecho, y una ruta alternativa más apacible sería la que conecta el Cabo de Tres Forcas (Marruecos) con la costa andaluza entre Málaga y Almería, en la cual la pequeña isla de Alborán (aproximadamente 36°N, 3°W) está a mitad de camino. Esta isla podría haber sido usada como escala. En esa ruta de Alborán, las corrientes y vientos se suceden desde Europa hacia África la mayor parte del año, y las altas montañas del Rif y Sierra Nevada permiten una vista permanente de la tierra durante la travesía. Por tanto, parece probable que la isla fuera conocida entre los cazadores-recolectores del Paleolítico, debido a su riqueza en mamíferos marinos, grandes peces y una zona extensiva intertidal, probablemente usada por el hombre debido a su riqueza en fauna comestible (Templado & Calvo, 2006). Además, restos arqueológicos encontrados en Marruecos oriental (el '*arpón de Tafortalt*', Camps, 1974) o evidencias de pesca al brumeo (en la cueva de Nerja, Málaga) durante el Magdalenense superior (~12-10 ky) sugieren la existencia de contactos entre las costas norte y sur del Mediterráneo occidental.

Los resultados obtenidos subrayan la necesidad de desarrollar análisis más detallados sobre las relaciones genéticas en ambos lados del Mediterráneo occidental, utilizando muestras cuidadosamente seleccionadas de individuos autóctonos. Estos estudios deberían incluir diferentes marcadores analizados en alta resolución genética, ya que las migraciones han sido numerosas en la región a través de un largo período de la historia humana, presumiblemente desde tiempos Paleolíticos. Muchos estudios se han centrado en el flujo génico reciente del norte de África en Iberia, aunque la atención científica debería ser ahora dirigida a un estudio detallado de la introducción de genes europeos en el noroeste de África a través del mar, con el fin de determinar su magnitud, escala temporal y métodos, y compararlo con aquellos movimientos terrestres desde el África oriental y el sudoeste de Asia. El origen africano del *Homo sapiens* 'moderno' no implica que el poblamiento del extremo noroeste de ese continente tuviera lugar desde el interior exclusivamente.

Capítulo IV

Linajes africanos en Iberia: señales de un flujo génico antiguo

Este capítulo está basado en la siguiente publicación: **Hernández C.L.**, Soares P., Dugoujon J.-M., Novelletto A., Rodríguez J.N., Rito T., Oliveira M., Melhaoui M., Baali A., Pereira L., Calderón R. (2015) Early Holocenic and historic mtDNA African signatures in the Iberian Peninsula: the Andalusian region as a paradigm. *PLoS One* (aceptado para su publicación, 17 sept. 2015).

Determinar el tiempo, la identidad y la direccionalidad de las migraciones en la cuenca mediterránea, el papel de las 'rutas migratorias' dentro y entre regiones de África, Europa y Asia, y los efectos de los comportamientos específicos dependientes del sexo de esos movimientos poblacionales, tiene importantes implicaciones para el conocimiento de la diversidad genética humana actual. Un componente crucial del mundo mediterráneo es su extremo occidental. Datos arqueológicos han permitido identificar claras señales de contactos transcontinentales antiguos entre las poblaciones norteafricanas e ibéricas que circundan la región del Estrecho de Gibraltar. El establecimiento del origen y los tiempos de las migraciones entre áreas geográficamente próximas constituye hoy un reto en el campo de la genética de poblaciones basada en marcadores uniparentales. Cuando se pretende evaluar las migraciones de norte a sur, los estudios de ADN mitocondrial (ADNmt) se han centrado en los linajes H1, H3 y V, y para la dirección opuesta, en los haplogrupos U6 y L, asumiendo que estos linajes son buenos representantes de la ancestría de cada orilla del Mediterráneo. En el presente trabajo hemos analizado secuencias mitocondriales completas pertenecientes a los haplogrupos U6, M1 y L en Andalucía –provincias de Huelva y Granada– y en la población bereber de Marruecos. Aquí se proporcionan datos e interpretaciones pioneras acerca del papel del noroeste de África y la Península Ibérica con respecto al origen, número de fundadores y dirección de la expansión de esos marcadores específicos. La entrada estimada de los linajes norteafricanos U6 a Iberia, ~10 ky, se ajusta a la de otros clados L, indicando que U6 y algunos linajes L se movieron conjuntamente desde África a Iberia en el Holoceno temprano. Del mismo modo, el análisis fundador pone de manifiesto que el gran nivel de afinidad de los linajes entre el norte de África e Iberia es el resultado de un proceso continuado, a lo largo del tiempo, que impide hacer interpretaciones simplistas. En particular, nuestro trabajo apoya la existencia, frecuentemente no reconocida, de un antiguo puente entre el Magreb y Andalucía.

Resumen

Determining the timing, identity and direction of migrations in the Mediterranean Basin, the role of "migratory routes" in and among regions of Africa, Europe and Asia, and the effects of sex-specific behaviors of population movements have important implications for our understanding of the present human genetic diversity. A crucial component of the Mediterranean world is its westernmost region. Clear features of transcontinental ancient contacts between North African and Iberian populations surrounding the maritime region of Gibraltar Strait have been identified from archeological data. The attempt to discern origin and dates of migration between close geographically related regions has been a challenge in the field of uniparental-based population genetics. Mitochondrial DNA (mtDNA) studies have been focused on surveying the H1, H3 and V lineages when trying to ascertain north-south migrations, and U6 and L in the opposite direction, assuming that those lineages are good proxies for the ancestry of each side of the Mediterranean. To this end, in the present work we have screened entire mtDNA sequences belonging to U6, M1 and L haplogroups in Andalusians –from Huelva and Granada provinces–

Abstract

and Moroccan Berbers. We present here pioneer data and interpretations on the role of NW Africa and the Iberian Peninsula regarding the time of origin, number of founders and expansion directions of these specific markers. The estimated entrance of the North African U6 lineages into Iberia at 10 ky correlates well with other L African clades, indicating that U6 and some L lineages moved together from Africa to Iberia in the Early Holocene. Still, founder analysis highlights that the high sharing of lineages between North Africa and Iberia results from a complex process continued through time, impairing simplistic interpretations. In particular, our work supports the existence of an ancient, frequently denied, bridge connecting the Maghreb and Andalusia.

1. Introducción

Determinar el tiempo, la identidad y la direccionalidad de las migraciones en el mar Mediterráneo a través de la historia humana, el papel de las 'rutas migratorias' dentro y entre regiones de África, Europa y Asia, y el efecto de los movimientos poblacionales específicos según el sexo, son todos ellos enfoques de especial interés en Paleoantropología, Prehistoria, Arqueología y Genética de Poblaciones (Souville, 1998; Straus, 2001; Currat et al., 2010; Brandt et al., 2015).

La proximidad geográfica entre África y las costas mediterráneas europeas, principalmente en su extremo occidental (14,3 km), ha permitido múltiples movimientos humanos de costa a costa. Como un componente crucial del mundo mediterráneo, el norte de África ha estado principalmente implicado en migraciones humanas dentro de la región y fuera de ella; también ha recibido migrantes de la vecina Europa mediterránea y del sudoeste de Asia. Todos esos movimientos han contribuido sustancialmente al desarrollo de intercambios humanos, culturales y comerciales. El norte de África se distingue por su compleja prehistoria e historia, proporcionando ambas un distintivo registro arqueológico (Lubell, 2001). Las industrias prehistóricas que se han registrado sucesivamente en esa franja geográfica (desde el Ateriense al Capsiense, a través de Iberomauriense) no siguen una secuencia lineal de reemplazamiento, sino que suscitan un interesante debate respecto a la continuidad o discontinuidad poblacional (Irish, 2000; Barton et al., 2013). La región ha sufrido también cambios climáticos extremos, con oscilaciones entre fases áridas que no permitieron la supervivencia humana, a episodios húmedos que transformaron parte del vecino desierto del Sahara en un paisaje verde.

Se cree que durante varios períodos de la historia humana pudieran tener lugar migraciones transcontinentales, aunque la mayoría de investigadores proponen un límite superior al final del Paleolítico (Straus, 2001). Un argumento comúnmente usado en contra del flujo génico prehistórico se basa en las peligrosas corrientes y vientos que se crean alrededor del Estrecho de Gibraltar, el cual une el mar Mediterráneo con el Océano Atlántico. Sin embargo, otras áreas más orientales, como la de Alborán –la cual se encuentra bordeada por el sudeste

de Iberia y noreste de Marruecos– proporcionan una travesía más segura, aunque también más larga (Souville, 1998). Claras señales de contactos transcontinentales antiguos han sido identificados a partir de restos arqueológicos, como el ‘*arpón de Taforalt*’, un utensilio Iberomauriense contemporáneo con el Magdaleniense español, detectado en el noreste de Marruecos (Camps, 1974). Posteriormente, los fuertes paralelismos en la introducción, adopción y expansión del ‘paquete Neolítico’ (las mismas especies domesticadas, similar tecnología, etc.) en el sur de Iberia y el Magreb, se interpretan como evidencias de tangibles y estratégicas relaciones poblacionales entre ambas orillas (Manen et al., 2007; Linstädter et al., 2012).

Los genes maternos africanos han dejado huellas tanto en el sur de Iberia como en la fachada atlántica de la Península (ver **Capítulo III** y Pereira et al., 2005b). Comparativamente, los marcadores son menos abundantes para los genes paternos, aunque se han observado trazas visibles del linaje E-M81 del cromosoma Y (conocido como el ‘marcador bereber’) en el *pool* génico de la población andaluza (Ambrosio et al., 2010a, 2010b). Un estudio reciente referido a la distribución geográfica de los genes autosómicos de las inmunoglobulinas en el *locus GM* a través del Mediterráneo ha puesto de manifiesto la relativa alta frecuencia (4% como valor promedio) del haplotipo subsahariano *GM 1,17 5** en la región andaluza y en la Iberia atlántica, en comparación con otras poblaciones de la Europa mediterránea (Fortes-Lima et al., 2014). Estos interesantes resultados son concordantes con los patrones de variación de los haplogrupos mitocondriales africanos a través de la Península Ibérica. Otros estudios genéticos desarrollados en población mediterránea occidental han revelado una contribución notable de linajes maternos euroasiáticos (~80% para algunas poblaciones) en el norte de África (Cherni et al., 2009; Ennafaa et al., 2009; Rhouda et al., 2009; Ottoni et al., 2010). Estas señales permiten diferenciar a las poblaciones africanas más septentrionales de las subsaharianas (Cherni et al., 2005; Coudray et al., 2009; Fadhlouzi-Zid et al., 2011; Bekada et al., 2013).

El estudio de linajes africanos específicos en la Europa mediterránea ha conducido al análisis de genomas mitocondriales completos (mitogenomas), especialmente, en la Península Ibérica y su población insular (Maca-Meyer et al., 2003b; Pereira et al., 2010a; Cerezo et al., 2012; Pardiñas et al., 2014; Secher et al., 2014). Estos clados mitocondriales africanos, que tienen historias evolutivas distintivas, están asociados a patrones particulares de dispersión y origen geográfico. Las distribuciones de U6 y M1 sugieren que estos marcadores serían componentes relevantes de las poblaciones norteafricanas, encontrándose el U6 más representado en el noroeste de África, y el M1 entre población del noreste (Pennarun et al., 2012). El macro-haplogrupo L, donde se localiza la raíz de la filogenia mitocondrial humana, exhibe, sin embargo, una incidencia particularmente alta y una amplia distribución por el África subsahariana, reduciéndose la misma hacia el norte del continente (Salas et al., 2002; Soares et al., 2012).

El **Capítulo III** se ha centrado en evaluar los patrones globales de variación del ADNmt en las regiones occidental y oriental de Andalucía, representadas por

los individuos autóctonos de Huelva y Granada, respectivamente. Los resultados han puesto de manifiesto una alta diversidad genética, acompañada de una contribución africana heterogénea dentro del territorio: 8,9% y 1,7% de linajes U6; 5,7% y 0,8% de linajes L, en Huelva y Granada, respectivamente (ver **Tabla III-2**). Estos datos, junto con lo expuesto más arriba, proporcionan un buen soporte científico para considerar a estas poblaciones autóctonas como candidatas preferentes para el estudio de las migraciones humanas transcontinentales y de las relaciones genéticas en el extremo occidental del Mediterráneo. El área seleccionada es compleja y de gran importancia estratégica. Las poblaciones que circundan la región marítima del Estrecho de Gibraltar, tanto las localizadas en el norte como las del sur, exhiben marcadores de flujo génico recíproco.

De acuerdo con esos resultados previos, hemos integrado 62 secuencias completas mitocondriales de población andaluza y de bereberes de Marruecos (de las poblaciones de Asni, Bouhria y Figuig, Coudray et al., 2009) en una base de datos actualizada de linajes U6, M1 y L con el fin de: (1) explorar los patrones filogeográficos de esos haplogrupos y sus sub-ramas derivadas, fundamentalmente, a través de la región mediterránea, (2) refinar sus filogenias actuales y estimar edades de coalescencia y, (3) obtener información fundamentada acerca del papel del norte de África y la Península Ibérica en cuanto al origen, número de fundadores y dirección de expansión de esos linajes. También se presentan inferencias de dinámica poblacional basada en secuencias U6 completas. En resumen, los análisis desarrollados en este Capítulo pretenden profundizar en el conocimiento del impacto genético de los movimientos poblacionales y acontecimientos demográficos que han ocurrido en los últimos milenios en el extremo occidental del Mediterráneo, fundamentalmente, aquellos referidos a la herencia mitocondrial africana.

2. Material y métodos

2.1. Poblaciones y selección de muestras

Se ha analizado el genoma mitocondrial completo de muestras cuidadosamente seleccionadas del sur de España (Andalucía, n= 32) y del norte de África (bereberes de Marruecos, n= 30) pertenecientes a los linajes U6, M1, L1, L2 y L3. Las muestras ibéricas provienen de un conjunto de muestras autóctonas, no emparentadas (n= 750; ver **Capítulo II**, sección 4). La misma estrategia se aplicó a la población marroquí (n= 217; Coudray et al., 2009) para evaluar su autoctonía. El trabajo de colección de las muestras y el proceso de muestreo de los individuos bereberes fue desarrollado personalmente por el Prof. J. M. Dugoujon. La información sobre las localidades de muestreo en Marruecos se detalla en Coudray et al., 2009. Tanto el muestreo como los protocolos moleculares fueron aprobados por el Comité de Bioética de la *Université Paul*

Sabatier III de Toulouse (Francia) y se obtuvo el consentimiento informado de todos los participantes.

En Andalucía, llevamos a cabo un incremento del conjunto de muestras respecto a lo analizado en el **Capítulo III** (procedimiento detallado en la sección **II-4**). Con este propósito, se diseñaron dos ensayos TaqMan de genotipado para la detección de las muestras U6 (A3348G) y L/M (T10873C) (ver condiciones y secuencias en **Anexo II, Tabla S3**). En total, se evaluó la presencia de individuos pertenecientes a linajes africanos en las 750 muestras autóctonas andaluzas. Se detectaron 50 individuos portadores de alguno de los haplogrupos U6, M1 y L, y se seleccionaron 32 para la secuenciación completa del genoma mitocondrial. Esta selección se realizó sobre la base de los haplotipos de la región control. De esta manera, todos los individuos que presentaron haplotipos distintivos fueron seleccionados. Teniendo en cuenta el gran tamaño muestral manejado y las precauciones en la selección de las muestras, se puede afirmar que nuestro estudio comprende una fracción elevada de la diversidad mitocondrial de los linajes africanos en el sur de Iberia.

Por otro lado, se secuenciaron también los mitogenomas de 30 individuos bereberes de Marruecos (7 Asni, 7 Bouhria y 16 Figuig). De nuevo, la selección de estas 30 muestras –a partir de un total de 80 individuos pertenecientes a linajes africanos– se basó en información de la región control.

2.2. *Secuenciación de mitogenomas y análisis estadísticos*

El proceso de secuenciación del ADNmt completo de las 62 muestras se detalla en la sección **II-5.4**. Las secuencias completas fueron depositadas en GenBank (*Accession Numbers* KT819205 - KT819266). Las posiciones 16182C, 16183C y 16519, los *indels* y fenómenos de heteroplasmia se omitieron de los análisis filogenéticos.

Se construyeron mapas de contornos (*surface maps*) para obtener una visión más amplia de los patrones de distribución geográfica de los principales linajes africanos (ver metodología en la sección **II-6.7**). Con este objetivo, se recopilaron un total de 3.661 secuencias HVS-I (ver datos y referencias en la **Tabla S4**) a partir de una base de datos general compuesta por 34.229 muestras pertenecientes a Iberia, Europa, Oriente Próximo/Medio, y África. Esta compilación de secuencias (rango 16051-16400 de la rCRS, Andrews et al., 1999) incluye 317 secuencias U6, 361 M1 y 2.983 pertenecientes a ciertos sub-haplogrupos L (L1b, L2a, L2b, L3b, L3d, L3f, L3h1b). Todas las poblaciones consideradas para nuestros análisis tienen un tamaño poblacional ≥ 50 . Cabe destacar aquí que las poblaciones africanas seleccionadas para los mapas de interpolación de frecuencias (ver **Figura S2**) se dispersan siguiendo los patrones de distribución espacial y asentamiento observados hoy en el Continente (ver Linard et al., 2012 para más detalles).

Las secuencias HVS-I adscritas a los haplogrupos U6a y L1b se analizaron mediante *median-joining networks* específicos (Bandelt et al., 1999) usando 209 y

175 muestras, respectivamente, pertenecientes a regiones estratégicas (Iberia peninsular, Italia y sus islas, junto con población africana). Además, se calcularon los valores de diversidad de estos sub-haplogrupos mediante el número medio de diferencias por pares de secuencias (MNPD) (ver sección II-6.1).

Para la reconstrucción filogenética, basada en secuencias completas, se realizó una compilación de 2.182 mitogenomas (ver **Dataset S1**). La filogenia de U6 se basó en las 20 nuevas secuencias analizadas en el presente estudio, junto con otras 246 previamente publicadas. Los análisis para M1 parten de nuestras cinco secuencias y otras 109 tomadas de la literatura. Finalmente, la filogenia del clado L comprende 1.802 secuencias completas, incluyendo 37 nuevas procedentes de nuestras muestras. El macro-haplogrupo L fue dividido en los sub-clados L1, L2 y L3. Tras la aproximación inicial para la construcción de la *network* (Bandelt et al., 1999), la filogenia fue refinada a mano siguiendo un principio de parsimonia. Las ambigüedades en la construcción del árbol se resolvieron ponderando las mutaciones mitocondriales con respecto a su ocurrencia en la filogenia global humana (Soares et al., 2009). Debido a que no se han considerado fenómenos *indel*, nuestra filogenia reconstruida no concuerda, en ciertos casos puntuales, con lo expuesto en *PhyloTree* (van Oven & Kayser, 2009). Un ejemplo sería la asociación entre las sub-ramas L3b, L3c, L3d y L3f.

Los TMRCA para haplogrupos y sub-haplogrupos se calcularon mediante el estadístico ρ –empleando todos los polimorfismos del genoma y solo las mutaciones sinónimas– y también mediante estimas de Máxima Verosimilitud (ML) (ver detalles en la sección II-6.2). Para estos cálculos, la compilación de mitogenomas se detalla en el **Dataset S1**. Para las estimas de edad basadas en ρ , se empleó una tasa de mutación para el genoma mitocondrial completo de 1 sustitución/3.624 años y una tasa de mutación sinónima de 1 sustitución /7.884 años (Soares et al., 2009). Los errores estándar se estimaron de acuerdo con Saillard et al. (2000). Las estimas por ML se calcularon usando el programa PAML 3.13 (Yang, 1997) y se transformaron en escala temporal usando el mismo reloj molecular que en el caso anterior (Soares et al., 2009).

2.3. Análisis fundador (FA) y Bayesian Skyline Plots (BSPs)

Para analizar el rango temporal y la direccionalidad de las migraciones humanas entre el África subsahariana, el norte de África y Europa, empleamos el *Análisis Fundador* (FA, Richards et al., 2000) (ver sección II-6.3). Este método asume una división estricta entre las hipotéticas poblaciones ‘fuente’ y ‘sumidero’ que se basa en dos criterios ($f1$ y $f2$) para identificar a las secuencias fundadoras y tener en cuenta parcialmente la homoplasia y las migraciones de retorno entre regiones (*back migrations*). Por tanto, se trata de asegurar que las identidades de secuencia no se encuentran en los *tips* de la filogenia de la población fuente. El primer paso es reconstruir una *network* HVS-I (rango 16051-16400 de la rCRS) con los datos poblacionales presentados en la **Tabla S4**. Posteriormente se identifican a los fundadores y descendientes, empleando una herramienta informática desarrollada por el equipo de la Dra. L. Pereira (Alves et al., 2012).

Estos fundadores candidatos se identifican en las *networks* buscando nodos compartidos entre las poblaciones fuente y sumidero. Los dos criterios $f1$ y $f2$, consideran que los nodos o haplotipos compartidos deben tener al menos una ($f1$) o dos ($f2$) ramas derivadas de la hipotética población fundadora en la población sumidero. La diversidad privativa de cada fundador es convertida en años usando el estadístico ρ , asumiendo una tasa de mutación de 16.677 años (Soares et al., 2009). Cada fundador será asociado probabilísticamente a diferentes tiempos de migración hipotéticos, empleando una aproximación Bayesiana tal y como se ha realizado anteriormente (Richards et al., 2000). En este contexto, tuvimos en cuenta dos aproximaciones, la primera de ellas, definida por intervalos de 200 años entre migraciones desde 0 a 50 ky con el objetivo de visualizar la distribución global de las edades de los fundadores sin ningún modelo predeterminado. Esto proporciona un escenario que permite definir mejor los modelos de migración (Soares et al., 2012; Rito et al., 2013). Usando esta información, junto con datos paleoclimatológicos, arqueológicos e históricos, definimos modelos de migración empleando tiempos predeterminados de migración. Esto permite estimar estadísticamente el porcentaje de linajes que se encuentran asociados a cada una de las migraciones establecidas. Los análisis se desarrollaron con dos tiempos de migración, uno en el Holoceno temprano (definido como 8 ky) y el otro, en tiempos históricos (definido como 0,5 ky). Un modelo simplista que use estos dos eventos permitirá clasificar a los linajes fundadores en uno de los tiempos de migración y definir el porcentaje de linajes que entraron en cada uno de ellos. Las estimas de incertidumbre se calcularon siguiendo lo descrito en otros trabajos (Soares et al., 2012; Rito et al., 2013).

Para evaluar el flujo génico, los FA se desarrollaron usando información de los linajes U6 y L asumiendo varios escenarios. En primer lugar, se analizó la llegada de linajes L desde el África subsahariana al norte de África, a la Europa mediterránea (Iberia, Italia y Grecia) y, más específicamente, a Iberia. Es recomendable usar toda África como población fuente del haplogrupo L ya que la mayor parte de los fundadores se encuentran en el África subsahariana.

Con el fin de evaluar la población fuente y sumidero más probables, ambos lados de la cuenca mediterránea fueron considerados independientemente como áreas fuente para las secuencias U6, tal y como se llevó a cabo previamente para los haplogrupos R0a, HV1 y R2, en cuanto a un posible origen de estos, en el Oriente Próximo o en la Península de Arabia (Al-Abri et al., 2012). Debido a que el haplogrupo M1 supone una contribución menor (tanto por frecuencia como por distribución geográfica) sobre la influencia africana en Europa, este linaje fue omitido de los análisis FA.

También se desarrollaron FA para secuencias completas U6 considerando el norte de África e Iberia como población fuente y sumidero y viceversa, usando un criterio $f1$. Debido a que el FA requiere un valor único de tasa de mutación, no se implementó una curva para la tasa de mutación, dependiente del tiempo, sino que se empleó el valor central entre la tasa de mutación respecto a la estima de fundador más joven y antigua, siguiendo lo propuesto en anteriores trabajos

(Fernandes et al., 2015). La tasa de mutación usada fue de 1 mutación cada 2.643 años para el genoma completo mitocondrial.

Finalmente, para evaluar cambios en el tamaño poblacional efectivo (N_e) a través del tiempo en tres áreas geográficas clave (África subsahariana, norte de África e Iberia) se obtuvieron *Bayesian Skyline Plots* (BSPs) para las secuencias completas pertenecientes al haplogrupo U6 (ver sección II-6.4). Los códigos y referencias de las muestras usadas se presentan en el **Dataset S1**.

3. Resultados

3.1. *Linajes mitocondriales africanos en Iberia y áreas próximas. Patrones filogeográficos*

La **Tabla IV-1** presenta la estructura de los haplogrupos del ADNmt de origen africano detectados en los 750 individuos autóctonos de Andalucía occidental y oriental. En la misma tabla se muestran los datos mitocondriales de 3 poblaciones bereberes de Marruecos publicadas previamente por Coudray et al., 2009. Los andaluces de Huelva registran una influencia distintiva africana (11,8% de la variabilidad total), comparado con lo encontrado en la región oriental de Andalucía (Granada), donde la proporción de los haplogrupos africanos es mucho menos pronunciada (3,6%). Los análisis de estructura poblacional nos revelan que existen diferencias significativas entre ambas sub-poblaciones andaluzas para estos perfiles mitocondriales africanos ($F_{ST} = 0,0604$; $P = 0,0090$). El notable incremento del tamaño muestral andaluz llevado a cabo en este Capítulo en relación con las 279 secuencias analizadas en el **Capítulo III**, refuerza la alta heterogeneidad detectada en la variación global del ADNmt entre el oeste y este de Andalucía.

Dentro de la Península Ibérica, las mayores frecuencias del haplogrupo U6 (7,50%) se detectan en la provincia de Huelva. Cabe destacar que los bereberes de Asni, asentados en las laderas de la Cordillera del Atlas también portan una mayor proporción de secuencias U6 (11,32%) que lo observado en las poblaciones de Bouhria (1,43%) y Figuig (3,19%), localizadas en el flanco noroeste de Marruecos.

El sub-haplogrupo U6a es prevalente en el sudoeste de España. Comparativamente, nuestros datos reflejan los valores más elevados en la muestra de Huelva (6,43%) y en bereberes Asni (9,43%). El número medio de diferencias por parejas (MNPD) dentro de U6a –basado en secuencias HVS-I– oscila entre 2,0 en el sur de Iberia (Pereira et al., 2004; Casas et al., 2006; presente estudio) hasta 2,2 en población nativa marroquí (Rando et al., 1998; Brakez et al., 2001; Falchi et al., 2006; Coudray et al., 2009; Rhouda et al., 2009; Turchi et al., 2009; Harich et al., 2010). El linaje U6bd se encuentra mucho más restringido a través de Iberia peninsular, y se observa, además, a bajas

frecuencias (0-1,47%) (Richards et al., 2000; González et al., 2003; Pereira et al., 2004).

	Andalucía occidental (Huelva, n=280)		Andalucía oriental (Granada, n=470)		*Asni (n=53)		*Bouhria (n=70)		*Figuig (n=94)	
	n	%	n	%	n	%	n	%	n	%
U6	21	7,50	7	1,49	6	11,32	1	1,43	3	3,19
U6a	18	6,43	5	1,06	5	9,43	1	1,43	3	3,19
U6bd	2	0,71	2	0,43	1	1,89	-	-	-	-
U6c	1	0,36	-	-	-	-	-	-	-	-
M1	1	0,36	3	0,64	2	3,77	3	4,29	2	2,13
L	11	3,93	7	1,49	12	22,64	9	12,86	42	44,68
L0	-	-	1	0,21	-	-	-	-	4	4,26
L1b	4	1,43	-	-	4	7,55	5	7,14	6	6,38
L2a	2	0,71	1	0,21	1	1,89	2	2,86	3	3,19
L2b	3	1,07	-	-	2	3,77	-	-	-	-
L2d	-	-	-	-	-	-	-	-	1	1,06
L3b	-	-	-	-	-	-	-	-	8	8,51
L3d	-	-	2	0,43	-	-	-	-	-	-
L3e	-	-	-	-	4	7,55	2	2,86	20	21,28
L3f1b	2	0,71	-	-	-	-	-	-	-	-
L3h1b	-	-	2	0,43	-	-	-	-	-	-
L3x	-	-	1	0,21	-	-	-	-	-	-
L5	-	-	-	-	1	1,89	-	-	-	-
Linajes africanos (n)	33	11,79	17	3,62	20	37,74	13	18,57	47	50,00
Linajes africanos (secuencias completas)	19		13		7		7		16	

Tabla IV-1. Tipos y frecuencias de los linajes mitocondriales africanos detectados en Andalucía. Los datos de población bereber de Marruecos (resaltados con un asterisco, Coudray et al., 2009) se muestran con fines comparativos.

Con respecto al linaje L, los sub-clados L1 (L1b), L2 (L2a, L2b) y L3 (L3d, L3f, L3h, L3x) permiten distinguir de nuevo entre el oeste (3,92%) y el este (1,28%) de Andalucía. En las muestras bereberes, L1b registra un valor promedio de un 7% (6-8%). En bereberes de Figuig, L3e aparece notablemente representado (21,3%) un dato que no encuentra además paralelismo en las otras dos poblaciones norteafricanas (3-8%) estudiadas. Los orígenes evolutivos de los linajes L0, L1, L2 y L3 se localizan en el África subsahariana (Kivisild et al., 2006a). La presencia de marcadores L en el norte de África indica que el desierto del Sahara no ha sido una barrera infranqueable para las poblaciones humanas que lo bordean (Harich et al., 2010; Podgorná et al., 2013; Fortes-Lima et al., 2014, entre otros).

El haplogrupo M1 se encuentra débilmente representado en Andalucía (Iberia: 0-2,1%) mientras que en población bereber su frecuencia oscila entre 2,1% y 4,3%.

La **Figura IV-1** presenta los patrones de distribución espacial de los linajes mitocondriales U6a y L1b detectados a partir de un gran número de poblaciones de la región mediterránea y otras regiones seleccionadas (ver localidades muestreadas en la **Figura S2**). Como se puede observar en la **Figura IV-1A**, los valores más elevados de U6a se encuentran principalmente en el Magreb y Mauritania. La radiación U6a alcanza Europa, principalmente a través de la Península Ibérica. La presencia de U6a en otras poblaciones mediterráneas europeas (p.e. Italia) es mucho menor (Ottoni et al., 2009). La **Figura S3** muestra la filogeografía del haplogrupo U6 como un todo.

Sobre el patrón espacial de L1b (**Figura IV-1B**) destaca su alta incidencia en el África occidental desde el Golfo de Guinea hacia el noroeste del Magreb. La presencia de L1b es perceptible en el sudoeste de Iberia (1,4%, presente estudio; Iberia: 0-2,8%) y también en el sur de Italia y Sicilia (0,7-1,1%) (Ottoni et al., 2009). Las **Figuras S4-S10** presentan los mapas de contorno para el haplogrupo M1 y algunos sub-clados dentro de L.

Las relaciones genéticas entre las secuencias U6a y L1b (haplotipos HVS-I) se presentan en la **Figura S11**. Parece notorio el gran nivel de identidad entre el norte de África y la Península Ibérica como se muestra en los grandes nodos. La *network* de L1b (**Figura S11B**) revela que este linaje se encuentra presente principalmente en el noroeste de África (con Marruecos como el mayor representante) y, también, en población ibérica (p.e. andaluces de Huelva y habitantes autóctonos de la isla portuguesa de Madeira), italiana y egipcia. La variación de los clados U6a y L1b y las estructuras *star-like* de sus *networks* serían consistentes con expansiones poblacionales humanas a través del África mediterránea y Europa. Sin embargo, debemos tener en cuenta el menor nivel de resolución que proporcionan las secuencias HVS-I comparado con el obtenido por la información de genomas completos (ver siguiente sección).

3.2. Secuenciación completa del ADNmt. Filogenias y estimas de coalescencia

La determinación de la compleja historia de los linajes maternos africanos, particularmente en Iberia y en otras poblaciones de la costa mediterránea norte, requiere un análisis más detallado. Hemos abordado este objetivo mediante el análisis de la variación mitocondrial total de 32 andaluces autóctonos de las provincias de Huelva y Granada, los cuales representan una fracción de la muestra total de Andalucía perteneciente a los haplogrupos U6, M1 y L (50 individuos de un total de 750 muestras andaluzas). Siguiendo las mismas condiciones, 30 muestras bereberes de Marruecos de Asni, Bouhria y Figuig han sido analizadas. La **Tabla S5** presenta los haplotipos mitocondriales completos de las 62 muestras.

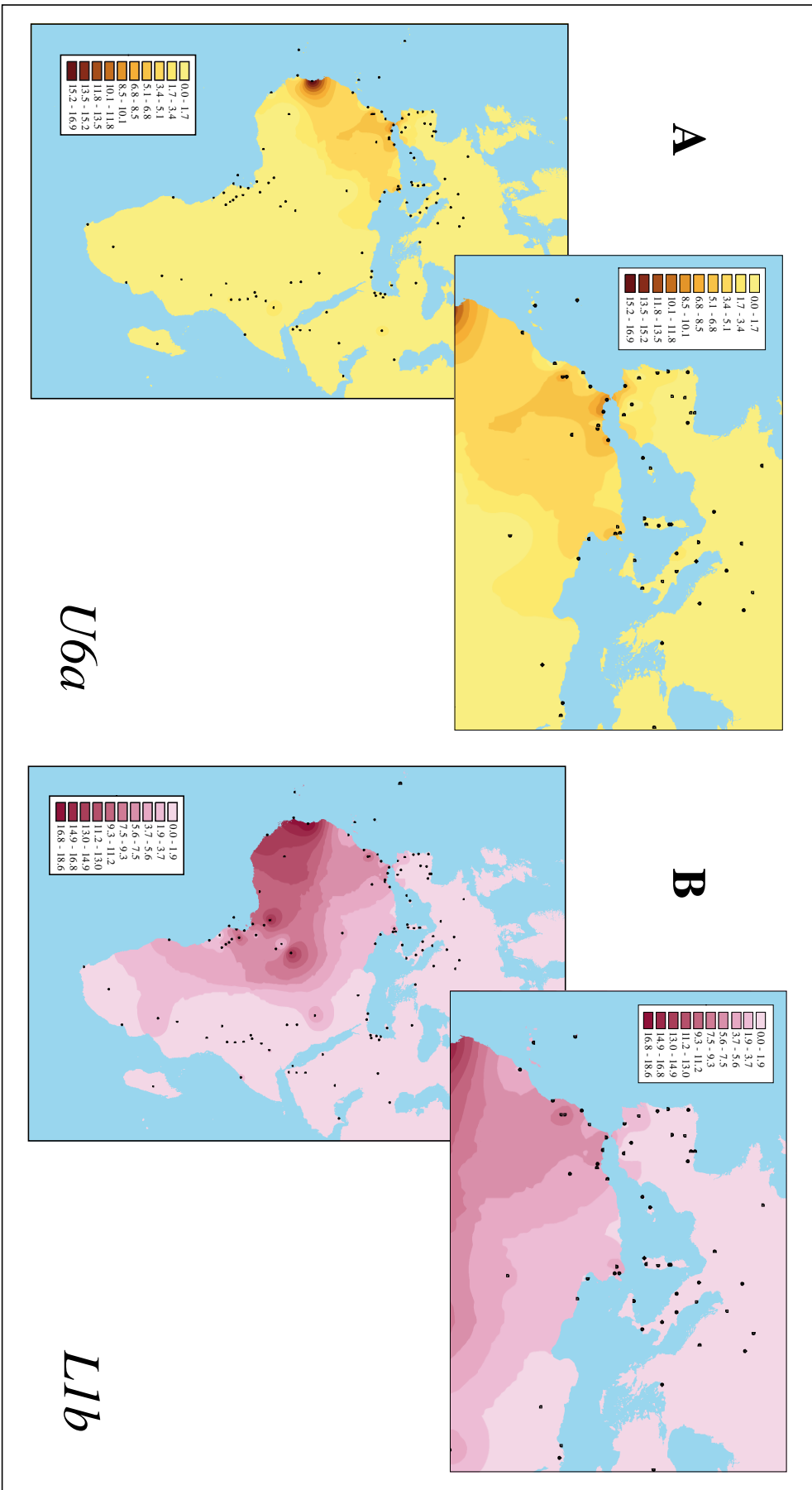
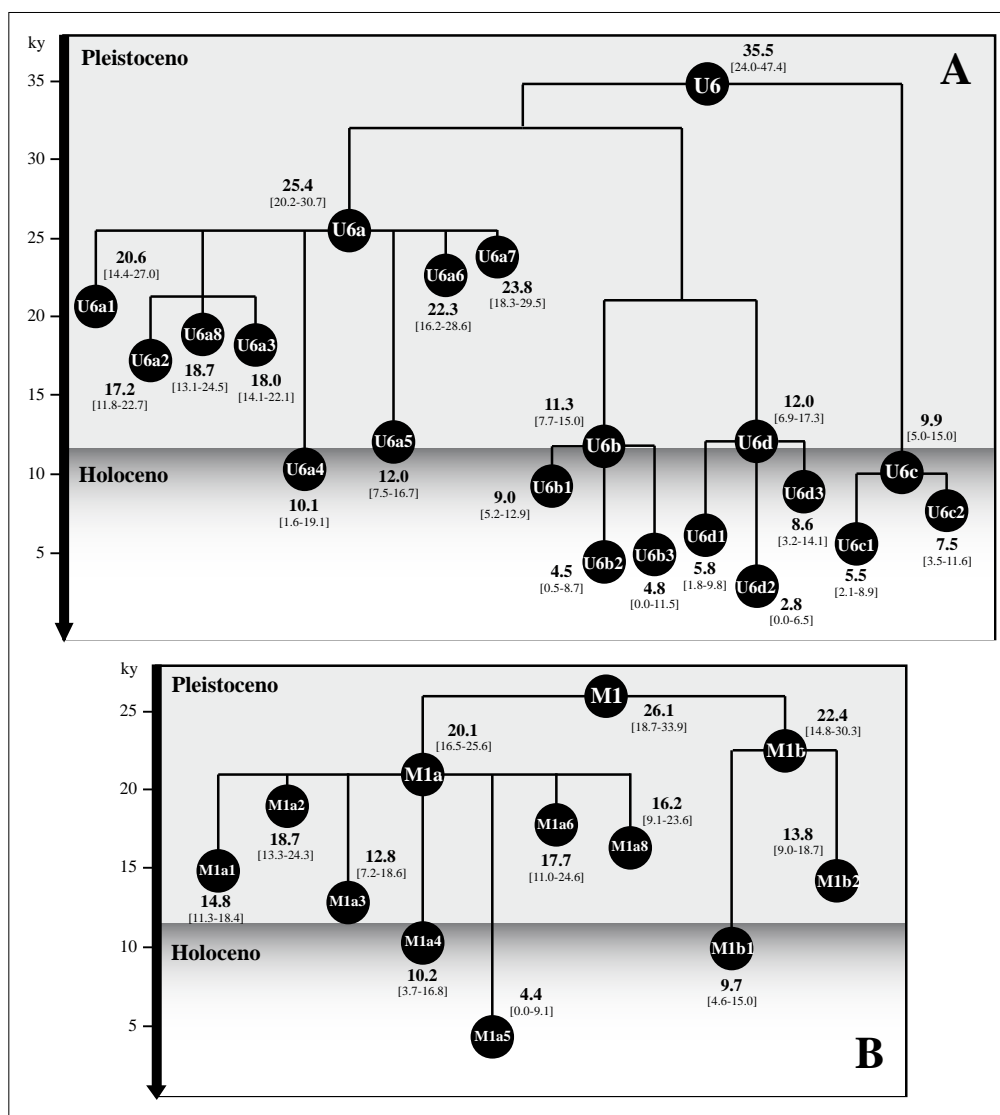


Figura IV-1. Mapas de contornos de los sub-haplogrupos U6a (A) y L1b (B). Las frecuencias se expresan en porcentajes (ver escala de color). La localización de las poblaciones y las referencias de las mismas se encuentran en la **Figura S2** y en la **Tabla S4**.

En la **Figura IV-2** se observan los árboles filogenéticos esquemáticos inferidos para los haplogrupos U6 y M1, con las estimas de edad (tiempos de coalescencia) calculadas por Máxima Verosimilitud (ML) así como los intervalos de confianza para todas las radiaciones. En los **Datasets S2-S6** se exponen los árboles filogenéticos actualizados y refinados para U6, M1, L1, L2 y L3, basados en un alto número de mitogenomas (n= 2.182; el número medio de muestras analizadas por haplogrupo son 436).

Figura IV-2. Filogenia simplificada de los haplogrupos U6 (A) y M1 (B). El árbol proporciona estimas de ML en miles de años (ky) y los intervalos de confianza al 95% para los principales clados y sus radiaciones.



El nodo superior del haplogrupo U6 (**Figura IV-2A**) coalesce a 35,5 ky (intervalos de confianza 24,0-47,4) en el Paleolítico Superior temprano (Zilhão & D’Errico, 1999). Este resultado es consistente con lo presentado en recientes estudios (Pereira et al., 2010a; Pennarun et al., 2012; Secher et al., 2014). El clado

U6a representa la radiación más profunda, hace 25,4 ky (20,2-30,7), desembocando en numerosas ramas y clados jerárquicos anidados. U6a es el clado U6 más frecuente, con una representación predominante de muestras magrebíes (Marruecos) (ver árbol filogenético completo en **Dataset S2**).

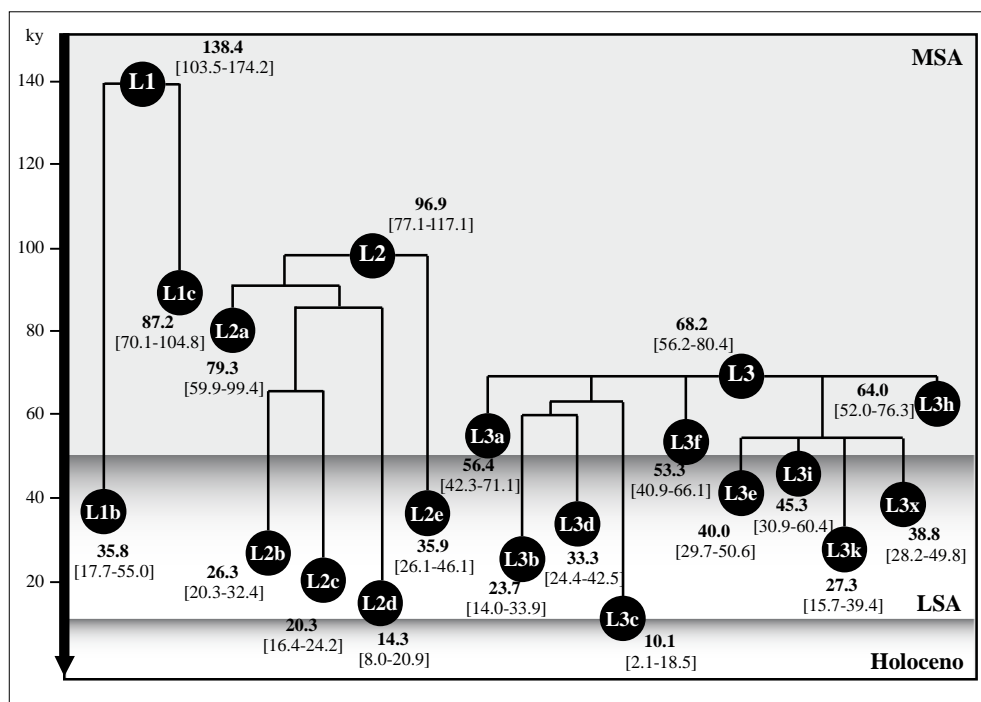
Una inspección más detallada de la filogenia de U6 (**Dataset S2**) revela que no resulta sencillo inferir si el norte de África o Iberia portan los linajes más basales. La mayoría de sub-haplogrupos tienen linajes que descienden directamente de los nodos. El caso más extremo, es una muestra marroquí que se encuentra en la raíz de U6a, mientras que el resto se asigna a diferentes sub-linajes. U6a6 podría ser un buen candidato norteafricano, ya que la mayoría de sus muestras pertenecen al Magreb (6 de Marruecos y 2 de Túnez). Este sub-linaje tiene una edad de coalescencia de 22,3 ky (16,2-28,6), coincidiendo con uno de los períodos de expansión de U6 en el norte de África (Pereira et al., 2010a). Dentro de este clado, dos muestras españolas (una de Granada, presente estudio, y otra de Asturias) son derivadas, procediendo de nodos que datan 4,8 ky (0,7-9,0) y 3,3 ky (0,0-7,3). La introducción más probable desde el norte de África estaría localizada en un período anterior a 4,0 ky.

Descendiendo aún más en la distribución de muestras a través de la filogenia U6, es posible afirmar que solo los linajes U6 derivados se encuentran ampliamente distribuidos a través del norte de África, Iberia, e incluso el África subsahariana atlántica. La dispersión hacia el resto de Europa es mucho más restringida. Nuestro árbol construido a partir de mitogenomas U6 muestra que U6a1 es predominantemente europeo, ya que contiene un número significativo de secuencias de individuos mediterráneos, principalmente de la costa noroeste, con una contribución notable de Iberia (21 de las 29 muestras europeas) y, además, tiene un clado ancestral en Portugal (*Accession Number* HQ651694). Solo unas pocas secuencias magrebíes forman el *cluster* U6a1 (7/39, 18.0%). En lo que concierne al U6a2 (17,2 ky [11,8-22,7]), éste se encuentra altamente representado en Etiopía y muy disperso en el Oriente Próximo. El U6a3 (18,0 ky [14,0-22,1]) comprende 44 mitogenomas e incluye algunas ramas típicas del África subsahariana, otras más comunes en población ibérica y judía, y una secuencia del Oriente Próximo. Dentro del sub-clado U6a3b y su derivado U6a3b1 se puede observar cómo las secuencias ibéricas y magrebíes están bastante relacionadas. Los clados U6c (9,9 ky [5,0-15,0]) y U6d (12,0 ky [6,9-17,3]) están presentes a bajas frecuencias en Iberia, Europa y el norte de África. El U6a5 (12,0 ky [7,5-16,6]) es común en el África subsahariana y está presente en Túnez e Italia (2 de las 6 muestras proceden de Sicilia), indicando una posible ruta desde Túnez hacia Italia, incluso cuando las muestras tunecinas e italianas no se encuentran en la misma sub-rama. Estos datos pueden estar relacionados con la destrucción de Cartago (149-146 BC) o con la expansión Islámica en Sicilia (902-1091). El U6a7a (7,0 ky [2,4-11,7]) y el U6b (11,3 ky [7,7-15,0]) tienen ramas basales dispersas por el norte de África, Iberia, Oriente Próximo y el África subsahariana atlántica. Algunas radiaciones de U6a7 han sido relacionadas recientemente con episodios históricos (Secher et al., 2014). La presencia de secuencias judías dentro de U6a7a podría trazar la diáspora judía (sefardí) desde España (siglo XV) hacia otros territorios de Europa occidental y fuera del

continente. Un individuo andaluz de Huelva enriquece este *cluster* con una secuencia U6a7a1b (*Accession Number* KT819220). Sin embargo, la señal filogeográfica más potente aparece asociada con el linaje U6b1a, con una edad de coalescencia estimada de 3,0 ky (1,0-5,1), datación próxima a la colonización de las Islas Canarias ($2,49 \pm 0,6$ ky) (Onrubia-Pintado, 1991). Con una alta frecuencia entre población nativa, U6b1a parece ser una rama hermana de la expansión magrebí (Pennarun et al., 2012; Secher et al., 2014).

La filogenia simplificada de M1 (**Figura IV-2B**) también muestra cierta complejidad, aunque notablemente menor que la del árbol de U6, y los resultados que se ofrecen son diferentes tanto respecto a su pasado evolutivo como a su distribución geográfica. El clado M1 tiene una edad de coalescencia menor que U6 (26,1 ky [18,7-33,9]), seguida próximamente de sus dos principales radiaciones M1a (21,0ky) y M1b (22,4 ky) (ver **Tabla S6**). La filogenia de M1 muestra más representantes de África oriental (Etiopía y áreas circundantes del Cuerno de África) y fuera de la cuenca mediterránea (p.e. Cáucaso y sudoeste de Asia) (ver distribución espacial de M1 en la **Figura S4**). Unas pocas muestras ibéricas se encuentran dispersas a través del árbol filogenético de M1 ($n= 9/114$), sin mostrar una asociación clara con secuencias magrebíes. Estos resultados apoyan llegadas puntuales a la Península Ibérica asociadas con migraciones menores desde el noroeste de África, donde M1 registra frecuencias elevadas (5,3% de media), pero también desde el Mediterráneo oriental y el Oriente Medio (ver filogenia completa de M1 en **Dataset S3**).

Figura IV-3. Filogenia simplificada del haplogrupo L. El árbol proporciona estimas de ML en miles de años (ky) y los intervalos de confianza al 95% para los principales clados y sus radiaciones. MSA: *Middle Stone Age*, LSA: *Late Stone Age*.



Las edades de coalescencia de los sub-haplogrupos L (L1, L2 y L3) se muestran en la filogenia esquemática recogida en la **Figura IV-3**. El principal *input* del haplogrupo L a Europa corresponde a L1b, que es la rama más joven (35,8 ky [17,7-55,0]) de su clado parental L1 (138,4 ky [103,5-174,2]). L1b sería, por tanto, contemporáneo a U6. La composición de L1b (ver árbol filogenético de L1 en **Dataset S4**) muestra una prevalencia de ramas africanas occidentales, aunque son visibles algunas interesantes radiaciones hacia latitudes más septentrionales.

Los marcadores más notables del linaje L dentro de Europa corresponden a los sub-clados L1b1a6 (6,9 ky), L1b1a8 (12,7 ky), L1b1a12 (7,8 ky) y pre-L1b1a16 (10,0 ky). Cabe señalar que nuestra estructura de ramas para L1b1a6 difiere de la mostrada en Cerezo et al., 2012. Este sub-clado ha sido enriquecido con una secuencia andaluza (de Huelva) y otras 3 marroquíes (una secuencia por cada población bereber). Otros datos novedosos de nuestro trabajo se refieren a las filogenias de L1b1a8 y L1b1a12, las cuales se han modificado de nuevo con respecto a Cerezo et al., 2012. El primer sub-haplogrupo contiene secuencias exclusivamente del sur y noroeste de España (Andalucía y Galicia, respectivamente) y Marruecos (bereberes de Figuig); el último sub-clado (L1b1a12) está compuesto, sin embargo, de muestras ibéricas y del resto de Europa. Por otro lado, el *cluster* pre-L1b1a16 (L1b1a9 para Cerezo et al., 2012) engloba mitogenomas del oeste del Mediterráneo [Iberia (JN214438 y JN214447) y Marruecos (EU092667)] y otros de la Europa mediterránea central (JN214460, JN225465 y JN225466). Por tanto, algunas ramas de L1b serían compatibles con una evolución *in situ* dentro de Europa tras la introducción inicial desde el continente africano vía norte de África. Como la mayoría de muestras europeas se encuentran dentro de L1b1, el tiempo inicial de migración se limita a la edad de este clado (11-13 ky). Otras secuencias ibéricas y europeas se encuentran dispersas por los clados L2 y L3 mostrando robustas señales filogeográficas asociadas (ver filogenias completas en **Datasets S5** y **S6** y mapas de contorno en **Figuras S5-S10**). La posición de las muestras ibéricas dentro de las filogenias L2 y L3 es de gran complejidad. En la mayoría de los casos, las secuencias ibéricas no se agrupan con las norteafricanas, sino con las del Oriente Medio o de Europa. Además, las muestras españolas presentan una tendencia a asociarse en los árboles, a pesar de que procedan de distantes regiones peninsulares. Cabría la posibilidad de que alguna de estas secuencias mitocondriales españolas pudieran corresponder a descendientes de emigrantes a Hispanoamérica que retornaron a la Península tras mezclarse con población afro-americana. Las muestras portuguesas son escasas, y ninguna de ellas pertenece a L3. El origen europeo propuesto para el sub-clado L2a1k (Malyarchuk et al., 2008; Cerezo et al., 2012), con una edad de coalescencia de 10,3-10,6 ky, no está claro en nuestra filogenia reconstruida ya que hemos adscrito 3 muestras bereberes a esta rama. Por tanto, la presencia de dos ramificaciones paralelas, la europea y marroquí, oscurece el origen de este linaje. Teniendo en cuenta la diversidad más profunda de L2a1 en África, la situación más probable sería una introducción desde el norte de África a Iberia.

3.3. *Linajes mitocondriales africanos en el Mediterráneo: marco temporal y dispersiones*

Las filogenias examinadas muestran trazas antiguas de linajes africanos maternos en la Península Ibérica y nos permiten establecer límites superiores a la entrada de estos linajes en la región. Sin embargo, para descender en las estimas temporales y en la direccionalidad de las migraciones a través de la cuenca mediterránea, hemos empleado la metodología del análisis fundador (FA). En primer lugar, aplicamos el FA a la diversidad de las secuencias HVS-I con lo que se consigue equilibrar el menor poder informativo de la región control con el mayor número de muestras disponibles. La **Figura IV-4** muestra los resultados de FA para el haplogrupo L. El norte de África, Europa mediterránea e Iberia se asumieron como poblaciones sumidero frente al África subsahariana como población fuente. La introducción más intensa de linajes L en el norte de África (**Figura IV-4A**) parece haber ocurrido en dos puntos, uno de los cuales es muy reciente (pico a 0 ky), mientras que el otro se detecta alrededor del período del Óptimo Climático del Holoceno, HCO (6,6 ky – criterio *f1* – y 11,8ky – criterio *f2*). Cuando se imponen dos períodos de migración, el primero a 0,5 ky (representando la trata de esclavos desde el África subsahariana occidental al Magreb y otros acontecimientos contemporáneos) y otro a 8,0 ky (hacia la mitad del HCO) es posible determinar (ver **Figura S12**) que el período más antiguo es responsable de la migración de 2/3 de los linajes subsaharianos L presentes hoy en el norte de África; otros eventos más recientes introdujeron 1/3 de los mismos. El patrón de llegada de linajes L a Iberia o a la Europa mediterránea es similar, es decir, los episodios prehistóricos fueron los mayores contribuyentes de la presencia subsahariana en estas dos áreas geográficas (**Figura IV-4B, C**).

Una inspección detallada de los fundadores L introducidos en el norte de África muestra que la mayoría de ellos alcanzaron esta región en el período más antiguo (ver **Figura S13**). En cuanto a los linajes que llegaron a la Europa mediterránea, se observa que el haplotipo L1b claramente domina el *pool* de los linajes L. Además, el 79,5% (*f1*) y el 81,5% (*f2*) de los linajes que sugieren una entrada en el Holoceno temprano (el Holoceno representa los últimos 11.700 años antes del año 2.000 AD, Walker et al., 2009) a Europa también muestran una entrada en el mismo período al norte de África, indicando que la secuencia de migración fuera África subsahariana – norte de África – norte del Mediterráneo. De hecho, se detectó una cercanía genética de los mitogenomas L entre el norte de África y Europa, especialmente, una identidad de linajes con Iberia.

Por el contrario, solo el 24,2% (*f1*) y el 20,3% (*f2*) de los linajes L que entraron en Europa en el período histórico concuerdan con aquellos que llegaron al norte de África, lo que sugiere varias rutas independientes de introducción. La entrada de estos linajes a Europa directamente desde el África subsahariana en el período histórico podría estar relacionada con el tráfico de esclavos en ese Continente.

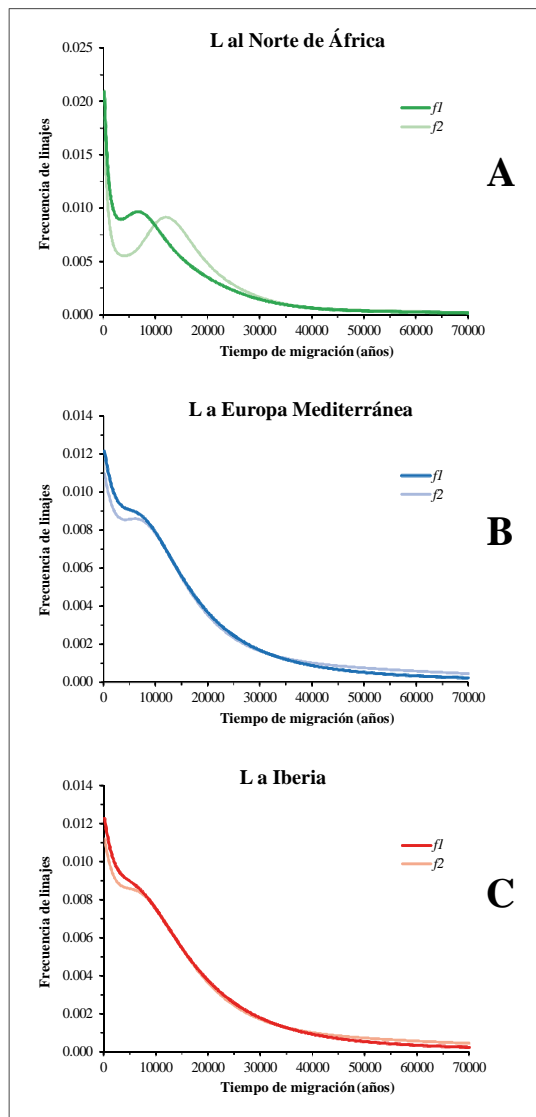
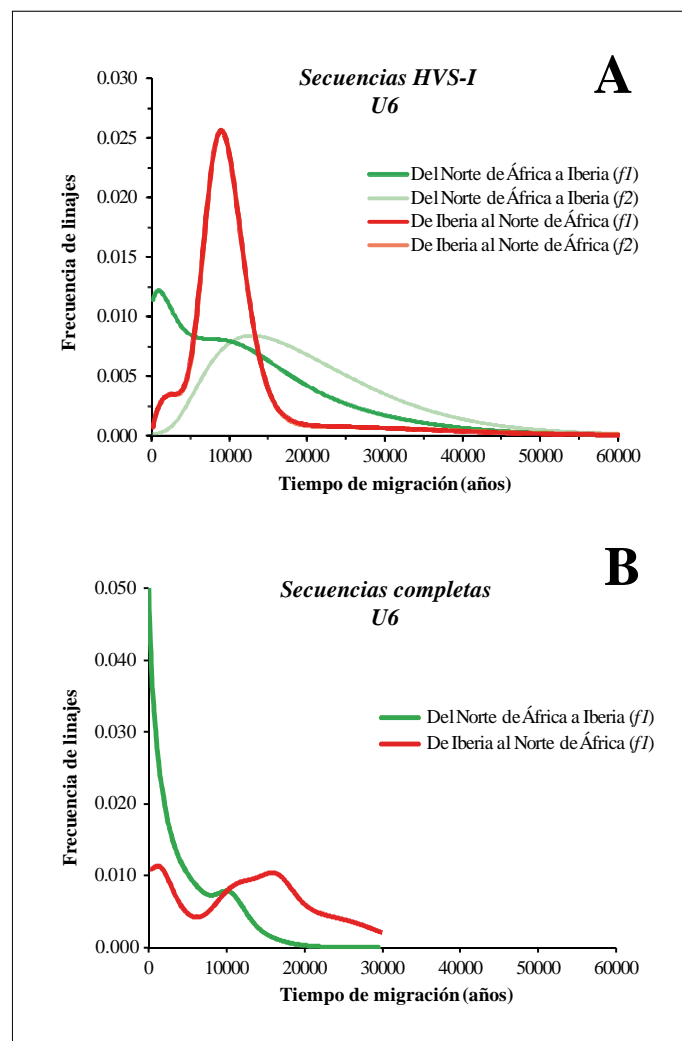


Figura IV-4. Análisis fundador para el haplogrupo L. Los gráficos muestran las distribuciones probabilísticas de *clusters* fundadores a través de tiempos de migración, según los criterios $f1$ y $f2$, para las expansiones de linajes L al norte de África (A), Europa mediterránea (B) e Iberia (C).

Los resultados de los FA para el haplogrupo U6 (basados en secuencias HVS-I) se muestran en la **Figura IV-5A**. Asumiendo la dirección más aceptada de migración desde el norte de África hacia Iberia, el criterio $f1$ muestra un pico principal, hacia 0,8-1,0 ky, precedido de una meseta hacia los 10 ky aproximadamente, mientras que $f2$ solo muestra un pico hacia los 13 ky. La migración en la dirección opuesta, es decir, de Iberia al norte de África, muestra un pico principal en los 9 ky y un pico menor hacia los 3 ky, para ambos criterios, $f1$ y $f2$ (ambas curvas se encuentran solapadas). Estos resultados nos indican que la diversidad de las secuencias HVS-I de U6 no es particularmente informativa para determinar qué lado de la cuenca mediterránea es la fuente más probable de migración y cuál sería la población sumidero para U6. La filogenia y distribución geográfica de U6 (ver secciones 3.1 y 3.2) apoyan esta afirmación.

Figura IV-5. Análisis fundador para el haplogrupo U6. Los gráficos muestran las distribuciones probabilísticas de *clusters* fundadores de U6 para secuencias HVS-I (A) y genomas completos (B), a través de tiempos de migración, escaneados a intervalos de 200 años desde 0-60 ky.



Dado que muchos polimorfismos de la HVS-I (como 16189, 16278, 16311) son altamente recurrentes dentro del haplogrupo U6, y confieren por tanto alta reticulación, se intentó determinar si las secuencias completas añadían más información para resolver este problema (**Figura IV-5B**). El norte de África y la Europa mediterránea muestran una notable imbricación de linajes a lo largo de casi todos los clados de U6, por lo que desarrollamos un FA considerando cada región como fuente/sumidero y viceversa. Este análisis pretende esclarecer dos puntos, en primer lugar, conocer qué región posee un mayor nivel de diversidad privada, lo que sería una indicación del punto de origen del haplogrupo U6, y en segundo lugar, estimar la edad de llegada a la población sumidero o receptora.

Cuando se comparan los fundadores a ambos lados del Mediterráneo, un mayor nivel de diversidad se observa en el norte de África, lo que evidencia que la dirección más probable de migración fue desde el norte de África a Iberia. Cuando se prueba este modelo (**Figura IV-5B**) se observan dos picos de migración, el más antiguo hacia los 10 ky y uno reciente, próximo al presente.

Para el pico más antiguo, U6a3, U6a6, y U6a8 serían los principales fundadores norteafricanos.

Cuando se intenta probar la dirección inversa de migración, esto es, de Iberia al norte de África, observamos un pico profundo hacia 20 ky y otro en el pasado reciente. Esta edad aparentemente más profunda de migración, aceptando un origen ibérico, podría ser reflejo, por el contrario, de una expansión más antigua de clados privados en el norte de África. Se debe tener en cuenta que el elevado nivel de identidad de linajes entre el norte de África e Iberia es el resultado de un proceso de intercambio genético complejo y continuado a través del tiempo, lo que imposibilita un análisis simplista.

La entrada estimada de los linajes U6 del norte de África a la Península Ibérica hacia 10 ky correlaciona bien con otros linajes africanos como L1b1a6, como se ha mostrado en este trabajo y en otros (Cerezo et al., 2012). Esto nos indicaría que los linajes U6 se movieron desde África hacia Iberia en el Holoceno temprano junto con algunos clados L.

También se analizó si Andalucía, cuyo extremo más oriental –el Reino de Granada– fue la última posesión islámica de Iberia, podría mostrar una proporción diferente de los linajes U6 a través de los dos picos de migración, desde el norte de África, en comparación con lo observado en general dentro de Iberia (ver **Figura IV-6**). El FA basado en secuencias HVS-I atribuye una edad de introducción de los linajes más antigua (en una proporción de 0,3 para eventos históricos frente a 0,7 para eventos post-glaciales) y proporciones similares (0,5) para Andalucía, mientras que el contenido de las secuencias completas muestra resultados equivalentes para el *pool* global ibérico, siendo los linajes U6 andaluces principalmente introducidos durante tiempos históricos (con una proporción de 0,7).

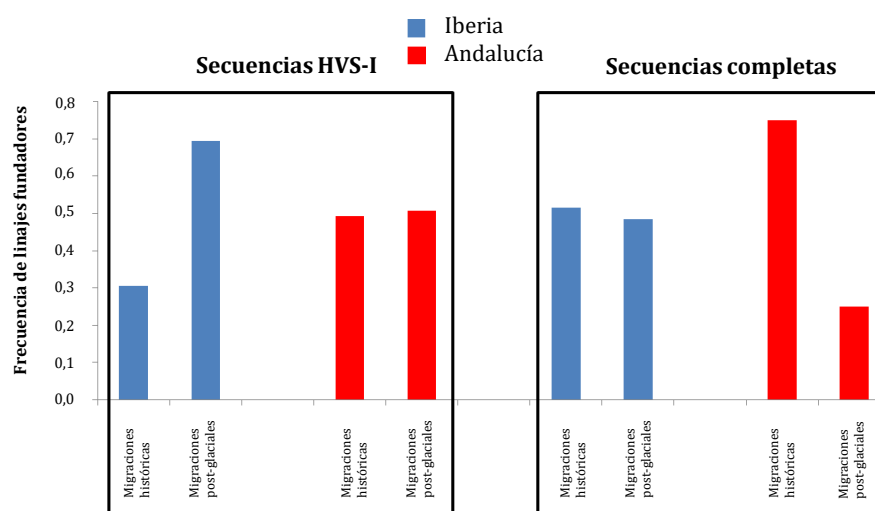


Figura IV-6. Proporción de linajes fundadores U6 en Andalucía y en Iberia. El análisis está basado en secuencias HVS-I (izquierda) y en mitogenomas (derecha). Se consideraron dos períodos migratorios (eventos históricos y post-glaciales).

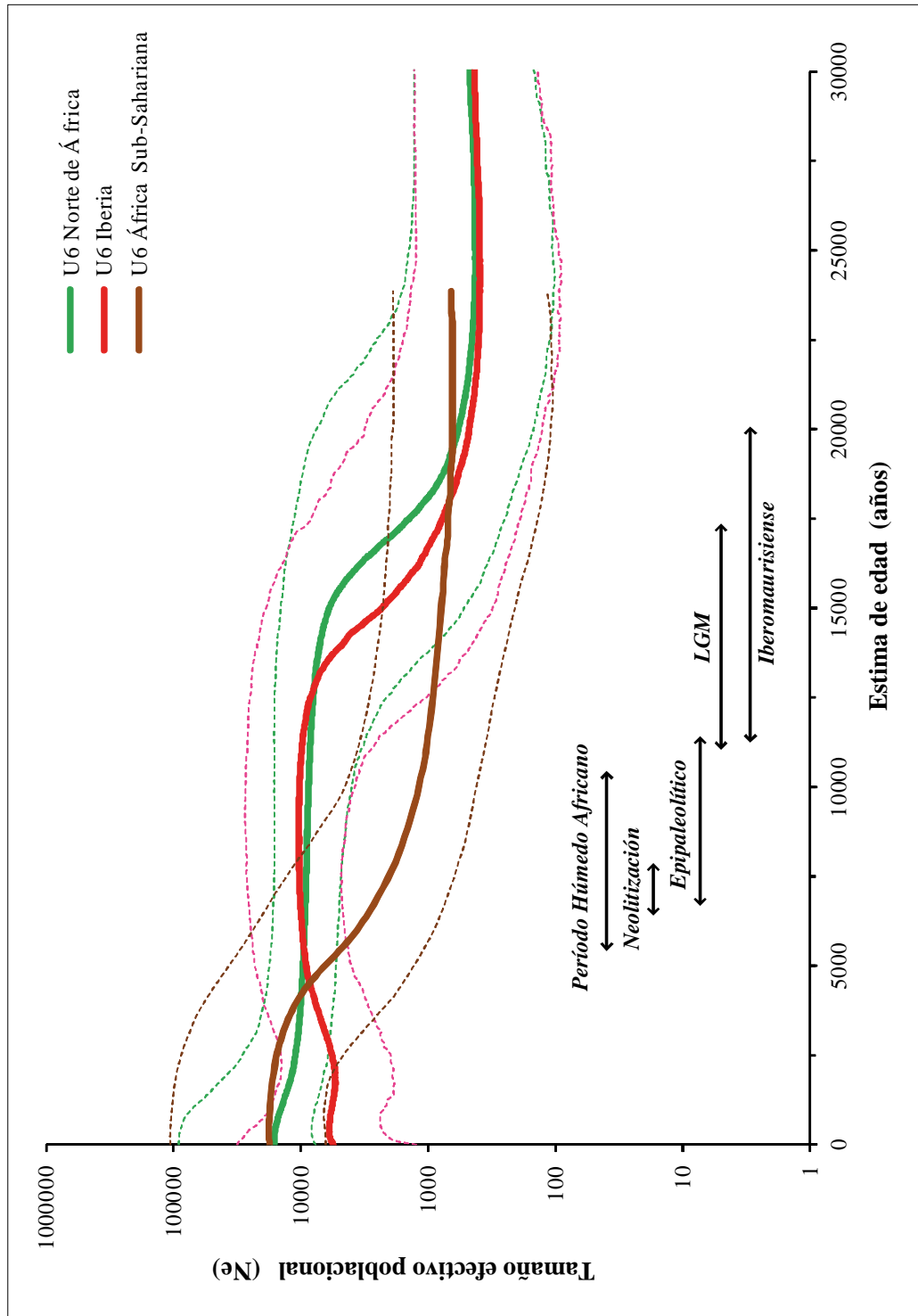


Figura IV-7. Bayesian Skyline Plots (BSPs) de genomas completos U6. Se indican los cambios temporales en el N_e en el África subsahariana (marrón), norte de África (verde) e Iberia (rojo). Las líneas sólidas representan las medianas a escala logarítmica del N_e en el eje Y dentro de cada región. Las líneas punteadas muestran los intervalos de confianza al 95%.

Con el objetivo de añadir más información acerca de qué región, si el norte de África o Iberia, muestra las señales más profundas en términos de diversidad U6, hemos construido *Bayesian Skyline Plots* (BSPs) basados en genomas completos (**Figura IV-7**). Mediante estas representaciones, hemos inferido las tendencias temporales de tamaño efectivo poblacional (N_e) para regiones geográficas clave: África subsahariana, norte de África e Iberia. Como se puede observar, la expansión de los linajes U6 ocurre primero en el norte de África, en el período post-glacial, seguida muy cercanamente por la expansión de las secuencias ibéricas (~2 milenios, aunque las fechas deben ser consideradas con precaución, ya que expansiones en las poblaciones sumidero pueden reflejar eventos demográficos en la población fuente, y también, por los respectivos intervalos de confianza). Estas expansiones ocurrieron durante el Iberomauriense (20,0-11,5 ky BP) seguidas por una fase de estabilidad a largo plazo. La expansión del *set* subsahariano ocurrió mucho más tarde, esto es, dentro del Período Húmedo Africano (AHP, 10,0-5,5 ky BP), probablemente como consecuencia de la migración del linaje U6 desde el norte de África, facilitada por las condiciones climáticas favorables de este período.

Cabe destacar que existe una disminución de los linajes U6 en Iberia en períodos más recientes (desde los 5-3 ky, y posteriormente se alcanza una meseta) que no se corrige con la llegada de linajes U6 durante la conquista romana o islámica de la región, tal y como indican los resultados de los FA.

4. Discusión

El intento de discernir el origen y el momento (tiempo) de las migraciones entre regiones cercanas, que comparten afinidades genéticas próximas, es un reto en el campo de la Genética de Poblaciones basada en marcadores uniparentales (Fernandes et al., 2015) y una compleja tarea para los estudios basados en un genotipado masivo autosómico (Hellenthal et al., 2014). La cuenca mediterránea se encuentra entre las zonas de mezcla genética más intensas del globo, debido a la posible ocurrencia de migraciones en ambas direcciones a lo largo de miles de años. A pesar de ello, unas pocas migraciones prehistóricas e históricas habrían tenido un papel predominante para esos episodios de flujo génico, siguiendo expansiones poblacionales y dispersiones culturales o políticas.

Los estudios basados en ADNmt se han centrado en el análisis de los linajes H1, H3 y V cuando se pretende determinar las migraciones desde el norte hacia el sur; mientras que los haplogrupos U6 y L se usan en la dirección opuesta, asumiendo que estos linajes son buenos representantes de la ancestría a cada lado de la cuenca mediterránea. Sin embargo, dadas las bajas frecuencias de los linajes U6 a lo largo del Mediterráneo, e incluso en muchas poblaciones bereberes (cuya ancestría se asume que representan), la posibilidad de que hubiera sucedido una migración norte-sur a través del mar Mediterráneo no ha sido evaluada científicamente. En este trabajo, gracias al incremento considerable

de secuencias completas U6 y L en Iberia (centradas en Andalucía, región que presenta altas frecuencias de estos clados) y en el noroeste de África, hemos proporcionado evidencias filogeográficas y de FA que arrojan luz a este punto.

Habitualmente, la historia genética de U6 se ha asumido como una migración desde el sudoeste de Asia a través del norte de África (Olivieri et al., 2006). Esta hipótesis se basa en el origen global de los sub-clados U en el sudoeste de Asia, región que habría sido el centro de la distribución geográfica de dichos sub-clados U: Europa, la India, Asia central, África oriental y el norte de África. Dos posibles escenarios pueden ser propuestos para explicar la aparición del primer haplotipo U6 (portando las mutaciones 3348 y 16172): i) estas mutaciones surgieron en la región fundadora, pero no dejaron ningún legado genético en sus actuales poblaciones; ii) ellas se originaron en algún lugar del norte de África, tras la llegada del haplotipo fundador U6. Dentro del norte de África, el U6 es solo frecuente en su extremo occidental (así como en el sudoeste de Europa). Además, todas las ramas basales se encuentran virtualmente restringidas a esta región (U6b, U6c y U6d), lo que podría indicar su origen occidental. Sin embargo, no puede excluirse el clado principal U6a, el cual muestra una gran riqueza de sub-clados en el noroeste de África (Pennarun et al., 2012) aunque algunas de sus ramas derivadas también incluyen secuencias de poblaciones del África oriental y de Oriente Medio (p.e. U6a2).

Una posible explicación para la concentración de U6 en el noroeste de África es su virtual desaparición de Egipto, dado el flujo génico continuo desde el Oriente Próximo así como debido a la extensa barrera física del desierto de Libia entre el Delta del Nilo y Túnez. Por tanto, esto convierte a los egipcios en una población genéticamente más próxima a los eurasiáticos que a otros norteafricanos, como se ha observado para el genoma nuclear (Fadhlaoui-Zid et al., 2013). Sin embargo, como hemos confirmado aquí con las filogenias actualizadas basadas en mitogenomas, las ramas U6b, U6c y U6d se encuentran entrelazadas entre el norte de África e Iberia. Esto conduce a que la mayoría de herramientas de inferencia filogeográfica básica sean ineficientes para determinar su origen. Por ello, desde una aproximación puramente filogeográfica, un tercer escenario sería teóricamente posible. El U6 podría haber emergido en Europa desde la raíz del haplogrupo U, como parece haber ocurrido para el U5 (Soares et al., 2010), pero, en contraposición a esta rama hermana, el U6 ocupa una región muy restringida dentro de Europa. Además, la distribución del U6 es limitada incluso dentro de la Península Ibérica, dado que este área ha sido una fuente probable de migraciones al resto de Europa durante el período post-glacial y el Neolítico (Gamble et al., 2004; Bocquet-Appel et al., 2005) y que incluso el U6 se encuentra prácticamente ausente fuera de la Europa mediterránea. Estas consideraciones implican que un origen europeo para el U6, aunque no imposible, parece poco probable. El incremento de U6 hace 20 ky, como ha sido detectado en el BSP, podría ajustarse a la emergencia de la industria Iberomauriense en el Magreb (Pereira et al., 2010a; Barton et al., 2013), sugiriendo también una evolución de U6 en el norte de África. Los FA confirman que el norte de África posee un mayor nivel de diversidad privada.

Además, nuestros resultados del FA revelan evidencias de migraciones de vuelta (*back-migrations*) al norte de África desde Iberia. Algunas sub-ramas del haplogrupo mitocondrial H han revelado movimientos post-glaciales hacia el continente africano desde la Península (Cherni et al., 2009; Ottoni et al., 2010). Otras fundadas evidencias, incluyendo datos arqueológicos e históricos, apoyan un escenario de flujo génico, presumiblemente bidireccional, entre el norte de África y la Península Ibérica. Marcados paralelismos entre las industrias Iberomauriense y el Magdaleniense español (~20-12,9 ky cal BP) (Villaverde et al., 2012) del Paleolítico superior podrían haber sido el resultado de la exploración y la ocurrencia de movimientos humanos de costa a costa. Tanto el Estrecho de Gibraltar como el mar de Alborán (una ruta marítima alternativa, al este del Estrecho) habrían representado escenarios clave para explicar estos contactos humanos tempranos (Souville, 1998; Linstädter et al., 2012). Además, la neolitización del sur de Iberia y del Magreb mediterráneo (7,6-6,9 ky) ha sido considerada como el mismo proceso integrativo (Manen et al., 2007; Linstädter et al., 2012). Los episodios históricos (el dominio compartido de Iberia y el Magreb durante el Imperio Romano y la expansión occidental del Islam) habrían fortalecido los vínculos entre ambas costas del Mediterráneo occidental. En tiempos antiguos, los movimientos humanos a través de las regiones fueron dependientes de fases favorables en las condiciones ambientales, las cuales no fueron ni numerosas ni duraderas.

El patrón inferido para el L (especialmente para L1b) proporciona argumentos a favor de migraciones antiguas hacia Iberia en el período post-glacial. Los linajes L son originarios del África subsahariana y los FA muestran su entrada en el norte de África hacia los 10-15 ky. Dentro de África, las distancias geográficas entre los centros densamente poblados son considerables, y resultan escasas las áreas óptimas para el asentamiento humano. Muchas de las rutas que conectan las regiones más pobladas se encuentran separadas por áreas principalmente desiertas (p.e. entre Etiopía y Nigeria) y, algunas de ellas, atraviesan la gran barrera física del desierto del Sahara (p.e. entre la cuenca del Níger y Marruecos y el Delta del Nilo y Túnez) que puede ser cruzada solo bajo condiciones ambientales adecuadas. Migraciones tempranas a gran escala a través del Sahara/Sahel hacia el norte de África han sido asociadas al MIS 3 (*Marine Isotope Stage 3*, ≈50-45 ky BP) y al Período Húmedo Africano (AHP). El río más occidental que discurre a través del Sahara hacia el Mediterráneo, el Irharhar, fue una ruta probable de migración humana que conectó el oeste y noroeste de África (Coulthard et al., 2013). Otras posibles rutas de migración tempranas desde el este al noreste de África se extienden a través del África occidental sin bordear la franja mediterránea. Estos linajes L llegaron a Iberia poco después de la migración inicial subsahariana (el pico más temprano se sitúa alrededor de los 10 ky) y el FA parece apoyar, además, el papel del norte de África como parada intermedia ya que la mayoría de los linajes (~80%) que llegaron a Iberia en este período coinciden con los que llegaron al norte de África poco tiempo antes.

Por tanto, el modelo más parsimonioso para los procesos demográficos y migratorios más antiguos a través del Mediterráneo, basado en evidencias mitocondriales, sería el siguiente: a partir de los 20 ky, los linajes U6 tuvieron

una expansión extensiva en el noroeste de África, asociada con la emergencia de la industria Iberomauriense en el Magreb; este *pool* mitocondrial fue enriquecido por linajes subsaharianos L, especialmente, los L1b que comenzaron a llegar al norte de África al principio del AHP (~11-5,5 ky BP) (Castañeda et al., 2009). Los linajes U6 y L se introdujeron desde el noroeste de África hacia Iberia en el período post-glacial, más probablemente durante el *Younger Dryas* o principio del Holoceno. La apertura de las comunicaciones trans-saharianas con el AHP también permitió la migración desde el norte de África de las secuencias U6 y su expansión local en el África subsahariana, detectada por los BSPs.

Los picos recientes de migración identificados en los FA para los linajes U6 y L podrían asociarse con la dominación islámica de Iberia y el comercio de esclavos, respectivamente. El FA de las secuencias HVS-I atribuye una proporción comparativamente menor de secuencias introducidas recientemente cuando se comparan con las relacionadas con el período post-glacial: 1/3 frente a 2/3, respectivamente. Los análisis desarrollados por Cerezo et al. (2012) acerca de la filogeografía de las secuencias L en Iberia y el resto de Europa apuntan a una introducción en tiempos recientes (romanización, expansión islámica y tráfico de esclavos) de un 65%, frente a un 35% en tiempos más remotos (~11 ky). Por otro lado, nuestros FA basados en mitogenomas U6 se encuentran más próximos a estos resultados, atribuyendo una proporción 50:50 de las secuencias para ambos períodos. En el caso de las secuencias andaluzas, la fracción explicable por movimientos recientes se eleva al 70%.

Las estimas de migraciones a través del Mediterráneo y el desierto del Sahara basadas en datos mitocondriales no concuerdan generalmente con lo inferido a partir de información derivada de genotipados masivos, siendo las dataciones de estos estudios considerablemente más recientes. Henn et al. (2012) estimaron que la migración desde el África subsahariana hacia el norte de África tuvo lugar hace 40 generaciones (~1,2 ky), coincidiendo con el tráfico de esclavos trans-sahariano, mientras que su estima para el asentamiento del norte de África es previa a los 12 ky. Moorjani et al. (2011) señalaron que entre un 1-3% de la ancestría africana observada en el sur de Europa mostraba una edad media de mezcla desde hace 55 generaciones, dataciones que coinciden con un flujo norteafricano al final del Imperio Romano y las subsecuentes migraciones árabes. Otras metodologías, como las desarrolladas por Hellenthal et al. (2014), que permiten esclarecer varios eventos migratorios, conducen a edades de mezcla subsahariana en el rango de 890-1754 CE, por lo que de nuevo se apunta a la expansión islámica y tráfico de esclavos como principales acontecimientos. La actual limitación en cuanto a las estimas de edad para mezclas poblacionales empleando la variación genómica, fundamentadas en el fenómeno de la recombinación, nos revelan la continua utilidad de los marcadores uniparentales para clarificar la historia evolutiva humana, abarcando 200.000 años. En particular, nuestro trabajo apoya la existencia, frecuentemente negada, de un antiguo puente que ha conectado el Magreb con Andalucía a través del Estrecho de Gibraltar, pero localizado posiblemente más hacia el este, con una escala en la pequeña isla de Alborán.

En este estudio hemos analizado con detalle la historia de los genes africanos femeninos en el Mediterráneo basándonos en evidencias demográficas, arqueológicas y paleoclimáticas. Futuras aproximaciones genómicas nos ayudarán a profundizar en el conocimiento de los procesos de poblamiento de esta área geográfica. En este contexto, se requieren adecuados procesos de muestreo y de selección de muestras, esfuerzos en la compilación de datos genéticos con objetivos comparativos y poderosas herramientas estadísticas que permitan una evaluación precisa del marco temporal, de los episodios migratorios entre el norte de África y Europa, de su impacto en el *pool* génico europeo, del papel de las barreras físicas al flujo de población y de las rutas más probables de dispersión. Sin duda, la región que comprende el extremo occidental del Mediterráneo tendrá una importancia estratégica para abordar estas cuestiones.

Capítulo V

La historia del clado H en el Mediterráneo occidental

Este capítulo está basado en la siguiente publicación: **Hernández C.L.**, Dugoujon J.-M., Novelletto A., Rodríguez J.N., Cuesta P., Calderón R. (2015) Mitochondrial DNA clade H in southern Iberia supports ancient human genetic exchange in the western extreme of the Mediterranean (*manuscrito en preparación*).

Tradicionalmente, los estudios acerca del clado mitocondrial H en la Península Ibérica se han centrado en su franja norte –la cornisa cantábrica. Por otro lado, el norte de África alberga notables señales maternas eurasiáticas, principalmente compuestas por el linaje H. Se cree que la mayor parte de estas radiaciones llegaron al continente africano gracias a un movimiento poblacional ocurrido tras el fin del LGM y originado en el norte ibérico. Resulta notorio que el haplogrupo H no haya sido estudiado en profundidad en el área que se encuentra a medio camino entre el origen de las más relevantes sub-ramas, H1 y H3 (la región franco-cantábrica) y el límite más meridional de su expansión (el Magreb). En este Capítulo hemos llevado a cabo un estudio en alta resolución filogenética en 337 individuos autóctonos pertenecientes a dos poblaciones del sur de Iberia (Andalucía) mediante secuenciación de la región control mitocondrial y análisis de SNPs seleccionados de la región codificante. Ambas poblaciones muestran un perfil típicamente europeo occidental para la composición interna del clado H, pero los andaluces orientales (Granada) manifiestan también interesantes señales relacionadas con el Mediterráneo oriental. La estructura del linaje H a lo largo del Mediterráneo revela diferencias significativas entre los extremos occidental y oriental. Los resultados presentados aquí, junto con los referidos a otros clados mitocondriales, permiten trazar unas raíces temporales profundas en cuanto a las relaciones entre las orillas opuestas del Estrecho de Gibraltar, mostrando una prevalencia de los eventos prehistóricos frente a los históricos para explicar la proximidad genética entre Iberia y el Magreb. El peso diferencial del componente europeo materno en el norte de África (30-40%) con respecto a las trazas africanas en Europa (~15% en la población andaluza de Huelva), es un indicador de que factores demográficos han definido la asimilación de genes a través de Gibraltar. Nuestros datos presentan argumentos clave a favor de uniones prehistóricas y, por tanto, de un pasado evolutivo imbricado entre Iberia y el norte de África. Por ello, este Capítulo supone un apoyo al papel del Estrecho de Gibraltar –en muchas ocasiones presentado como una barrera geográfica infranqueable– como un puente entre continentes, culturas y civilizaciones.

Resumen

Traditionally, studies on mitochondrial DNA clade H in the Iberian Peninsula have been centered in its northern fringe –the Cantabrian cornice. By the other side, the North of Africa enclose major Eurasian traces, mainly composed by lineage H. It is assumed that most of these maternal radiations reached the African continent by means of a post-glacial population movement originated in northern Iberia. Noteworthy, haplogroup H has never been studied in depth in the halfway area between the origin of the main H sub-branches, H1 and H3, (the Franco-Cantabrian region) and the southernmost expansion limit (the Maghreb). Here, we have performed a high-resolution analysis of 337 autochthonous individuals from two populations of southern Iberia (Andalusia) by control region sequencing and coding SNPs screening. Both populations (Huelva and Granada) recorded a typical western European profile for the internal composition of clade H, but eastern Andalusians (Granada) showed also interesting population traces from the eastern

Abstract

Mediterranean. Clade H structure along the Mediterranean region revealed significant differences between the western and eastern end. The results show here, together with those referred to other mitochondrial clades, allow us to trace deep temporal roots on the relationships between both shores of the Strait of Gibraltar, being the prehistoric events more decisive than the historic ones. The dissimilar weight of the European maternal component in North Africa (30-40%), with respect to the African traces in Europe (~15% in the Iberian population from Huelva), unveiled that demographic factors have defined the assimilation of genes across Gibraltar. Our data present key arguments in favor of prehistoric bonds and an overlapping evolutionary past of Iberia and the North of Africa. Therefore, this study supports the role of the Strait of Gibraltar –many times stated as an unsurmountable geographic barrier – as a bridge between continents, cultures and civilizations.

1. Introducción

El macro-haplogrupo mitocondrial H ha sido un foco de atención para los estudios de diversidad genética humana durante más de una década. Concretamente, el interés se ha centrado en el análisis de los patrones de distribución espacial de los linajes H y en otras facetas asociadas a su historia evolutiva. Trabajos pioneros como el de Richards et al. (1998) dieron a conocer algunas propiedades que caracterizan a este clado, como es su filogenia *star-like*, compuesta por un nodo central principal (el cual alberga poblaciones de orígenes geográficos muy diversos) con multifurcaciones que surgen a partir de él. Tanto ese modelo de filogenia como la edad de coalescencia estimada para el clado H (~21 ky), han dado paso a señalar que este haplogrupo habría estado involucrado en una expansión poblacional post-glacial desde el sudoeste de Europa hacia el resto del Continente (Torroni et al., 1998).

El haplogrupo H claramente domina el *pool* mitocondrial de Europa (~40% en términos generales, con frecuencias que disminuyen progresivamente hacia el este del Continente). Sin embargo, bajo esa aparente uniforme presencia en Europa, subyace una compleja estructura de sub-linajes con especificidades locales concretas a través del continente europeo y de la cuenca mediterránea. Estos resultados comenzaron a manifestarse en un escenario científico con estudios filogenéticos más detallados que permitieron diseccionar el clado H por medio de análisis de SNPs codificantes (Loogväli et al., 2004; Pereira et al., 2005a) o secuenciación completa (Achilli et al., 2004). La clasificación de individuos en sub-ramas solamente basada en el uso de información de la región control es en muchos casos poco fiable, debido a la presencia recurrente de algunos polimorfismos y a la ausencia de sitios diagnósticos en ese fragmento mitocondrial (Loogväli et al., 2004). Por tanto, el estudio de patrones locales y específicos dentro del haplogrupo H requiere un incremento en la resolución filogenética. En esta línea, algunos trabajos han puesto de manifiesto que: (1) el

número de ramas internas del haplogrupo H es significativamente mayor que el encontrado en otros haplogrupos mitocondriales comunes en Europa (Loogväli et al., 2004), (2) la estructura del clado H en áreas orientales –tanto en el Oriente Medio/Próximo como en el Cáucaso– muestra marcadas diferencias cuando se compara con los perfiles registrados en la Europa occidental (Roostalu et al., 2007).

El área vasca y otras regiones cercanas a ella del norte de la Península Ibérica han sido, y son en la actualidad, los territorios candidatos para el estudio en sus poblaciones autóctonas, de la composición del clado H en la Europa occidental (Álvarez-Iglesias et al., 2009; García et al., 2011; Behar et al., 2012a). Esto se debe a su presumible papel como población fuente del repoblamiento post-glacial europeo indicado por algunos sub-haplogrupos H. De este modo, picos de frecuencia de los sub-clados H1 y H3 caracterizan a estas poblaciones. En clara correspondencia, los vascos albergan en su genoma un gran número de sub-ramas nativas de H1 y H3. Algunas de ellas (como H1j1, H1t1 y H3c2a) junto con la H2a5, representan ~40% de la diversidad del haplogrupo H en este territorio, encontrándose ausentes en otras poblaciones (Behar et al., 2012a).

Los linajes H también se han extendido fuera de Europa. Un ejemplo claro lo encontramos en el norte de África, donde se asume que el clado H es el principal responsable de la influencia europea a lo largo del Magreb (Cherni et al., 2009; Ottoni et al., 2010). En esta línea, la mayoría de las secuencias norteafricanas H pertenecen a los sub-haplogrupos H1 (42%) y H3 (13%) (Ennafaa et al., 2009), un patrón que sigue la misma estructura que en el vecino territorio de la Península Ibérica. Sin embargo, recientes estudios han mostrado que una gran diversidad genética se encuentra enmascarada en los perfiles H norteafricanos (Ottoni et al., 2010).

Asumiendo que parte de las secuencias H africanas fueran consecuencia de una ola expansiva post-glacial hacia el sur del Mediterráneo, desde el refugio franco-cántabro, resulta crucial un análisis detallado de este clado materno en aquellos territorios localizados a lo largo de la ruta de migración. Por ello, aquí presentamos por primera vez una visión detallada del haplogrupo H (analizado en un alto nivel de resolución filogenética) en el sur de la Península Ibérica (Andalucía). La región de Andalucía se encontraría a medio camino entre el área de origen y expansión de estas secuencias y el límite sur de ese movimiento de introducción a África. En este contexto, es fundamental conocer si el territorio de Andalucía actuó como escala para los linajes descendientes H en su ruta hacia el Magreb, o si por el contrario los genes maternos experimentaron una migración directa y no-mediada hacia el continente africano a través del Estrecho de Gibraltar o su región marítima circundante. Con este propósito, hemos seleccionado dos poblaciones autóctonas andaluzas de las provincias de Huelva y Granada. Cada una de estas poblaciones se encuentran situadas en los extremos geográficos de la región: Huelva en el extremo occidental y Granada en la vertiente oriental. Con anterioridad, ambas sub-poblaciones andaluzas han sido genéticamente caracterizadas para marcadores haploides y autosómicos por nuestro equipo (Calderón et al., 2006; Coudray et al., 2007; Ambrosio et al.,

2010a, 2010b, 2012; Fortes-Lima et al., 2014; Reales et al., 2014). El estudio de la diversidad mitocondrial global ha revelado una significativa diferenciación genética entre los andaluces occidentales y orientales (ver **Capítulo III**). La estructura poblacional observada se basa principalmente en el peso diferencial de los linajes africanos (U6, M1 y L) los cuales se encuentran más representados en el oeste que en el este. El análisis por secuenciación completa de esos linajes africanos detectados en el genoma andaluz (ver **Capítulo IV**) ha mostrado: (1) contactos transcontinentales prehistóricos entre el noroeste de África e Iberia los cuales son responsables de la mayoría de las trazas maternas africanas de Europa, (2) el relevante papel de Iberia como población fuente para migraciones de vuelta (*back-migrations*) que han portado algunas ramas específicas, desde tiempos prehistóricos, hacia el norte de África. La distribución de linajes mitocondriales concretos en poblaciones humanas contemporáneas testifican acontecimientos pasados de migración y flujo génico con direcciones definidas y de consecuencias demográficas (Cerezo et al., 2012; Pennarun et al., 2012; Achilli et al., 2013; Costa et al., 2013).

Nuestros resultados expuestos en el presente Capítulo, junto con los mostrados en el **Capítulo IV**, acerca de la presencia, complejidad filogenética y datación de los linajes mitocondriales africanos en Andalucía, deberían ser considerados como una completa visión de la singularidad del *pool* matrilineal andaluz en comparación con otras poblaciones ibéricas y mediterráneas en general.

2. Material y métodos

2.1. Poblaciones y selección de muestras

Los detalles del muestreo poblacional se exponen en el **Capítulo II**, sección 2. Como se ha comentado a lo largo de esta Tesis Doctoral, la definición de la población a investigar y un profundo conocimiento de las características específicas de la misma, unido a un adecuado muestreo geográfico y tamaño muestral, representan pilares fundamentales en estudios de diversidad genética humana. Cuando estas consideraciones se tienen en cuenta, se proporciona una especial robustez a los resultados e interpretaciones de este tipo de trabajos.

2.2. Caracterización molecular del clado H

Partiendo del total de muestras (n= 750, ver detalles en la sección **II-4**) se identificaron 337 individuos andaluces pertenecientes al haplogrupo H (110 de Huelva y 227 de Granada).

Para la evaluación de los polimorfismos T7028C (haplogrupo H) y G3010A (sub-haplogrupo H1) se diseñaron dos ensayos de genotipado TaqMan (ver metodología en la sección **II-5.3**). Las condiciones de las reacciones así como el

diseño de los *primers* y sondas se presentan en el **Anexo III, Tabla S7**. En todas las muestras se secuenció la región HVSI y parte de la HVSI (ver sección **II-5.1**). Las 224 nuevas secuencias control se depositarán en GenBank una vez que el trabajo sea aceptado para su publicación.

Asimismo, se evaluaron 21 SNPs codificantes dentro del haplogrupo H en el conjunto total de las muestras identificadas como H (n= 337). Es importante puntualizar que hemos seleccionado específicamente aquellos polimorfismos que definen las sub-ramas H, más comunes en el sudoeste de Europa. Las variantes moleculares fueron genotipadas usando la tecnología SNaPshot (ver sección **II-5.5**). Se desarrollaron 3 reacciones *multiplex*, adaptando el diseño y los protocolos propuestos por Álvarez-Iglesias et al., 2009 (ver **Tabla S8**). Además de estos SNPs, la posición A3796G (haplogrupo H1b1) fue testada por medio de PCR-RFLP (digestión con *Acil*).

El incremento en resolución filogenética alcanzado en esta otra parcela de la investigación desarrollada nos ha permitido una reducción en el porcentaje de H* (muestras H no-clasificadas). Las muestras definidas como H* incluyen secuencias basales con respecto a la rCRS (Andrews et al., 1999), y otras secuencias mitocondriales que no pueden ser adscritas a ningún sub-haplogrupo H, debido a la falta de definición. Las frecuencias de muestras H* presentadas en el **Capítulo III** [28,8% (Huelva) y 39,3% (Granada)] se redujeron respectivamente a 10,9% y 13,6% en el presente estudio.

2.3. *Análisis estadísticos y filogenéticos*

El alineamiento y chequeo de las secuencias, así como el cálculo de los parámetros estadísticos derivados de las mismas se ajustó a lo descrito en la sección **II-6.1**. Las mutaciones 16182C, 16183C, 16519 no fueron consideradas para los análisis filogenéticos.

Las diferencias en perfiles H entre pares de poblaciones –dentro de Andalucía– se evaluaron mediante el índice F_{ST} , teniendo en cuenta las secuencias mitocondriales y la estructura de sub-haplogrupos. Del mismo modo, estas diferencias dentro del sur de Iberia se analizaron mediante un test χ^2 (ver sección **II-6.5**).

Para llevar a cabo los análisis comparativos, construimos una base de datos actualizada que incluye 93 poblaciones (con un tamaño muestral ≥ 50) que cubren desde el Mediterráneo occidental (con un especial interés tanto en Iberia como en el norte de África) hasta el Oriente Próximo (ver **Tabla S9**). Esta fuente de información fue cuidadosamente seleccionada desde trabajos previamente publicados. La compilación de poblaciones siguió una serie de criterios. En primer lugar, la adscripción de las muestras al haplogrupo H es inviable mediante el uso exclusivo de información procedente de la región control. Por ello, solo tuvimos en cuenta aquellas secuencias mitocondriales que hubieran sido genotipadas también para la mutación T7028C. Del mismo modo, la definición de los sub-haplogrupos H requiere habitualmente el análisis de algún polimorfismo

de la región codificante (p.e. G3010A para H1, T6776C para H3). Hemos seleccionado información de la región control (HVS-I, rango 16051-16400) y de la región codificante del genoma mitocondrial en un nivel comparable al evaluado por nosotros. La base de datos construida (n= 18.622 individuos, de los cuales 6.011 pertenecen al clado H) fue usada en dos niveles de resolución. Primero, evaluamos la distribución del haplogrupo H, dentro de una escala amplia, usando la base de datos completa. Después, solo seleccionamos aquellos estudios (n= 71, aquellos subrayados en la **Tabla S9**) que diseccionan el clado H con cierto detalle filogenético, es decir, mediante el genotipado de variantes codificantes, para los subsiguientes análisis de la estructura interna del clado. Cuando más de un estudio en alta resolución de la misma población ha sido publicado, las muestras se agruparon en una única población. Con respecto al norte de África, hemos combinado las poblaciones árabes y bereberes siguiendo lo propuesto en anteriores trabajos según los cuales las diferencias en la estructura del haplogrupo H en poblaciones magrebíes se basan en factores geográficos y no culturales (Ennafaa et al., 2009).

La estructura interna del haplogrupo H en Europa y la región mediterránea se evaluó mediante un método estadístico de naturaleza multivariada, el HCA y un análisis molecular de la varianza, AMOVA (ver secciones **II-6.5** y **II-6.6**).

Para aquellos sub-linajes que alcanzaban una representación moderada (n> 35), se estimó su edad de coalescencia (TMRCAs) empleando el estadístico ρ (*rho*) (detalles en la sección **II-6.2**). Además, construimos *networks* específicas para los sub-clados H1 y H3 (sección **II-6.1**) con el fin de desvelar la naturaleza de los presuntos contactos poblacionales entre Iberia y el norte de África.

Por último, la distribución geográfica del haplogrupo H y de alguno de sus sub-clados principales fue representada en mapas de interpolación de superficies (*surface maps*). La localización de las poblaciones utilizadas se presenta en la **Figura S14**. La existencia de gradientes o de variaciones clinales para el clado H y ciertos sub-linajes fue analizada mediante análisis de autocorrelación espacial. La sección **II-6.7** expone los detalles acerca de la metodología y programas utilizados.

3. Resultados

3.1. *Disección del haplogrupo mitocondrial H en Andalucía*

Las frecuencias de los sub-haplogrupos H observadas en Andalucía se muestran en la **Tabla V-1**. La incidencia global del clado H entre los andaluces (45%; 337/750) es similar a la encontrada en otras poblaciones ibéricas (~42% valor promedio). Esas distribuciones de frecuencias dentro de Andalucía muestran diferencias significativas (χ^2 -test, $P < 0,05$), revelando por tanto señales de estructuración genética. Este escenario es similar al observado cuando se analiza la presencia e incidencia de los linajes maternos africanos (U6, M1 y L) siendo

considerablemente mayor entre los andaluces occidentales (~12%) en relación con lo detectado en sus vecinos del este de la región (3,62%) (ver **Capítulos III y IV**). Los datos derivados del análisis del cromosoma Y no son tan discriminantes en este sentido (Ambrosio et al., 2010b).

Haplogrupo	Andalucía W (Huelva, n _T =280)		Andalucía E (Granada, n _T =470)	
	n	%	n	%
H	110	39,29	227	48,30
H*	12	10,91	31	13,66
H1	48	43,64	73	32,16
H1*	38	34,55	58	25,55
H1a	-	-	1	0,44
H1a1	1	0,91	-	-
H1b1	-	-	5	2,20
H1c3	4	3,64	-	-
H1e1a3	-	-	5	2,20
H1k	1	0,91	-	-
H1t1a1	-	-	1	0,44
H1ah1	3	2,73	-	-
H1ba	-	-	1	0,44
H1bf1	1	0,91	2	0,88
H2	3	2,73	3	1,32
H2a1	3	2,73	1	0,44
H2a2b	-	-	2	0,88
H3	16	14,55	41	18,06
H3*	14	12,73	30	13,22
H3c	-	-	10	4,41
H3c2	2	1,82	1	0,44
H4	3	2,73	12	5,29
H4a1	1	0,91	12	5,29
H4a1a4b1	2	1,82	-	-
H5	5	4,55	17	7,49
H5*	4	3,64	6	2,64
H5a	-	-	5	2,20
H5a3a1	-	-	5	2,20
H5a4a	1	0,91	1	0,44

Haplogrupo	Andalucía W (Huelva, n _T =280)		Andalucía E (Granada, n _T =470)	
	n	%	n	%
H6	8	7,27	16	7,05
H6a	6	5,45	9	3,96
H6a1a1a	-	-	4	1,76
H6a1a7	2	1,82	2	0,88
H6a1b4	-	-	1	0,44
H7	5	4,55	7	3,08
H7*	5	4,55	6	2,64
H7h	-	-	1	0,44
H10	3	2,73	-	-
H10a1	1	0,91	-	-
H10e	2	1,82	-	-
H11	-	-	5	2,20
H11*	-	-	4	1,76
H11a	-	-	1	0,44
H13	-	-	9	3,96
H17	1	0,91	5	2,20
H17*	1	0,91	3	1,32
H17c	-	-	2	0,88
H18	5	4,55	1	0,44
H20	-	-	4	1,76
H20*	-	-	1	0,44
H20a	-	-	3	1,32
H27	1	0,91	1	0,44
H36	-	-	1	0,44
H82	-	-	1	0,44

Tabla V-1. Frecuencias de los sub-linajes H en Andalucía. La primera fila (en cursiva y sombreado) indica la frecuencia global del clado H en cada población estudiada. A partir de ahí, las frecuencias se refieren exclusivamente al haplogrupo H. Los asteriscos (*) indican las muestras no clasificadas a este nivel de resolución o que pertenecen al paragrupo H*.

De las 337 secuencias mitocondriales, el número de ramas internas identificadas dentro del clado H fueron 41, de las cuales, 24 muestran frecuencias

$<0,01$. Las estimas de diversidad del haplogrupo (\pm SD) variaron entre $0,847\pm 0,027$ y $0,888\pm 0,013$ en Andalucía occidental y oriental, respectivamente. En una escala geográfica más amplia (Península Ibérica), el valor menor de diversidad dentro del haplogrupo H se registra en vascos ($0,646\pm 0,041$) (Behar et al., 2012a). Las relaciones filogenéticas entre las secuencias mitocondriales andaluzas se muestra en la **Figura S15**.

Los sub-clados H1 y H3 engloban un 52,81% de la diversidad interna de H en Andalucía [43,64% (Huelva) y 32,16% (Granada) para H1, y 14,55% (Huelva) y 18,06% (Granada) para H3]. El rango de estos resultados mantiene la proporción de lo expuesto en el **Capítulo III** y son correspondientes a lo encontrado en otras poblaciones ibéricas/europeas occidentales. El sub-haplogrupo H1 es particularmente frecuente en la región franco-cantábrica (~60% de toda la variación de H en población vasca de España, Behar et al. 2012a). Esta proporción se reduce lentamente hacia el sur de la Península. En las poblaciones vecinas norteafricanas, las frecuencias de H1 son generalmente comparables a las detectadas en Iberia: 52% en Marruecos y 37% en Libia, con valores intermedios en Argelia (49%) y Túnez (37%) (Cherni et al., 2009; Ennafaa et al., 2009; Fadhlouzi-Zid et al., 2011; Bekada et al., 2013).

Del mismo modo que para el H1, el sub-haplogrupo H3 también alcanza su pico de frecuencia en la Península Ibérica (valor promedio 15%; en vascos la proporción es del 22,4% dentro del haplogrupo H). Curiosamente, su presencia es escasa o ausente en el sur de Italia (Ottoni et al., 2009) así como entre poblaciones del Oriente Próximo y Medio y población del Cáucaso (Roostalu et al., 2007). En población magrebí, el H3 varía desde un 8% en Marruecos hasta el 19% en Túnez.

La **Figura V-1** presenta los mapas de contornos basados en las frecuencias del haplogrupo H y de sus sub-clados H1 y H3 en Europa, cuenca mediterránea y áreas vecinas. Los mapas ponen en evidencia una estructura espacial para los tres clados considerados, con frecuencias que disminuyen desde el oeste al este europeo, incluyendo la región mediterránea (los correlogramas de Moran tienen una significación corregida de Bonferroni de 0,0000). La especificidad del patrón de correlación es especialmente visible para el H1 cuando únicamente se incluyen a las poblaciones europeas. La forma (perfil) del correlograma exhibe un gradiente lineal (ver **Figura V-2**) sin puntos de inflexión producidos por un supuesto movimiento migratorio direccional. Este tipo de patrón está caracterizado por valores de autocorrelación positivos (y significativos) para distancias cortas y valores significativos negativos entre localidades, que de acuerdo con la distancia entre las mismas se incrementa (ver más detalles en Legendre & Fortin, 1989).

El sub-haplogrupo H1 se caracteriza por tener la complejidad interna más elevada. Nuestros datos de Andalucía indican coherentemente que: (1) H1 es relativamente menos diverso en el oeste [$0,865\pm 0,442$ (Huelva)] que en el este de la región [$0,940\pm 0,0132$ (Granada)], (2) la mayoría de los sub-clados H1 no aparecen simultáneamente en las dos sub-poblaciones andaluzas (p.e. H1a1, H1c3, H1k y H1ah1 solo fueron detectados en la muestra de Huelva, mientras que

H1b1, H1e1a3, H1t1a1 y H1ba se observan en Granada) y, (3) los linajes H1 se encuentran generalmente en Andalucía en frecuencias <5%.

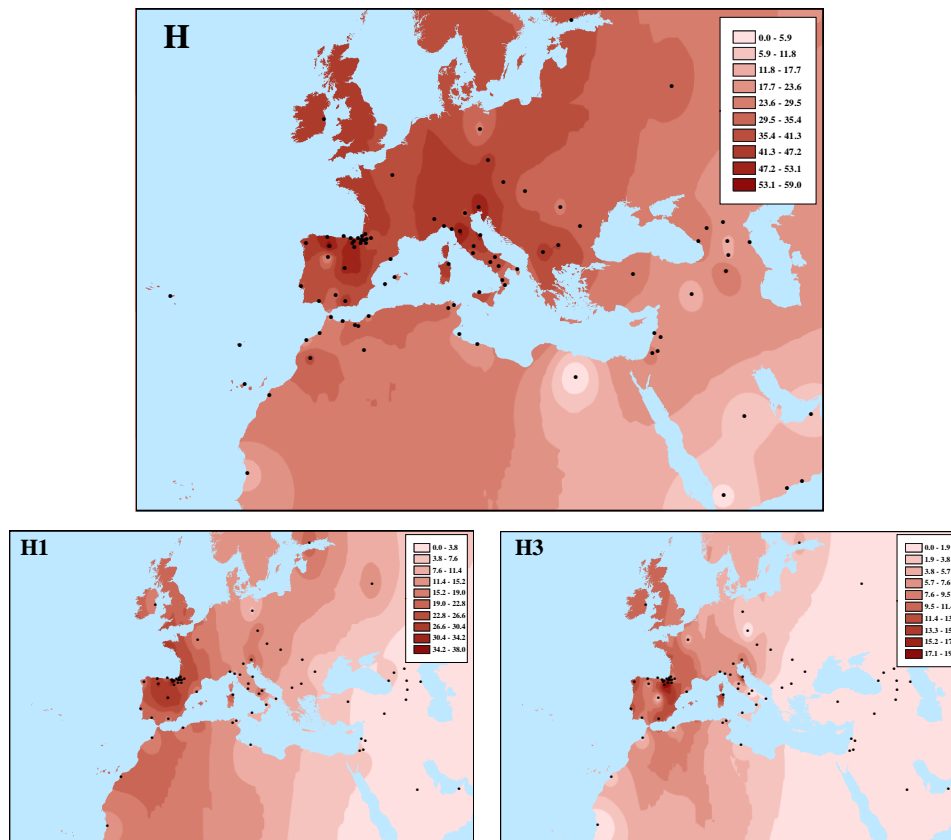
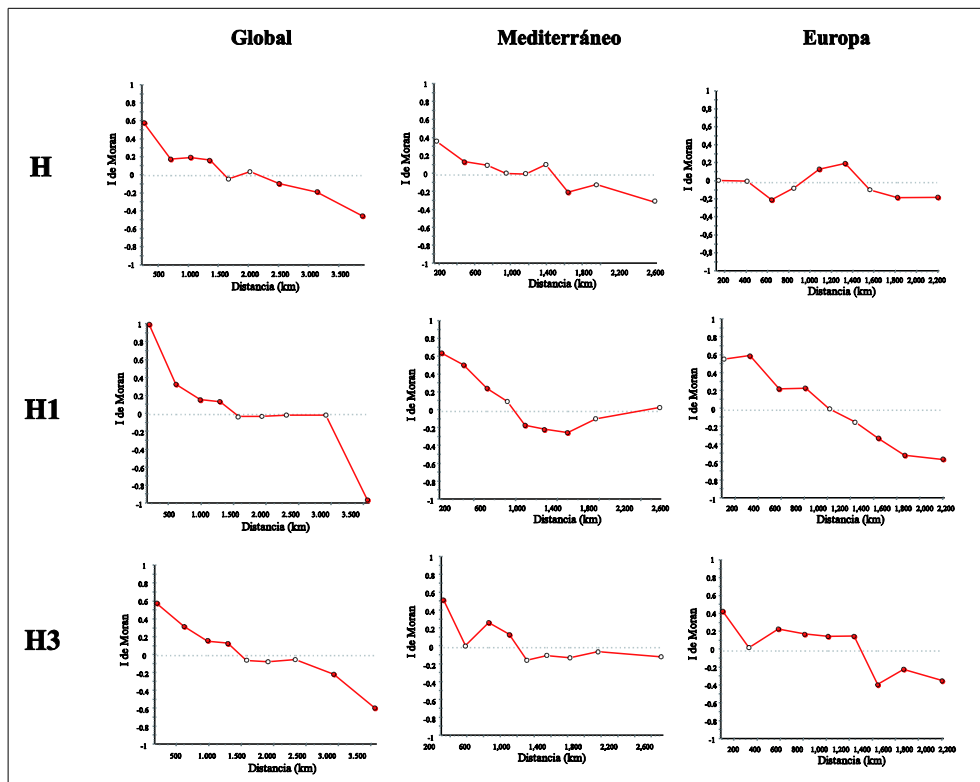


Figura V-1. Distribución geográfica del clado H y de sus sub-haplogrupos H1 y H3. Las frecuencias (%) se muestran en la escala de color. Ver códigos y referencias en la **Tabla S9**.

Otros estudios previos sobre la variación de la estructura del haplogrupo H en la Península Ibérica (ver Álvarez-Iglesias et al., 2009, entre otros) apuntan a H1c3, H1k y H1ba como linajes mitocondriales restringidos a las poblaciones ibéricas atlánticas (p.e. portugueses y gallegos). Los nuevos datos presentados aquí enriquecen y amplían este rango geográfico hacia el extremo sur de la Península. Los análisis moleculares detectaron también la presencia del linaje H3c en andaluces de Granada (4,41%) y su rama derivada H3c2 en otro individuo autóctono (0,44%). A este respecto, Behar et al. (2012a) indicaron que la subrama H3c2a parece estar especialmente restringida a la población vasco-parlante y otras poblaciones geográficamente vecinas (3,30% de la variación del haplogrupo H). Cabe destacar que tanto el H1k como el H3c han sido observados en el sur de Italia (Achilli et al., 2004; Ottoni et al., 2009; Brisighelli et al., 2012) y en la Europa central y oriental (Brandstätter et al., 2008; Karachanak et al., 2012). La rara variante H1bf1, detectada en frecuencias <1% en nuestra muestra

andaluza, se considera específica de Iberia (Pereira et al., 2006; Álvarez-Iglesias et al., 2009; García et al., 2011).

Figura V-2. Correlogramas de los análisis de autocorrelación espacial para el clado H y los sub-clados H1 y H3 en el conjunto de poblaciones general, en el Mediterráneo y en Europa. Las clases de distancia significativas se indican con círculos rojos ($P < 0,05$).



Los haplogrupos H5 y H6 son los siguientes contribuyentes dentro del clado H en Europa. En el sur del Continente, las frecuencias medias de H5 y H6 varían entre 4-6% y 3-10% dentro del linaje H, respectivamente. De acuerdo con estos resultados, nuestros datos de Andalucía revelan que ambos haplogrupos explican el 13,64% de la diversidad de H, con una mayor representación de H5 en Granada (7,50%) y de H6 en el caso de Huelva (7,27%). En la Europa central y oriental (poblaciones eslavas), el H5 oscila entre 2,5-3,2%, y uno de sus sub-clados más frecuentes, el H5a, se piensa que pudiera tener un origen centroeuropeo (Mielnik-Sikorska et al., 2013). En poblaciones que habitan el alto Cáucaso (Roostalu et al., 2007), el H5* es el sub-clado más común dentro de H (20%). Por otro lado, tanto el H6a como el H6b muestran patrones de distribución distintivos, con una relativa alta frecuencia de H6a en Europa, y de H6b en población de Oriente Próximo y centro de Asia (ver Loogväli et al., 2004; Malyarchuk et al., 2006; Pereira et al., 2006). El sub-clado H6a representa un componente relevante en la variación estructural del clado H en el sur de Iberia. El sub-clado H6b no fue detectado en las muestras de Andalucía.

Otras sub-ramas del haplogrupo H, como H11, H20, H36 y H82 solo fueron observadas entre los andaluces de Granada, no detectándose entre sus vecinos occidentales (Huelva). El linaje H11 es un caso especial, ya que la frecuencia observada (2,20%) paraleliza a la registrada en el noroeste de África (Marruecos y Argelia <1,75%) (Rando et al., 1998; Ennafaa et al., 2009; Bekada et al., 2013). Sin embargo, el H20 (1,76%, presente estudio) es considerado un haplogrupo más raro en Iberia, con muy pocos representantes en su extremo noroccidental (Galicia) así como en el territorio vecino de León, en la meseta central (Álvarez-Iglesias et al., 2009; Ennafaa et al., 2009). Finalmente, el H36 (16070) y el H82 (16220), descritos aquí solo con información de la región control, registran frecuencias no polimórficas (<1%) tal y como sucede en otras poblaciones ibéricas, europeas y del Oriente Próximo y Medio (Pereira et al., 2005a; Roostalu et al., 2007; Ottoni et al., 2009; Brisighelli et al., 2012; Badro et al., 2013, entre otros).

El haplogrupo H13 se distribuye por el arco norte de la Península, extendiéndose desde Galicia (1,6%) a Cataluña (5%). En el extremo meridional (Andalucía), el H13 se encuentra presente en una frecuencia del 4% en Granada. Una frecuencia superior (10%) se registra en la Italia peninsular y en Cerdeña, mientras que en el norte de África la proporción es mucho menor (2-5%, Cherni et al., 2009; Ennafaa et al., 2009).

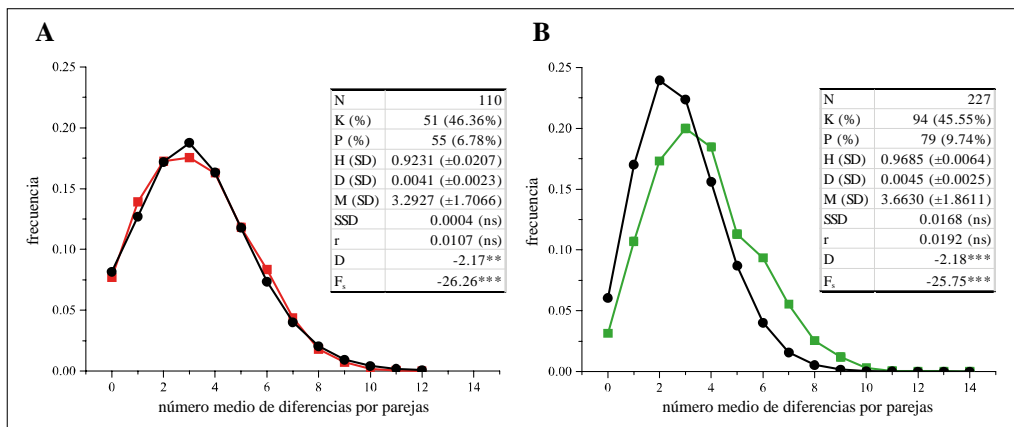
Como material suplementario, se presentan mapas de superficie mostrando las distribuciones de frecuencias de algunos sub-haplogrupos H específicos (**Figuras S16-S19**). La localización geográfica de las poblaciones consideradas se indica en la **Figura S14**.

El perfil del haplogrupo H observado en Andalucía se encuentra estructurado genéticamente ($F_{ST} = 0,007$; valor- $P = 0,018$) aunque en los mapas genéticos (ver más abajo), ambas poblaciones autóctonas se posicionan en el mismo *cluster*. El test χ^2 también apoya la diferenciación genética interpoblacional ($\chi^2 = 30,91$; g.l. = 16; $P = 0,014$). El análisis de los residuos corregidos tipificados identifica a los sub-haplogrupos H18, H10, H1 y H13 (en orden descendente de significación) como los mayores contribuyentes (95% I.C.) asociados a las diferencias observadas. Este interesante resultado se aproxima a lo encontrado usando secuencias mitocondriales H ($F_{ST} = 0,007$; $P = 0,027$). Otro dato relevante emerge del análisis de las secuencias mitocondriales y de los polimorfismos de la región codificante. Encontramos 158 haplotipos diferentes entre los 337 individuos H (46,9%), y solo un pequeño porcentaje (11%; 17/158) de los mismos es compartido entre ambas poblaciones andaluzas. La **Tabla S10** muestra la caracterización molecular de las muestras autóctonas H analizadas y la asignación a sub-clados específicos.

En la **Figura V-3** se muestra la información estadística relativa a las secuencias control andaluzas pertenecientes al haplogrupo H. Escenarios de expansiones poblacionales recientes para el linaje H en este territorio se ven soportados por los valores significativos y negativos de los índices de neutralidad de Tajima (D) y Fu (F_s). La suma de desviaciones al cuadrado entre las distribuciones *mismatch* observadas y esperadas (no significativa) y el valor del

índice de Harpending (r) apoyarían este resultado. En términos comparativos, las diversidades de haplotipos estimadas en el sur de Iberia ($0,9231 \pm 0,021$ - $0,9685 \pm 0,006$, presente estudio) son mayores que las encontradas en su extremo septentrional (Cornisa Cantábrica) [p.e. Galicia: $0,800 \pm 0,038$; Cantabria: $0,875 \pm 0,042$, Álvarez-Iglesias et al., 2009]. En el norte de África, los valores oscilan entre $0,860 \pm 0,060$ en Marruecos y $0,970 \pm 0,010$ en Túnez (Ennafaa et al., 2009).

Figura V-3. Distribuciones *mismatch* para Huelva (A) y Granada (A) y parámetros estadísticos de las secuencias del ADNmt. Para ambas poblaciones, las distribuciones esperadas se muestran con una línea negra.



K: número de secuencias diferentes; P: sitios polimórficos; H: diversidad haplotípica; D: diversidad nucleotídica; M: número medio de diferencias por parejas; SSD: suma de desviaciones al cuadrado entre distribución esperada y observada; r: índice de Harpending; D y F_s de neutralidad de Tajima y Fu (ns: no-significativo, **p < 0,01; ***p < 0,001).

La **Figura V-4** presenta las redes de haplotipos para los linajes H1 y H3 basadas en un conjunto de poblaciones ibéricas (n= 12) y norteafricanas (n= 5) que han sido analizadas en un alto nivel de resolución filogenética para el haplogrupo H (ver poblaciones subrayadas en la **Tabla S9**). Las conexiones maternas entre Iberia y el Magreb recaen en una presencia paralela, aunque con diferencias en el peso real comprendido por cada sub-haplogrupo (siendo la contribución de Iberia la principal). Ambas *networks* muestran filogenias a modo de estrella (*star-like*), con un nodo central, representado por todas las poblaciones usadas en el *dataset* para el H1 (46,5% de las secuencias totales H1) y, casi todas, para el H3 (56,8% de las secuencias H3). Ambas genealogías se caracterizan por una baja proporción de haplotipos compartidos entre Iberia y África (~11%). La mayor parte de los nodos terminales norteafricanos surge directamente de los nodos basales de H1 y H3, lo que apunta a varias radiaciones independientes y casi exclusivas de África.

Cabe destacar que la ocurrencia de haplotipos idénticos entre individuos ibéricos es más frecuente (28,9% de las secuencias H1; 25,0% de H3). Algunas identidades han sido detectadas entre las secuencias andaluzas y norteafricanas, pero las genealogías no revelan que estas conexiones sean preferenciales.

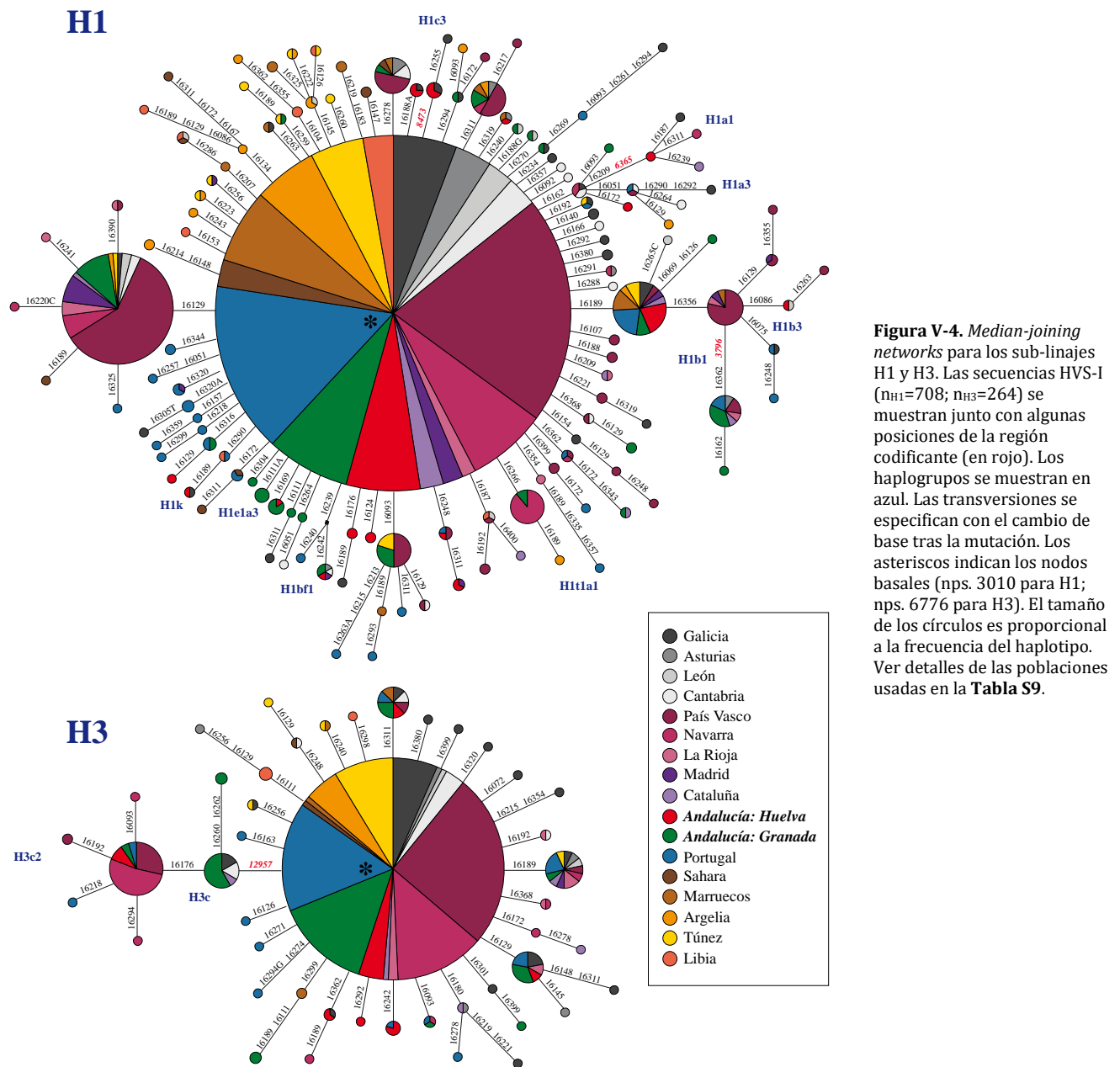


Figura V-4. Median-joining networks para los sub-linajes H1 y H3. Las secuencias HVS-I ($n_{H1}=708$; $n_{H3}=264$) se muestran junto con algunas posiciones de la región codificante (en rojo). Los haplogrupos se muestran en azul. Las transversiones se especifican con el cambio de base tras la mutación. Los asteriscos indican los nodos basales (nps. 3010 para H1; nps. 6776 para H3). El tamaño de los círculos es proporcional a la frecuencia del haplotipo. Ver detalles de las poblaciones usadas en la **Tabla S9**.

3.2. Estructura genética-poblacional en el Mediterráneo

El Análisis Jerárquico de Cluster (HCA) mostrado en la **Figura V-5** analiza la variabilidad en la composición interna del haplogrupo H en 71 poblaciones seleccionadas de Europa, la cuenca mediterránea y Oriente Medio y Próximo (ver datos en **Tabla S9**). Es interesante observar cómo la dispersión de las poblaciones en el espacio bidimensional se ve dominada por el eje horizontal (el Factor 1 explica el 71,89% de la varianza total), mientras que el Factor 2 solo engloba el 10,77% de la variabilidad observada. El HCA conforma cuatro *clusters*,

siendo los haplogrupos H*, H1, H3, H13 y H5 (ver vectores en el gráfico) los que significativamente definen la topología genética de las poblaciones. Cuando se computa la descomposición de la inercia de los dos primeros ejes, el coeficiente de inercia fue 0,812 tras la consolidación. Este valor es coherente con el número de ramificaciones principales observadas en el árbol, y constituye una prueba de que una elevada proporción de los datos se ve explicada por estos cuatro agrupamientos o clusters.

El *cluster* 1 (C1) es el más numeroso (n= 24 poblaciones), con una notable representación de muestras ibéricas (n= 14) junto a otras europeas (no-ibéricas) y norteafricanas de Marruecos, Argelia y Libia. El C1 se encuentra caracterizado por H1 (media= 0,413; $P= 0,006$) y H3 (media= 0,152; $P= 0,000$). Las dos poblaciones andaluzas estudiadas (destacadas en cursiva) son parte de este agrupamiento. El C2 (n= 10) puede ser descrito como el *cluster* 'vasco' (n= 8), con el haplogrupo H1 mostrando una muy alta significación estadística (media= 0,642; $P= 0,000$) en la definición genética del *pool* mitocondrial vasco. Por el contrario, el haplogrupo H5 domina mayoritariamente el C3 (media= 0,095; $P= 0,016$), el cual comprende un alto número de poblaciones (n= 20) con un componente principal italiano (n= 9), y otras poblaciones de Oriente Próximo/Medio y de la región del Cáucaso. Cabe destacar que dos poblaciones norteafricanas de Túnez (#88, #89 de la **Tabla S9**) se localizan también en C3. Los tunecinos se ven más fuertemente influenciados por el Oriente Próximo, en términos de su composición mitocondrial del linaje H (Ennafaa et al., 2009) en comparación con sus vecinos del África noroccidental. Finalmente, el C4 (n= 17) está caracterizado por H* (media= 0,583; $P= 0,000$) y, más débilmente, por los haplogrupos H13 y H20. Este *cluster* se encuentra compuesto por un mosaico de poblaciones, principalmente con un origen en Oriente Medio, el Cáucaso y el sudoeste de Asia (n=12) y con otras muestras (n= 5) de Italia y Europa central. Los haplogrupos H10, H3 y H1 están particularmente infrarrepresentados en el C4.

Los resultados del Análisis Molecular de la Varianza (AMOVA) (ver **Tabla V-2**), basados en los mismos datos que el HCA, proporcionan más argumentos a favor de una estructura genética. En primer lugar se observa una diferenciación significativa en la composición del haplogrupo H entre las poblaciones de los extremos occidental y oriental del espacio mediterráneo ($F_{CT}= 0,1513$, $P < 0,001$). En contraposición, se detecta también una estructura poblacional aunque débil entre Iberia y el norte de África (Magreb), indicada por el valor $F_{CT}= 0,0106$, $P= 0,0352$. Este resultado concuerda con los datos generalmente encontrados sobre la estructura del linaje H en el norte de África (Ennafaa et al., 2009). Cuando el test solo se aplica a Iberia, no se encontraron diferencias significativas entre las poblaciones del norte y del sur de la Península ($F_{CT}= 0,0102$). Un escenario similar fue observado cuando se compara el área vasca (como representante puntual del norte de Iberia) con Andalucía, en el extremo geográfico opuesto ($F_{CT}= 0,0212$).

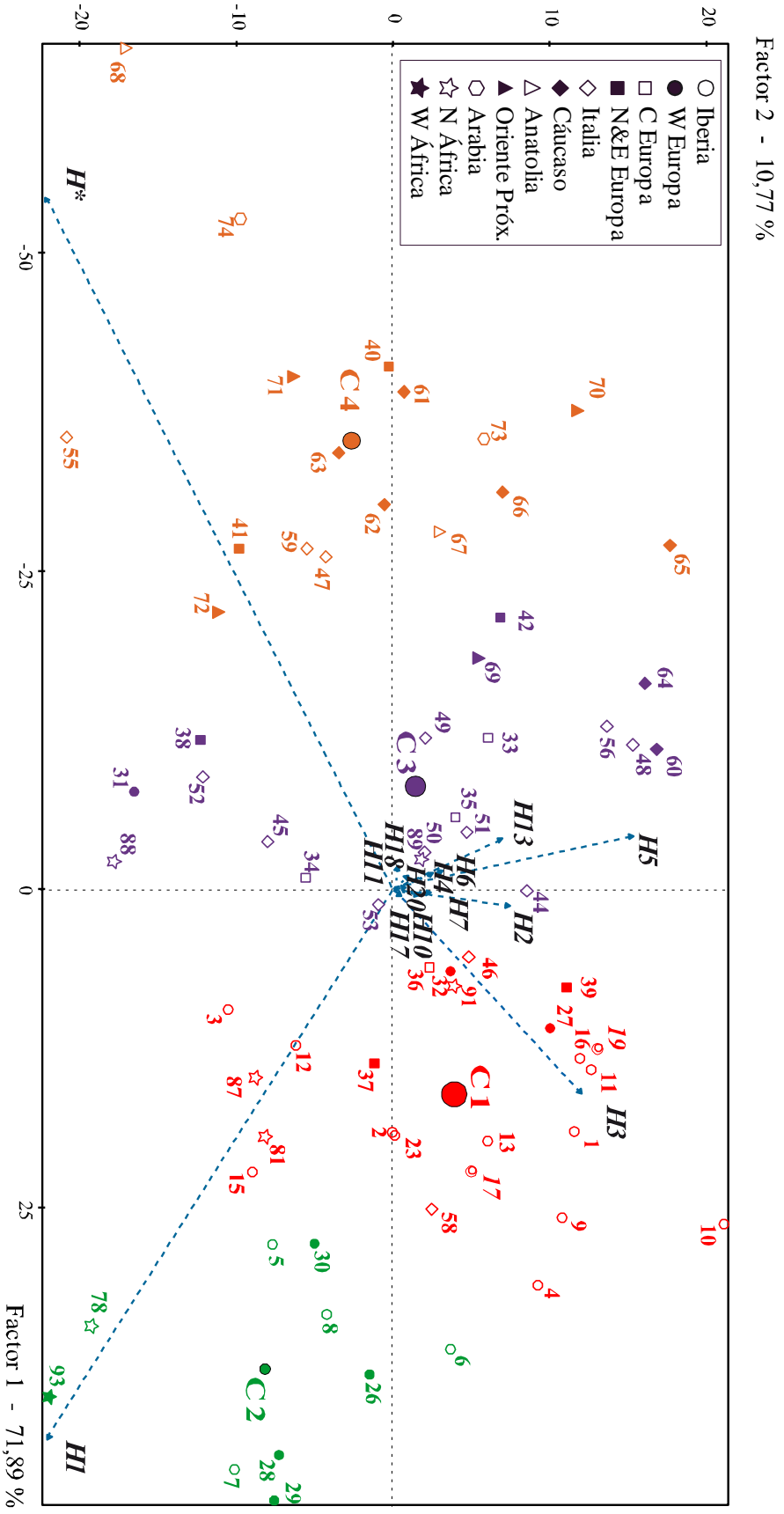


Figura V-5. Análisis Jerárquico de Cluster (HCA) de 71 poblaciones en las cuales la estructura del clado H fue estudiada. Solo se han tenido en cuenta estudios de alta resolución filogenética (ver poblaciones subrayadas en la **Tabla S9**). Los sub-clados se muestran como vectores. El código de color indica la pertenencia a los 4 *clusters* definidos por el método multivariante. Cada grupo geográfico tiene un símbolo característico (ver leyenda). Las poblaciones se muestran como códigos (ver **Tabla S9**). Huelva corresponde a la población I7 y Granada a la 19.

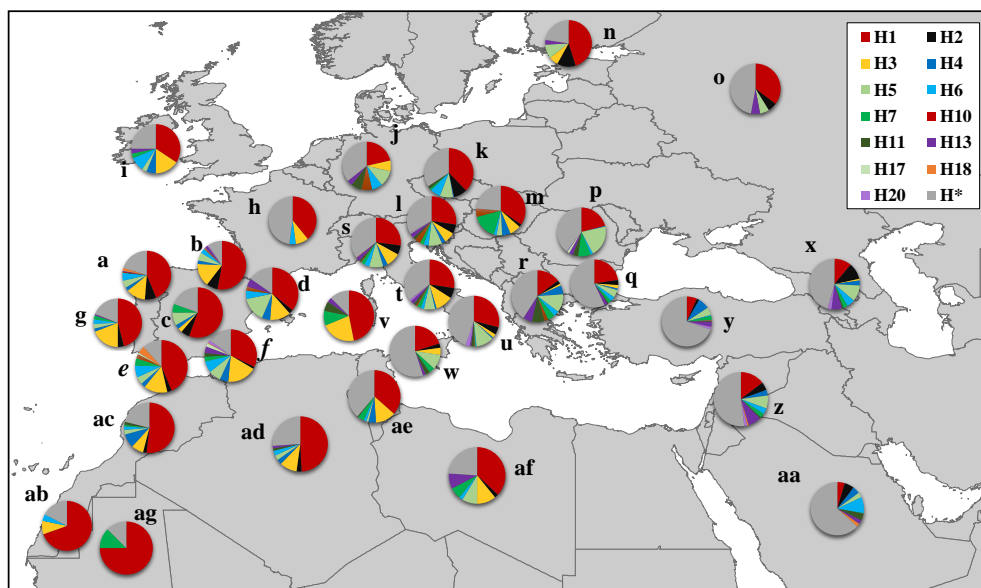
Tabla V-2. Análisis molecular de la varianza (AMOVA) de acuerdo con diferentes escenarios geográficos. Las poblaciones se tomaron de la **Tabla S9**.

<i>Mediterráneo oriental</i>			
<i>Mediterráneo occidental</i>	$F_{CT} = 0,15130^{***}$	15,13	
	$F_{SC} = 0,02568^{***}$	2,18	
	$F_{ST} = 0,17310^{***}$	82,69	
<i>Italia</i>		<i>Magreb</i>	
<i>Iberia</i>	$F_{CT} = 0,05207^{***}$	5,21	$F_{CT} = 0,01059^{*a}$ 1,06
	$F_{SC} = 0,02284^{***}$	2,17	$F_{SC} = 0,02096^{***}$ 2,07
	$F_{ST} = 0,07373^{***}$	92,63	$F_{ST} = 0,03133^{***}$ 96,87
a valor- $P = 0,03519 \pm 0,00598$			
<i>Norte de Iberia</i>		<i>Área Vasca</i>	
<i>Andalucía</i>	$F_{CT} = 0,01022$ (n.s.)	1,02	$F_{CT} = 0,02117$ (n.s.) 2,12
	$F_{SC} = 0,01921^{***}$	1,9	$F_{SC} = 0,01944^{***}$ 1,9
	$F_{ST} = 0,02923^{***}$	97,08	$F_{ST} = 0,04020^{***}$ 95,98

n.s.: no-significativo; * $P < 0,05$; *** $P < 0,001$

Un resumen visual de la estructura geográfica del linaje H se presenta en la **Figura V-6**. Como ha sido ya enfatizado, H1 y H3 explican conjuntamente más de la mitad de la diversidad interna al clado H en la Europa occidental (Iberia), un patrón que puede ser extrapolado al norte de África. Sin embargo, a medida que nos movemos hacia el este del Mediterráneo, otras sub-ramas de H adquieren más relevancia en el panorama genético.

Figura V-6. Representación gráfica de las frecuencias de los sub-clados H detectados en Europa y el Mediterráneo.



Poblaciones tomadas de la **Tabla S9**. a: noroeste España (pob. 1-3); b: área franco-cantábrica (pob. 4-13, 26-30); c: Madrid (15); d: Cataluña (16); e: Huelva (17); f: Granada (19); g: Portugal (23); h: Francia (31); i: Irlanda (32); j: Alemania (33); k: República Checa (34); l: Austria (35); m: Hungría (36); n: Finlandia (37); o: Rusia (38); p: Rumanía (39,40); q: Bulgaria (41); r: Macedonia (42); s: Italia norte (44-48); t: Italia centro (49-51); u: Italia sur (52,53,55,56); v: Cerdeña (58); w: Sicilia (59); x: Cáucaso (60-66); y: Anatolia (67,68); z: Oriente Próximo (69-72); aa: Arabia (73,74); ab: Sahara (78); ac: Marruecos (81); ad: Argelia (87); ae: Túnez (88,89); af: Libia (91); ag: Mauritania (93).

3.3. Estimación de edades de coalescencia

Las edades de coalescencia para algunos sub-haplogrupos H se muestran en la **Tabla V-3**. Con la excepción de H2, H11 y H20, los cuales presentan las edades estimadas más antiguas [17.789 (Hg H20) - 23.648 (Hg H2) años BP], los restantes sub-clados presentan edades que coinciden con el período post-glacial final o con el Holoceno temprano. Cabe destacar que la profundidad temporal de H1 y H3 es más reciente en el norte de África que en Iberia, y, por tanto, su origen más probable se encuentre en esta última área geográfica. El linaje H1 posee una edad de 13,4 ky en Iberia y de 10,5 ky en el norte de África. Los valores correspondientes para H3 son 9,5 y 8,1 ky, respectivamente. Estos resultados podrían ser explicados por una introgresión prehistórica desde Iberia hacia el norte de África, probablemente, como consecuencia del mismo movimiento que provocó el repoblamiento de Europa (Cherni et al., 2009). Estos episodios de flujo génico prehistóricos se ajustarían a escenarios propuestos de estructura poblacional en el Mediterráneo occidental (Currat et al., 2010). La baja frecuencia de haplotipos H compartidos entre las poblaciones del Mediterráneo occidental revelarían una radiación rápida hacia el sur.

Sub-clado	n	ρ (HVS-I)	σ	Edad (años)	95% I.C.
H1	1524	0,99	0,21	16.480	[9.687-23.273]
H1 Iberia	591	0,80	0,20	13.375	[6.892-19.859]
H1 Norte de África	191	0,63	0,11	10.478	[6.756-14.200]
H2	244	1,42	0,70	23.648	[899-46.397]
H3	391	0,78	0,20	13.085	[6.442-19.728]
H3 Iberia	236	0,57	0,15	9.469	[4.420-14.518]
H3 Norte de África	37	0,49	0,17	8.113	[2.526-13.700]
H4	129	0,57	0,13	9.437	[5.243-13.632]
H5	324	0,61	0,14	10.140	[5.686-14.594]
H6	177	0,78	0,19	13.002	[6.931-19.074]
H7	120	0,66	0,19	10.979	[4.756-17.202]
H10	69	0,88	0,41	14.743	[1.286-28.201]
H11	59	1,27	0,41	21.200	[7.823-34.577]
H13	120	1,00	0,20	16.677	[10.154-23.200]
H20	46	1,07	0,50	17.789	[1.353-34.225]

Tabla V-3. Edades de coalescencia estimadas para los principales sub-clados de H. Los valores del parámetro ρ se han basado en las secuencias HVS-I detalladas en la **Tabla S9** y fueron transformados a años usando el reloj molecular corregido propuesto por Soares et al., 2009.

4. Discusión

Para refinar el conocimiento previo en el campo de la Genética Evolutiva Humana, son necesarios estudios en alta resolución genealógica. Aproximaciones detalladas filogenéticas han permitido la disección del aparentemente homogéneo clado mitocondrial europeo H en varias sub-ramas que reflejan de

manera más ajustada las dispersiones del hombre moderno a lo largo del continente europeo y de la cuenca mediterránea.

La evidencia mitocondrial mostrada en el presente Capítulo, basada en el análisis interno del clado H, revela que las mayores diferencias entre las poblaciones europeas en un contexto europeo se detectan en un eje longitudinal (oeste-este). Sin embargo, la estructuración genética de Europa, que se deriva de estudios *genome-wide* se encuentra basada fundamentalmente en la latitud (Seldin et al., 2006; Lao et al., 2008) aunque otras investigaciones, también, han mostrado una estratificación según un eje este-oeste (Bauchet et al., 2007). Tras la colonización inicial del continente por el hombre anatómicamente moderno (AMH) los dos períodos que fundamentalmente condujeron a la definición y modelación de la estructura genética europea actual fueron el último glacial máximo (LGM) y la transición Neolítica (Pinhasi et al., 2012).

Muchos autores proponen que el retroceso de las poblaciones humanas hacia los refugios del sur de Europa durante el LGM y las subsiguientes expansiones hacia latitudes más septentrionales tuvieron un fuerte impacto sobre la diversidad genética europea (Gamble et al., 2004; Arenas et al., 2012). Datos derivados del ADN antiguo (ADNa) han sugerido que la actual estructura genómica europea fue constituida durante el Paleolítico superior y que la misma, deriva de una metapoblación que hubiera experimentado flujo génico en múltiples direcciones (Seguin-Orlando et al., 2014). En contraposición, otros trabajos apuntan a la transición Neolítica como el proceso principal que ha conformado el *pool* génico de los europeos actuales (Fu et al., 2012; Brotherton et al., 2013).

Desde un punto de vista de la diversidad del ADNmt, las acusadas diferencias en la composición de sub-linajes H entre el oeste y el este mediterráneo nos podría forzar a trazar una estructura para el clado H previa a los tiempos Neolíticos. Roostalu et al. (2007) han interpretado esos perfiles distintivos como el resultado de un flujo génico materno limitado entre los extremos de esta cuenca geográfica después del LGM. Nuestros datos indican, además, unos patrones significativamente diferenciales para el clado H entre las penínsulas ibérica e italiana. Este resultado podría ser interpretado como derivado de diferentes sustratos del Paleolítico superior: Magdalenense (17-12 ky) y Epigravetiense (17-10 ky), respectivamente. El hecho de que los mayores valores de diversidad haplotípica de Europa se encuentren en las penínsulas meridionales (Iberia e Italia) reforzaría la hipótesis de una estructuración genética profunda en el tiempo (Auton et al., 2009). En este contexto, ambas áreas se consideran como sitios refugio presentando un registro de continuidad ocupacional desde 17-12,4 ky (Verpoorte, 2009).

Migraciones humanas post-glaciales han sido inferidos a partir de la distribución de los haplogrupos mitocondriales H1, H3 y V (Torroni et al., 1998, 2001; Pereira et al., 2005a) y los linajes del cromosoma Y R1b y I1b2 (Semino et al., 2000; Cinnioğlu et al., 2004; Rootsi et al., 2004). Asimismo, la cultura asociada

con la principal expansión démica ocurrida en Europa durante el LGM fue el Magdalenense (22-13 ky) (Gamble et al., 2006). Sin embargo, otros estudios proponen edades más recientes que las previamente propuestas para los linajes H (~18 ky), H1 y H3 (~11 ky) y, por tanto, una reinterpretación del escenario. Siguiendo a Soares et al. (2010) el haplogrupo H llegó justo tras el LGM a Europa y los sub-clados H1 y H3 reflejarían expansiones poblacionales relacionadas con el clima que habrían coincidido con el final del *Younger Dryas* y no con el LGM. Sin embargo, tal y como los autores señalan, esas edades podrían estar infra-representadas en relación a las verdaderas profundidades temporales, debido a efectos fundadores recientes. Un reciente estudio de ADN_a (Brotherton et al., 2013) ha revelado la discontinuidad genética para el haplogrupo H en Europa, con perfiles privativos encontrados exclusivamente en poblaciones de la Europa central pertenecientes al Neolítico temprano y otras muestras relacionadas a culturas más tardías (Neolítico medio y tardío e individuos de la Edad del Bronce). Los autores propusieron que el Neolítico fue el proceso principal para que el linaje H se convirtiera en el haplogrupo europeo predominante, no reconciliando sus resultados con las expansiones post-glaciales previas al período Holoceno.

Movimientos bidireccionales a través del Estrecho de Gibraltar y/o su región marítima están soportados por los resultados presentados aquí junto con los expuestos en el **Capítulo IV**. En ambos casos, se observan raíces temporales profundas para estos eventos de flujo génico entre Iberia y África. Además, la presencia de haplogrupos específicos de Continente tales como el H para Europa y el U6, M1 y L para África en costas opuestas nos estarían indicando que esos eventos migratorios no habrían sido tan esporádicos como previamente se había considerado. Sin embargo, diferentes historias demográficas habrían afectado a la asimilación de aquellas variantes procedentes del continente vecino en la población sumidero (*sink population*). Claros ejemplos los encontramos en la contribución de los linajes maternos africanos en las poblaciones ibéricas alcanzan, como mucho, el ~15% de la diversidad total mitocondrial (valor encontrado en la población andaluza de Huelva, ver **Capítulo III**); mientras que la mayoría de las variantes mitocondriales detectadas en el norte de África tienen un origen europeo. En conclusión, las migraciones en ambos sentidos (hacia el sur y hacia el norte del Mediterráneo) no serían comparables en términos de su impacto real en el *pool* génico de las poblaciones humanas actuales asentadas en las dos orillas de esa cuenca geográfica.

Las expansiones humanas relacionadas con el clima son responsables de la presencia de la mayoría de las variantes maternas europeas detectadas en el norte de África. Además, los análisis en alta resolución manifiestan la existencia de perfiles locales para el clado H en la franja norte de África, y una aparente falta de identidad con las secuencias ibéricas, en general, y con las andaluzas, en particular. Por tanto, se puede inferir bien que la población que se movió hacia el sur desde el norte de Iberia no se habría asentado en el sur de la Península ni se habría mezclado con la población que habitaba esa área. Nuestros resultados apoyarían inmersiones rápidas y tempranas en el tiempo desde Iberia hacia el

norte de África como ha sido ya postulado por otros autores (Ennafaa et al., 2009; Rhouda et al., 2009). Conexiones antiguas entre ambas orillas han sido inferidas a partir de algunos restos arqueológicos, entre los que destaca el '*arpón de Taforalt*' (Camps, 1974). En este contexto, algunos autores han enfatizado las similitudes entre las industrias del Paleolítico superior en el noroeste africano y el sur de Iberia (Iberomauriense vs. Magdalenense, respectivamente) (Linstädter et al., 2012). Evidencias derivadas del análisis de diferentes sistemas genéticos muestran que el *pool* actual del Mediterráneo occidental se constituyó mayoritariamente en tiempos pre-Neolíticos (Currat et al., 2010). Sin embargo, algunos autores han propuesto que el proceso de neolitización también hubiera presentado paralelismos en ambos lados de las costas (Manen et al., 2007; Linstädter et al., 2012).

La siguiente difusión de estas secuencias H desde áreas costeras del noroeste de África hacia el resto del continente africano no fue homogénea, sino que estuvo fuertemente influenciada por la deriva genética (Ottoni et al., 2010; Pereira et al., 2010b). Por otro lado, las marcadas diferencias entre la estructura del haplogrupo H entre los extremos occidental y oriental de la franja norteafricana pueden ser un reflejo del despoblamiento que caracteriza –incluso en la actualidad (ver Linard et al., 2012)– la costa entre Túnez y Egipto y que, por tanto, habría prevenido el flujo génico horizontal. En este contexto, Marruecos y Argelia se encuentran más influenciados por sus vecinos ibéricos, mientras que las poblaciones del noreste, como Libia, albergan considerables trazas maternas del Oriente Próximo. Túnez, como territorio intermedio, muestra perfiles H muy similares a los de Italia. Por tanto, hemos observado a lo largo de un eje longitudinal mediterráneo una alta identidad genética entre las penínsulas europeas meridionales (Iberia/Italia) y sus vecinos africanos (Marruecos y Argelia/Túnez). El papel de estas áreas-refugio post-glaciales de Europa y las migraciones humanas forzadas hacia el sur podrían haber provocado estos contactos trans-mediterráneos. Algunos episodios históricos han sido explorados también para explicar las conexiones entre Italia y Túnez para otros marcadores genéticos (Fortes-Lima et al., 2014).

Las asociaciones entre las poblaciones humanas de las costas norte y sur del Mediterráneo occidental son particularmente fuertes cuando se analiza la herencia materna en esta área. Este resultado apunta a un papel preferente de las mujeres en el transporte de genes a través de Gibraltar y su región marítima. Sin embargo, otras evidencias genéticas del sur de Iberia apoyan escenarios de migración entre el norte de África y la Península. Nuestro equipo ha encontrado interesantes señales africanas en nuestras poblaciones andaluzas para los linajes del cromosoma Y (Ambrosio et al., 2010a, 2010b), así como otras evidencias de conexiones transcontinentales para marcadores autosómicos (Calderón et al., 2006; Fortes-Lima et al., 2014; Reales et al., 2014).

En un contexto geográfico más restringido, cada fracción del genoma materno analizada por nosotros ha revelado patrones micro-geográficos dentro de Andalucía (ver **Capítulos III y IV**). El mismo resultado surge del análisis en alta

resolución del clado H. Estos patrones paralelos pueden ser bien entendidos desde un planificado proceso de muestreo. La **Tabla V-4** muestra la estructura de sub-linajes H para nuestras dos poblaciones andaluzas en comparación con los perfiles detectados en el otro extremo geográfico de Iberia (Cornisa Cantábrica).

El extremo occidental de Andalucía (Huelva) muestra afinidades para los perfiles H con otras poblaciones occidentales europeas, principalmente con Galicia y Portugal. En la **Tabla V-4** se puede observar esa similitud genética fundamentada esencialmente en la presencia de ciertos sub-clados H concretos y en el valor de diversidad del haplogrupo H. Las conexiones genéticas entre la franja atlántica de Iberia han sido destacadas recientemente para la diversidad mitocondrial global y para otros marcadores autosómicos (ver **Capítulo III** y Fortes-Lima et al., 2014).

Es interesante destacar que otra de las poblaciones que compone la franja cantábrica del norte de España, como son los vascos, a pesar de registrar la mayor frecuencia del clado H (~55%), presenta la menor diversidad asociada. La mayor parte de vascos autóctonos definidos como H (~57%) pertenecen a nodos basales de H1 o bien a ramas nativas y específicas del área vasca (H1*).

Por otro lado, la población oriental de Granada muestra una estructura típicamente ibérica del clado H, enriquecida con otros sub-linajes menos comunes que señalan influencias de otros territorios del continente europeo y que provocan el alto valor de diversidad interna de H en esa población (ver **Tabla V-4**). Los principales linajes H del Mediterráneo oriental que contribuyen al genoma de los andaluces de Granada son H13 y H20, ambos con una edad de coalescencia alrededor de 16-17 ky (ver **Tabla V-3**). La llegada de estos linajes podría ser explicada por una llegada a través del Mediterráneo desde el Oriente Próximo/Anatolia. El proceso de introducción del Neolítico en Iberia habría implicado una colonización pionera marítima (Zilhão, 2001) junto con migraciones dirigidas a ciertas localizaciones favorables para el asentamiento y el desarrollo de la agricultura (Fernández & Gómez, 2009). Apoyando esta hipótesis, el valor más elevado de H13 en Iberia se localiza en Cataluña (en el extremo noreste de la Península), que, por su posición geográfica, es considerada como uno de los corredores para la dispersión del Neolítico temprano ibérico (Zapata et al., 2004). Por otro lado, se encontraron ciertas identidades entre los andaluces orientales con poblaciones actuales y pretéritas de la Europa central (basadas en los haplogrupos H7, H11 y H82). Por lo tanto, se observan diferentes componentes del linaje mitocondrial pan-Europeo H en el sur de Iberia indicando potenciales influencias poblacionales.

Haplogrupo	Cornisa cantábrica			Andalucía		
	Galicia ¹	Cantabria ¹	País Vasco ²	Occidental (Huelva) ³	Oriental (Granada) ³	
H*	10,48	5,77	12,23	10,91	13,66	
H1	37,91	46,15	59,71	43,64	32,16	
H1*	30,65	36,54	56,83	34,55	25,55	
H1a	1,61	7,69	0,72	-	0,44	
H1a1	0,81	-	-	0,91	-	
H1b1	-	-	-	-	2,20	
H1c3	2,42	-	-	3,64	-	
H1e1a3	-	-	-	-	2,20	
H1k	0,81	-	-	0,91	-	
H1t1a1	-	-	-	-	0,44	
H1ah1	-	-	1,44	2,73	-	
H1ba	1,61	-	-	-	0,44	
H1bf1	-	1,92	0,72	0,91	0,88	
H2	6,45	3,85	7,19	2,73	1,32	
H2*	4,03	3,85	7,19	-	-	
H2a1	2,42	-	-	2,73	0,44	
H2a2b	-	-	-	-	0,88	
H3	20,16	17,31	13,67	14,55	18,06	
H3*	18,55	13,46	11,51	12,73	13,22	
H3c	1,61	3,85	0,72	-	4,41	
H3c2	-	-	1,44	1,82	0,44	
H4	7,26	5,77	1,44	2,73	5,29	
H4a*	4,03	1,92	-	-	-	
H4a1	3,23	3,85	1,44	0,91	5,29	
H4a1a4b1	-	-	-	1,82	-	
H5	6,45	9,62	0,72	4,55	7,49	
H5*	-	3,85	-	3,64	2,64	
H5a	6,45	5,77	-	-	2,20	
H5a3a1	-	-	0,72	-	2,20	
H5a4a	-	-	-	0,91	0,44	
H6	4,84	5,77	2,88	7,27	7,05	
H6a	-	-	2,16	5,45	3,96	
H6a1a1a	4,84	3,85	-	-	1,76	
H6a1a7	-	-	0,72	1,82	0,88	
H6a1b4	-	1,92	-	-	0,44	
H7	0,81	-	-	4,55	3,08	
H7*	0,81	-	-	4,55	2,64	
H7h	-	-	-	-	0,44	
H10	1,61	5,77	-	2,73	-	
H10*	0,81	5,77	-	-	-	
H10a1	0,81	-	-	0,91	-	
H10e	-	-	-	1,82	-	
H11	-	-	-	-	2,20	
H11*	-	-	-	-	1,76	
H11a	-	-	-	-	0,44	
H13	1,61	-	-	-	3,96	
H17	-	-	2,16	0,91	2,20	
H17*	-	-	2,16	0,91	1,32	
H17c	-	-	-	-	0,88	
H18	0,81	-	-	4,55	0,44	
H20	-	-	-	-	1,76	
H20*	-	-	-	-	0,44	
H20a	-	-	-	-	1,32	
H27	1,61	-	-	0,91	0,44	
H36	-	-	-	-	0,44	
H82	-	-	-	-	0,44	
n _H	124	52	139	110	227	
n	282	135	255	280	470	
%H	43,97	38,52	54,51	39,29	48,30	
Diversidad	0,85±0,02	0,84±0,04	0,64±0,04	0,85±0,03	0,89±0,01	
Hg (SD)						

Tabla V-4. Perfiles de sub-haplogrupos H en el norte y sur de Iberia basados en estudios de alta resolución. En la última fila se indica el número de individuos pertenecientes al clado H y la diversidad interna de este haplogrupo ¹Alvarez-Iglesias et al., 2009; ²Behar et al., 2012a; ³presente estudio.

En conclusión, los resultados obtenidos sobre el clado H en Andalucía nos han permitido aportar argumentos clave para comprender que el mar Mediterráneo no ha actuado como una barrera al flujo génico humano y, ello es manifiestamente contrastado en el Mediterráneo occidental, que con el Estrecho de Gibraltar ha representado un puente entre continentes, culturas y civilizaciones. El estudio de la región de Andalucía en sus dos extremos geográficos, revela interesantes y desconocidas características de la historia evolutiva del Mediterráneo occidental. Futuras investigaciones basadas en el genotipado masivo del genoma de las poblaciones del sur de Iberia (incluyendo Portugal) junto a otras nativas del noroeste africano nos proporcionarán más detalles acerca del poblamiento del extremo más occidental del Mediterráneo y del papel de la Península Ibérica como receptora y emisora de migraciones humanas.

Capítulo VI

Discusión general

El estudio del ADNmt nos ha permitido desvelar claves acerca del pasado evolutivo de la población humana que habita el sur de la Península Ibérica. En concreto, hemos detectado señales de migraciones –dentro de Iberia y con otras regiones más distantes– y hemos evaluado las posibles barreras geográficas a este flujo génico, con el objetivo de valorar el papel del sudoeste de Europa en el poblamiento del espacio mediterráneo.

Para llegar a tener una mejor comprensión de la estructura poblacional humana es necesario integrar campos del conocimiento diferentes pero complementarios, y que no son estrictamente biológicos. Esta aproximación multidisciplinar busca reconciliar e interpretar visiones derivadas de acontecimientos prehistóricos e históricos, dinámica de población, comportamientos culturales y factores geográficos. Todas estas variables han podido conducir a fuertes efectos sobre el *pool* génico de las poblaciones humanas contemporáneas. El genoma mitocondrial nos ofrece una interesante visión del peso de estos factores desde el estudio de la herencia materna humana.

1. La diversidad mitocondrial en Iberia y en otros territorios y poblaciones del espacio mediterráneo

1.1. Los linajes maternos en Andalucía

Los elevados tamaños muestrales manejados y el uso de información molecular de alta resolución, garantizan un sólido escenario acerca de la variabilidad mitocondrial en Andalucía. Asimismo, la convergencia en los resultados obtenidos, desde el análisis de diferentes facetas del genoma mitocondrial en la población autóctona andaluza nos ha puesto en evidencia también, el adecuado proceso de muestreo que hemos llevado a cabo, un hecho que, sin duda, repercute en la robustez de las inferencias sobre historia evolutiva de la población objeto de estudio.

Los altos niveles de diversidad mitocondrial obtenidos en las poblaciones andaluzas de Huelva y Granada contrastan con lo observado en otras regiones y poblaciones de la Península Ibérica. También, las distribuciones *mismatch* de las secuencias de ADNmt y los test de neutralidad nos han indicado que la historia de las poblaciones autóctonas analizadas no se encuentra definida por escenarios de cuello de botella, aislamiento poblacional y otras manifestaciones de la deriva genética; por el contrario, las poblaciones estudiadas habrían experimentado procesos de expansión demográfica en su historia evolutiva reciente. Un escenario contrapuesto lo encontramos en la franja cantábrica del norte de España, cuyas poblaciones están caracterizadas por valores reducidos de diversidad genética (Salas et al., 1998; Martínez-Cruz et al., 2012). Por ejemplo, los patrones mitocondriales observados en vascos han llevado a proponer que procesos de fragmentación poblacional, aislamiento y deriva hayan sido

fundamentales para modelar el *pool* génico de sus poblaciones (García et al., 2011).

Nos encontramos, por tanto, frente a perfiles distintivos poblacionales en la franja cantábrica y en el extremo meridional de la Península Ibérica. Dado el gran protagonismo que ha experimentado el norte de Iberia en estudios sobre genética de poblaciones, los resultados obtenidos por nuestro equipo en Andalucía (referidos por ejemplo a su alta diversidad genética y a la presencia de señales de interesantes marcadores procedentes de otras áreas continentales) nos lleva a proponer que la atención científica pudiera ser centrada de manera definitiva en la población del sur peninsular. Como se ha expuesto en todo este trabajo, Andalucía ha actuado como puerta abierta y cruce de caminos para numerosas culturas y civilizaciones. Por tanto, cualquier intento de desvelar procesos de poblamiento de la Península Ibérica, en concreto, y del espacio mediterráneo, en general, nos conduce a la necesidad de otorgar un papel principal a la población autóctona andaluza.

Profundizando aún más en el estudio de este territorio y de su población, uno de los principales resultados destacados en la presente Tesis es la existencia de una diferenciación micro-geográfica dentro de Andalucía. Esta información surge tanto del análisis de la diversidad mitocondrial global (**Capítulo III**) como de linajes maternos específicos, los cuales, son considerados marcadores genéticos poblacionales (**Capítulo IV** y **Capítulo V**). La heterogeneidad genética en el sur de España sería un reflejo de esas 'dos Andalucías', diferenciadas sobre la base de factores geográficos, demográficos e históricos (ver sección **II-1.1**).

Los dos polos geográficos de Andalucía, Huelva en el oeste y, Granada, como representante del este, se caracterizan por un amplio espectro de haplogrupos mitocondriales que revelan patrones opuestos y, en cierto modo, complementarios. Mientras que la población occidental de Huelva muestra una elevada frecuencia y gran diversificación de los linajes de origen africano (p.e. U6, L), la de Granada destaca, sin embargo, por la diversidad interna del clado europeo por excelencia (el Hg H).

En cuanto a los linajes africanos, U6a y L1b son los sub-clados protagonistas en Huelva (ver **Capítulo III** y **Capítulo IV**). La presencia de haplogrupos africanos dentro de la Península se detecta principalmente en su extremo occidental, siendo Huelva la población con mayor contribución de estos clados (~15%). En la franja atlántica (Portugal y Galicia) también presentan un peso relevante. Este paralelismo ha llevado a la búsqueda de episodios que vinculen estos territorios, entre los cuales se ha propuesto que rutas comerciales y militares, como la '*Vía de la Plata*', actuasen como vehículo de genes (**Capítulo III**).

En el análisis filogenético del clado H, en Granada, nos encontramos con un sustancial descenso en las frecuencias de H1 (típico sub-haplogrupo de Europa occidental) en beneficio de la aparición de otros sub-linajes H más raros en Iberia (ver **Capítulo V**). Asimismo, en Granada se detectan clados asociados al Mediterráneo oriental (p.e. H13, H20) que no se han detectado en Huelva, lo que

puede ser consecuencia de la posición geográfica de Granada, más abierta a contactos con otras poblaciones mediterráneas.

1.2. Andalucía como puente entre Europa y África

Diversos y complejos procesos migratorios subyacen bajo la diversidad y distribución actual de los linajes africanos maternos en Europa. El ejemplo de la Península Ibérica es claro a este respecto, la cual muestra las trazas africanas más intensas de todo el continente europeo. La intrigante presencia de genes africanos entre las poblaciones ibéricas ha sido constatada sobre la base de diversos marcadores moleculares y, mucha de esa influencia africana, fundamentalmente encontrada en la vertiente más meridional de la Península, no puede ser asociada exclusivamente a episodios históricos. En concreto, en el **Capítulo III** hemos mostrado cómo existe una influencia más marcada de genes africanos en el oeste de Andalucía con respecto a lo observado en Granada (14,6% vs. 3,3%, respectivamente). Este patrón podría discrepar con lo que cabría esperar desde una explicación histórica. El este de Andalucía fue el último bastión musulmán en la Península (Reino de Granada, hasta el final del siglo XV). Además de esas altas frecuencias, en Huelva encontramos una gran diversificación de clados africanos con representación de las 3 radiaciones del haplogrupo U6 (U6a, U6b y U6c) y con relevantes *clusters* dentro de L (L1b, L2a y L2b). Las genealogías maternas (ver **Figura III-3**) nos revelan, además, la presencia de la población de Huelva en los nodos centrales y basales de los linajes U6 y L1b, lo que nos ha conducido a proponer una llegada temprana de estos clados a la Península así como la posibilidad de que algunos sub-linajes dentro de U6a se desarrollaran *in situ* dentro de Iberia y de ahí se hubieran re-expandido hacia el norte de África.

Estos interesantes resultados presentados en el **Capítulo III** conducen a la necesidad de incrementar el nivel de resolución genealógica, con el fin de desvelar el origen, la antigüedad (edades evolutivas) y los patrones de migración asociados a estos linajes mitocondriales ligados al continente africano.

En el **Capítulo IV** se ha mostrado que (1) la mayoría de los genes africanos maternos fueron introducidos en Iberia alrededor del Holoceno temprano con señales incluso de un flujo génico más profundo en el tiempo, (2) Iberia no debe ser considerada exclusivamente como una población sumidero (receptora de genes) sino también como una población fuente y relevante para la difusión sucesiva de genes africanos hacia el resto de Europa así como para la emisión de otras señales genéticas hacia sus vecinos del sur del Mediterráneo, (3) un número significativo de mitogenomas subsaharianos y magrebíes se integran en un marco mediterráneo, compartiendo linajes especialmente con la población vecina de la Península Ibérica. Además, las dispersiones hacia el resto Europa desde África fueron casi totalmente mediadas por Iberia, aunque encontramos trazas de migraciones directas, fundamentalmente hacia otra península meridional, como Italia [ver, por ejemplo, U6a5 (**Dataset S2**) y L3d1b1a (**Dataset S6**)].

Como manifestación de esos contactos transcontinentales, los marcadores con mayor peso son, sin duda, el U6a y el L1b. Dentro de Europa, la filogeografía de estos clados (**Figura IV-1**) gravita en Iberia, y más concretamente, en la vertiente sur y occidental. Por otro lado, el análisis detallado de las filogenias mitocondriales (ver **Datasets S2** y **S4**) nos ha mostrado que la edad evolutiva y la composición poblacional de ciertas sub-ramas (como U6a1, L1b1a6, L1b1a8, L1b1a12 y pre-L1b1a16) pueden ser interpretadas como resultado de una evolución *in situ* en la Península Ibérica, tras una introducción prehistórica desde el norte de África. Según esta hipótesis, las muestras magrebíes intercaladas en estos sub-clados serían marcadores de migraciones de vuelta (*back-migrations*) a África desde Iberia. Del mismo modo, los individuos del resto de Europa integrados en estos *clusters* nos estarían indicando que Iberia ha sido la puerta de entrada y posterior dispersión de linajes africanos hacia el continente europeo.

Estas entradas antiguas en dirección norte, así como las re-expansiones hacia las costas sur mediterráneas son soportadas, además, desde el estudio de las filogenias, y por el análisis fundador (*founder analysis*, FA) de los linajes L y U6. Algunos resultados poco concordantes (ver **Figura IV-5**) se fundamentan en las limitaciones del propio FA para indicar una clara direccionalidad de migración, cuando nos encontramos con situaciones de flujo génico continuado en el tiempo entre dos territorios, como es el caso del Magreb e Iberia.

Como se ha mencionado más arriba, varios estudios han puesto de manifiesto la dificultad de reconciliar la presencia de linajes africanos maternos en Iberia con eventos históricos, siendo su paradigma la dominación musulmana (González et al., 2003; Maca-Meyer et al., 2003b). Más recientemente, un análisis detallado de mitogenomas L ha ampliado el marco temporal para las dispersiones humanas entre África y Europa (Cerezo et al., 2012). Cabe destacar que, habitualmente, el estudio de genomas mitocondriales se centra en un linaje concreto, como por ejemplo el haplogrupo U6 (Pereira et al., 2010a; Secher et al., 2014) o el L (Cerezo et al., 2012). Sin embargo, nosotros hemos abordado, por primera vez, la evaluación conjunta de los principales marcadores africanos que se encuentran en Europa (U6, M1, L1, L2 y L3) en alta resolución filogenética. Nuestras evidencias genéticas, fundamentadas en una completa visión de la herencia materna africana, han mostrado que las relaciones poblacionales entre Europa y África son más antiguas de lo que previamente se pensaba. La mayor parte de los marcadores africanos parecen haberse introducido alrededor del Holoceno temprano. Asimismo, en esta investigación hemos proporcionado argumentos en contra de una visión simplista de las migraciones africanas hacia Iberia.

Otros datos mitocondriales –basados en análisis de muestras de ADN antiguo– apoyan la existencia de episodios migratorios previos a la invasión musulmana. Entre ellos se encuentra la presencia de un individuo perteneciente al linaje M1b1 en una necrópolis de Álava de los siglos VI-VII AD (Alzualde et al., 2006). La detección del linaje africano L3f en población contemporánea de Asturias es otra prueba de ello (Pardiñas et al., 2014). Por otro lado, los perfiles genéticos de especies animales domesticadas pueden ser interpretados como un reflejo indirecto de las migraciones humanas. De este modo, el análisis de linajes

mitocondriales de ganado bovino han puesto de manifiesto introducciones previas a la conquista islámica desde el norte de África a Iberia (Anderung et al., 2005). La presencia continuada del haplotipo 'T1' en diferentes yacimientos peninsulares del período Neolítico, y necrópolis romanas y medievales reforzarían la idea de numerosos contactos poblacionales de costa a costa a lo largo de los últimos milenios (Colominas et al., 2015).

En conclusión, el análisis de mitogenomas (alcanzando la máxima resolución filogenética), el incremento de muestras poblacionales disponibles en la literatura y el uso de metodologías cada vez más sofisticadas (**Capítulo IV**) ha confirmado los resultados e interpretaciones presentadas en el **Capítulo III** respecto a la presencia de haplogrupos africanos en Iberia, y, por tanto, en el continente europeo.

Sin embargo, no solo los linajes africanos nos revelan una entremezclada historia evolutiva de las dos orillas en el Mediterráneo occidental y jalonado por el Estrecho de Gibraltar, sino que esta información surge también desde el análisis del clado H (**Capítulo V**). El haplogrupo H es considerado como el marcador nativo europeo (*eurasiático*) por excelencia. En consecuencia, en el marco del Mediterráneo occidental, el clado H señalaría migraciones de norte a sur, en contraposición a los linajes U6, M1 y L.

Nuestros análisis han mostrado, de nuevo, similitudes entre las poblaciones asentadas en los territorios de Iberia y el noroeste de África, principalmente. En Iberia el clado H registra valores ~50% y se trata del linaje materno europeo más frecuentemente detectado entre las poblaciones norteafricanas. Los niveles se reducen en un gradiente desde el extremo occidental (Marruecos: 38%, Coudray et al., 2009) hasta Egipto (14%, Stevanovitch et al., 2003). No obstante, se debe tener en cuenta que la complejidad y diversificación de este macro-haplogrupo requiere un análisis detallado de sus sub-ramas para descender a patrones locales y regionales.

Iberia y el Magreb comparten, además de elevados valores del linaje H, paralelismos en su estructura interna, siendo H1 y H3 los sub-clados principales en ambas regiones. Los análisis filogeográficos y filogenéticos presentados en el **Capítulo V** confirman la hipótesis propuesta por numerosos autores, según la cual, estas dos radiaciones habrían llegado al Magreb por un movimiento masivo poblacional desde Iberia posterior al LGM (Cherni et al., 2009; Ennaffaa et al., 2009; Rhouda et al., 2009; Ottoni et al., 2010). Por tanto, el norte de África sería el extremo más meridional de estas expansiones post-glaciales que habrían repoblado el resto de Europa desde la Península Ibérica (Achilli et al., 2004; Pereira et al., 2005a).

Las edades de coalescencia estimadas para H1 y H3 en Iberia y el norte de África nos conducen a la misma interpretación, siendo similares para ambas regiones aunque moderadamente más altas en Iberia, indicando un origen más profundo en la Península. Por otro lado, el incremento en la resolución filogenética de las muestras norteafricanas permite observar la existencia de una divergencia local notable, es decir, la presencia de ramas nativas H africanas

(Ennafaa et al., 2009; Ottoni et al., 2010). Curiosamente, la proporción de haplotipos H1 y H3 compartidos entre Iberia y el Magreb es baja. Estos hechos están en correspondencia con la antigüedad en la introducción de estos marcadores maternos. De las dataciones y presencia de las muestras norteafricanas en los nodos basales de H1 y H3 también se revelan migraciones rápidas cruzando el Estrecho hacia el Magreb, sin un papel intermediario preferente de Andalucía.

Los resultados del **Capítulo V**, junto con lo expuesto en el **Capítulo IV**, apoyan movimientos bidireccionales a través del Estrecho de Gibraltar y las áreas marítimas circundantes. Para ambos casos nos encontramos con episodios de flujo génico con una notable profundidad temporal y que, dadas sus manifestaciones en el *pool* génico de estas poblaciones, no se pueden considerar esporádicos o puntuales. Estos resultados nos indican que las poblaciones que han habitado cada orilla del Estrecho han sido conscientes de la existencia del continente vecino desde tiempos remotos.

Las dos orillas, norte y sur, del Mediterráneo occidental han compartido una historia evolutiva entrelazada, y ello se está evidenciando desde la detección de señales en el genoma de sus poblaciones humanas contemporáneas. La proximidad geográfica entre las costas africana y europea ha permitido movimientos múltiples de costa a costa con especial relevancia en el Mediterráneo occidental y central. Contactos tempranos entre Iberia y el Magreb parecen ser evidenciados por las técnicas de navegación conocidas en el Mediterráneo desde tiempos antiguos. La introducción de linajes mitocondriales africanos en Iberia (y, por extensión, en Europa) requirió travesías a través del mar y, por tanto, el control de técnicas de navegación simples. El desarrollo de una intensa actividad marítima comenzó hace alrededor de 20 ky, siendo el mar Mediterráneo uno de los focos más tempranos de navegación humana. Sin embargo, la expansión de rutas marítimas más largas, utilizadas por navegantes en busca de recursos litorales, no tuvo lugar hasta el Neolítico (Broodbank, 2006; Zeder, 2008). En esta dirección parecen mostrarse los datos genéticos provenientes del estudio de diversos marcadores, los cuales revelan que la estructura genética que define a las poblaciones del Mediterráneo occidental es el resultado de procesos migratorios que comenzaron hace 20 ky. Currat et al. (2010) han sugerido que los patrones genéticos observados en las poblaciones asentadas alrededor del Estrecho de Gibraltar serían el reflejo de una estructuración poblacional profunda que no ha sido borrada por eventos más recientes (como por ejemplo el Neolítico).

En concreto, en el sudeste de España y noreste de Marruecos (región marítima de Alborán), la transición entre la economía cazadora-recolectora y la producción agrícola comenzó en el Epipaleolítico, correspondiendo con el Magdaleniense en España y el Iberomauriense en el norte de África. Existen numerosos paralelismos entre estas industrias (Linstädter et al., 2012), por lo que sería probable la ocurrencia de expediciones tempranas a través del mar de Alborán (Souville, 1998).

Numerosos autores destacan los paralelismos en los procesos de neolitización del sur de Iberia y el Magreb. Incluso, se ha señalado que los asentamientos iniciales del Neolítico temprano en la región meridional ibérica pudieran tener un origen norteafricano, debido a las analogías en cerámicas y ornamentos en yacimientos Neolíticos de Orán, el este del Rif y Andalucía; en instrumentos fabricados con huesos humanos en Nerja y el Magreb y, en fin, en el uso de las mismas especies domesticadas (Cortés Sánchez et al., 2012). Hay, por tanto, ciertas características del denominado '*paquete Neolítico*' del sur de la Península que no pueden ser explicadas exclusivamente por una influencia mediterránea, sino que supondrían introducciones desde el continente africano (Manen et al., 2007). Además, la coincidente aparición de las características Neolíticas en el sur de Iberia y norte de Marruecos no obedece a un gradiente temporal esperado de Europa al norte de África (Linstädter et al., 2012). Por el contrario, el desarrollo sincrónico del Neolítico en Andalucía y el norte de África sugiere un modelo rápido de expansión Neolítica, facilitado por el uso de tecnología marítima y de redes de comercio de materiales, previamente establecidas (Cortés Sánchez et al., 2012).

Más tarde, los acontecimientos históricos han reforzado los lazos entre el Magreb e Iberia, expresados por ejemplo en las unidades políticas y administrativas conjuntas experimentadas durante el Imperio Romano y la conquista musulmana. Esto ha llevado a proponer que el extremo occidental del Mediterráneo, el Estrecho de Gibraltar y el mar de Alborán no sean barreras sino puntos de unión entre dos masas de tierra (ver Liang et al., 2013). Así, el norte de África e Iberia representan conjuntamente un espacio autónomo y único, un '*bi-continente*', siguiendo la expresión del sociólogo G. Freyre (Braudel, 1966). Numerosos lazos políticos, socioculturales y económicos han ligado las costas norte y sur del Estrecho durante tiempos históricos. El continuo flujo migratorio entre Iberia y el Magreb implica que los procesos sociales y culturales de estos territorios no pueden ser correctamente entendidos si estos se estudian aisladamente.

Es interesante destacar que, tanto la Península Ibérica como el Magreb son consideradas como áreas excepcionales y con identidad propia en sus contextos continentales. El norte de África se encuentra condicionado por la existencia de dos barreras geográficas: hacia el norte, el mar Mediterráneo, y al sur, el desierto del Sahara, siendo la primera de ellas menos intensa, ya que se ha demostrado una afinidad genética más bien relevante de las poblaciones norteafricanas con las europeas. La particular historia demográfica del norte de África (respecto a tamaños demográficos, subdivisiones poblacionales y escenarios de cuello de botella); su composición poblacional (predominantemente de origen europeo); los intercambios intensos y continuados en el tiempo y la existencia de áreas desérticas extensivas (desde el Delta del Nilo a Túnez), son factores relevantes que han dejado previsiblemente efectos profundos en los perfiles genéticos que definen a sus poblaciones nativas. Además de las barreras geográficas, diferentes períodos climáticos han condicionado también el movimiento de las poblaciones humanas en el norte de África. Esta región ha sufrido una historia de cambios climáticos muy bruscos, con oscilaciones entre fases áridas, durante las cuales

gran parte del sub-continente era inhabitable, y periodos húmedos que convirtieron curiosamente al Sahara en un paisaje húmedo (Brooks et al., 2005).

Aunque los movimientos poblacionales se han ido sucediendo en el Mediterráneo occidental de norte a sur y viceversa, diferentes historias demográficas habrían influido en la asimilación de los genes en cada sentido. El peso de las trazas mitocondriales específicas de continentes (U6 y L para África/H para Europa) detectadas en costas opuestas del Mediterráneo no es equivalente. La potente señal de los marcadores europeos en el norte de África –comparado con la influencia africana en Europa– reflejaría un flujo génico más intenso hacia el sur, o bien, procesos demográficos distintivos en el sur de Europa y norte de África, como el menor tamaño demográfico registrado en norte de África durante los últimos milenios (Biraben, 1979; Biraben et al., 1997). La dinámica espacial y la historia demográfica de las poblaciones debe ser tenida en cuenta para explicar la actual estructura genética humana (Currat, 2012).

El estudio de la evolución de las tendencias temporales del tamaño efectivo poblacional (N_e) para el haplogrupo africano U6 ha revelado interesantes resultados (ver sección IV-3.3 y Figura IV-7). Se debe tener en cuenta que estas representaciones (BSPs, *Bayesian Skyline Plots*) habitualmente se construyen para linajes concretos. Pese a que un haplogrupo no equivale a poblaciones físicamente definidas, los BSPs nos ayudan a resolver cuestiones acerca de la temporalidad de las expansiones poblacionales y la evaluación de historias demográficas distintivas de cada linaje mitocondrial (Atkinson et al., 2009). En el caso del linaje U6, los perfiles de las gráficas obtenidas nos han revelado de nuevo que las condiciones climáticas son factores fundamentales para el movimiento y expansión de las poblaciones humanas. Así, el tamaño demográfico de U6 para el África subsahariana muestra un incremento asociado con la apertura de las comunicaciones en el Sahara durante el Período Húmedo Africano (AHP) (Figura IV-7).

1.3. Barreras y/o corredores geográficos en el Mediterráneo

En la anterior sección hemos señalado que los patrones mitocondriales definen al Estrecho de Gibraltar, y su región marítima circundante, como un puente entre dos continentes. A una escala más amplia, el espacio mediterráneo representa, asimismo, un lugar de confluencia de culturas y pueblos. En palabras del historiador Fernand Braudel, el Mediterráneo es una sucesión de paisajes, mares y civilizaciones aglutinadas unas sobre otras (Braudel, 1987).

Una de las ideas principales que hemos manejado a lo largo de esta Tesis es la permeabilidad del mar Mediterráneo como vehículo de genes y el hecho de que los movimientos habrían sido más efectivos por mar que por tierra. Este mar, más allá de un depósito nutricional, ha sido ante todo una *superficie de transporte* (Braudel, 1987). La historia mediterránea ha estado caracterizada por la ‘conectividad’ entre sus regiones, mediante el establecimiento de redes de comunicación. Dichas regiones han interactuado preferentemente a través del

agua mediante múltiples líneas de comunicación que han discurrido por aquellas zonas donde existe visibilidad de costa. Gracias a la naturaleza montañosa de muchas líneas de costa, la mayoría de la cuenca mediterránea es fácilmente conectable por áreas con visión de costa (ver **Figura VI-1**). Las travesías a larga distancia, de este a oeste o de norte a sur, podrían ser más rápidas y seguras si se realizaban a través de una cadena de islas o costas mediante 'navegación de cabotaje'. Las comunicaciones por tierra serían más complejas ya que solo existe un pequeño número de pasos de montaña o valles de ríos que podrían aproximarse al efecto de las costas. El entramado más avanzado de transporte terrestre se encuentra representado por las vías romanas (Horden & Purcell, 2000).

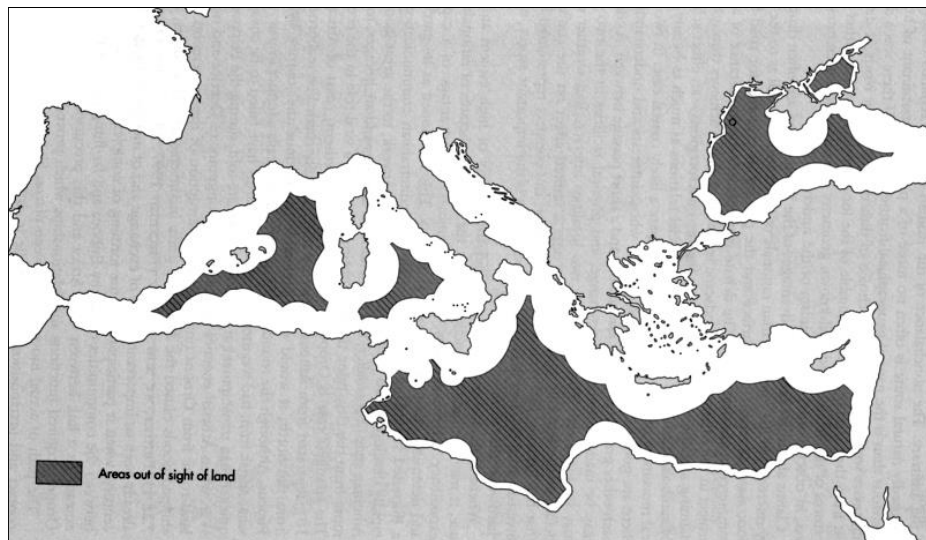
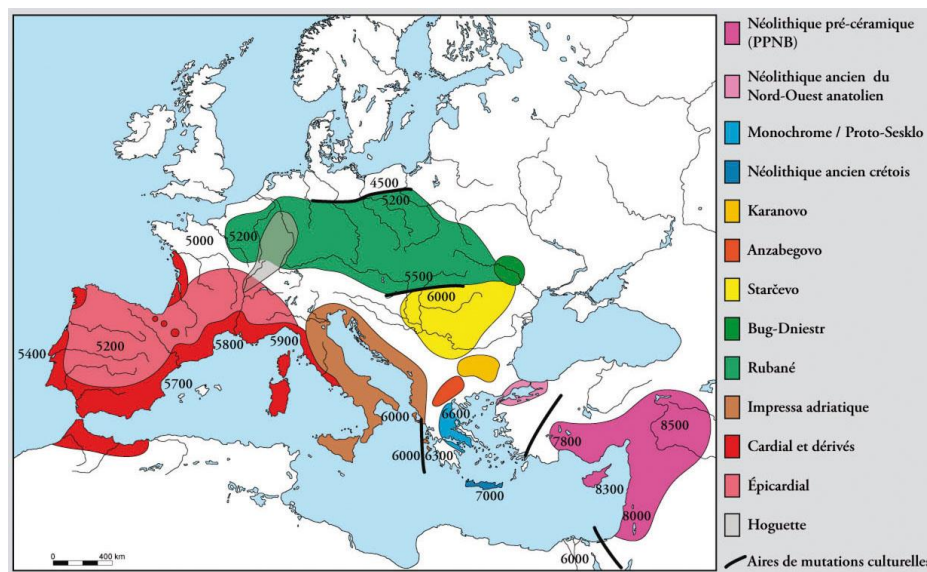


Figura VI-1. Áreas del Mediterráneo desde las cuales *no* hay visibilidad de la costa. Fuente: Horden & Purcell, 2000.

Antes de convertirse en un corredor, el mar fue durante mucho tiempo un obstáculo. Sin embargo, a partir del Paleolítico Superior, existen evidencias tempranas de navegación por la cuenca mediterránea (Straus, 2001; Broodbank, 2006) aunque fue a partir del Neolítico cuando la actividad marítima se estableció definitivamente en esta área. En este período, por primera vez, fueron las penínsulas del sur (Grecia, Italia e Iberia) las que actuaron como puertas de entrada de la neolitización hacia áreas internas (ver **Figura VI-2**). En un espacio de unos cuatro milenios, el mar Mediterráneo se convirtió en una autopista para la transmisión lateral de personas y tecnologías (Broodbank, 2006).

Figura VI-2. Propagación cronológica en Europa de la economía Neolítica. En color se destacan las culturas representadas. Adaptado de Guilaine, 2003.



Los movimientos marítimos desde el este pudieron introducir ciertos linajes maternos en Andalucía. La población andaluza de Granada, dada su posición geográfica más abierta a contactos trans-mediterráneos ha podido recibir influencias, y algunas detectables, dentro de la estructura del haplogrupo H, como por ejemplo los sub-clados H13 y H20 (ver **Capítulo V**). Estos marcadores pudieron haber llegado a Andalucía oriental a partir del Neolítico.

El estudio del clado H en alta resolución filogenética nos ha revelado, además de la afinidad genética entre Iberia y el Magreb, comentada en la sección anterior, una estructuración genética significativa en Europa y el Mediterráneo preferentemente en un eje longitudinal (oeste-este). Uno de los sub-haplogrupos que contribuyen más determinadamente a esta estructuración es el H1, que muestra un marcado gradiente (variación clinal) desde el Mediterráneo occidental al oriental.

En el marco de esta estructura poblacional mediterránea, hemos destacado a lo largo de este trabajo de Tesis cómo el norte de África no puede ser considerado como un territorio genéticamente homogéneo a lo largo de toda su extensión longitudinal. Por el contrario, una barrera genética es patente entre el cinturón del Magreb y el noreste de África. Ambos extremos de la franja norteafricana se encuentran fuertemente influenciados por sus vecinos de territorios septentrionales. El noroeste de África habría recibido intensas migraciones post-glaciales desde Iberia, acontecimientos que han determinado su estructura genética actual (Ennafaa et al., 2009). Por otro lado, el noreste africano ha recibido flujo génico del este africano gracias al corredor del valle del Nilo y del Oriente Próximo (Fadhlaoui-Zid et al., 2011). También es interesante enfatizar la fuerte afinidad genética que existe entre un punto intermedio de la franja norteafricana, Túnez, con sus vecinos europeos (Italia). Este patrón es muy marcado para la disección del linaje H (**Capítulo V**).

La discontinuidad genética en el norte de África se encuentra fundamentada en la existencia de áreas desérticas. La ausencia de sitios de abastecimiento entre el Delta de Nilo y Túnez ha afectado a los movimientos humanos desde tiempos remotos. La región noroccidental de Libia, la Cirenaica, fronteriza con Egipto, es un área aislada de sus vecinos orientales por 700 km de desierto hasta el Delta del Nilo. Su franja costera no es tampoco hospitalaria, existiendo una gran distancia hacia el territorio relativamente fértil de la Tripolitania, en el oeste. Por tanto, el desierto de la Cirenaica representaría una zona fronteriza que dificulta el transporte por tierra. Al contrario, el mar proporcionaría una red de comunicaciones más útil (Horden & Purcell, 2000).

Incluso hoy en día, estos factores son perceptibles en la distribución de la población humana en el continente africano. En la **Figura VI-3** se observa de manera muy ilustrativa los patrones de asentamiento poblacional en África, en los cuales influyen notablemente la presencia de áreas desérticas y la cercanía a masas de agua. Como indican Linard et al. (2012), el 90% de la población de África se concentra en el ~20% del territorio.

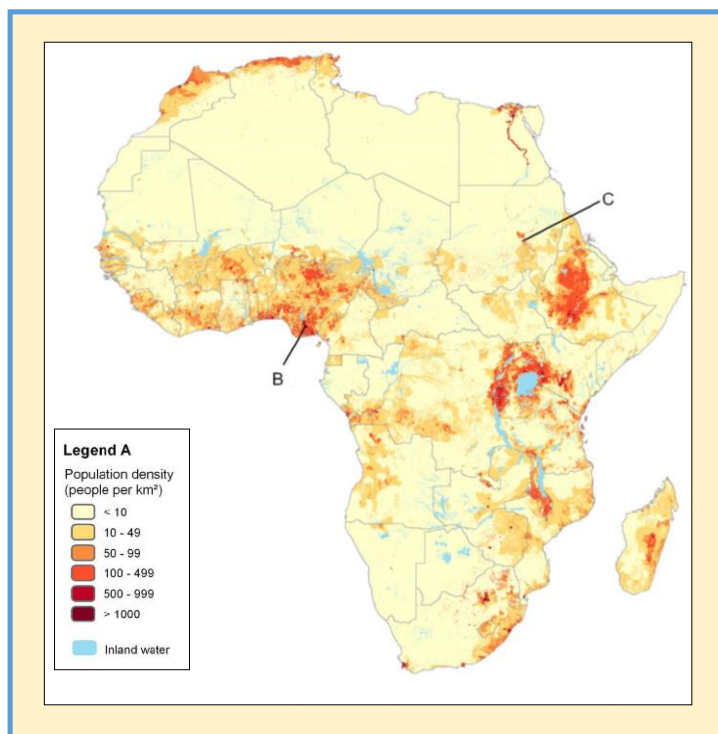


Figura VI-3. Distribución espacial de la población de África en el año 2010. Fuente: Linard et al., 2012.

Esas barreras geográficas poseen en muchas ocasiones un peso más relevante que los factores culturales. En el norte de África, se ha propuesto que la diferenciación genética se deba a la geografía y no a la cultura, ya que los patrones mitocondriales son dependientes de la localización espacial de las

poblaciones, y no de su pertenencia a grupos árabes o bereberes. Por tanto, se demuestra que la dominación árabe en la región tuviera un papel más cultural que demográfico (Ennafaa et al., 2009).

Otros marcadores que muestran esa distinción entre noroeste y noreste de África son los haplogrupos mitocondriales africanos U6 y M1 (**Capítulo IV**). El linaje U6 posee un carácter atlántico, estando prácticamente restringido al noroeste de África y sudoeste de Europa (ver **Figura S3**). El M1, por el contrario, tiene tendencia a mostrarse en el Mediterráneo oriental y en el este de África, con las mayores frecuencias registradas en las poblaciones de Etiopía y Kenia (ver **Figura S4**).

2. Consideraciones metodológicas

A lo largo de la presente Tesis hemos aplicado diferentes metodologías para el análisis de la diversidad mitocondrial en el sur de Iberia, en función del objetivo concreto a alcanzar.

La estrategia más adecuada para una evaluación global de los patrones de ADNmt comienza por la secuenciación de las regiones hipervariables (regiones control). De este modo, una pequeña fracción del genoma mitocondrial nos permite realizar una pre-clasificación de los individuos en haplogrupos y sub-haplogrupos. Sin embargo, en este punto nos encontraremos con dos problemas principales: (1) dado que la región control posee una alta tasa de mutación, el fenómeno de '*homoplasia*' o mutación recurrente hace difícil la asignación a linajes concretos y conduce, por tanto, a incertidumbres filogenéticas; (2) aquellos individuos pertenecientes a haplogrupos cercanos filogenéticamente a la secuencia de referencia -rCRS, haplogrupo H- tienden a mostrar muy pocas posiciones diagnósticas en la región control, por lo que es imprescindible efectuar más análisis genéticos.

El genotipado de ciertas variantes de la región codificante permite solucionar los problemas mencionados anteriormente, consiguiendo una clasificación inequívoca en haplogrupos mitocondriales. En el estudio sobre la variación global del ADNmt (**Capítulo III**) empleamos la técnica de PCR-RFLPs para contrastar la información obtenida a partir de las secuencias hipervariables. Esta técnica posee la ventaja de su sencillez y bajo coste económico, pero también muestra un reducido rendimiento, ya que no permite genotipar más de una variante a la vez. A pesar de las limitaciones que se asocian a los estudios basados en RFLPs, entre las cuales se encuentra la baja resolución molecular para distinguir variación a nivel sub-regional (Kivisild, 2015), cabe señalar que nuestros análisis presentados en el **Capítulo III** han permitido descubrir una micro-diferenciación dentro de Andalucía, patrón que se ha mantenido en los subsiguientes análisis genéticos a mayor resolución.

Estos resultados nos llevaron a la necesidad de centrarnos en dos parcelas de la diversidad materna responsables de la discontinuidad genética del territorio: los linajes africanos (**Capítulo IV**) y el clado H (**Capítulo V**). Dado que contábamos con un volumen total de muestras autóctonas superior al empleado en el **Capítulo III** (n= 279), planteamos localizar en el resto de dichas muestras a todos los individuos pertenecientes a estos haplogrupos de interés. Para ello, diseñamos ensayos de genotipado TaqMan, con los que realizamos una clasificación eficaz y rápida de los individuos, para posteriormente someterlos a análisis más detallados.

La secuenciación completa del genoma mitocondrial permite descender de la manera más refinada posible en la filogenia humana. Esta información de alta resolución es requerida para la definición de nuevas sub-ramas y permite la reconstrucción de los árboles con menores ambigüedades. La información proporcionada a partir de las secuencias completas es imprescindible para revelar patrones genéticos regionales, para valorar los cambios en los tamaños efectivos poblacionales femeninos a través del tiempo y para conocer comportamientos específicos de acuerdo con el sexo en la migración humana y mezcla poblacional (Kivisild, 2015). Todos estos enfoques han sido abordados en nuestro análisis de la herencia materna africana en Andalucía (**Capítulo IV**).

En el **Capítulo V** hemos podido comprobar cómo la aparente homogeneidad de algunos clados mitocondriales desaparece al incrementarse el nivel filogenético de análisis. Aunque sin alcanzar la máxima resolución posible, el genotipado múltiple de SNPs para la determinación de la estructura del clado H ha permitido una significativa reducción de las muestras sin clasificar (H*) y una precisa asignación a sub-haplogrupos. Este nivel analítico nos ha desvelado la diversidad que subyace tras los altos valores del haplogrupo H en Andalucía. Cabe destacar aquí que, dado el alto número de muestras pertenecientes al haplogrupo H (n= 337), resultaría inviable plantear la secuenciación completa de todas ellas, por lo que la metodología SNaPshot proporciona un balance adecuado entre rendimiento filogenético y las limitaciones temporales y económicas.

Durante el desarrollo del procedimiento experimental del **Capítulo V** nos encontramos con problemas no esperados en cuanto a la información genética obtenida. Dos polimorfismos diseñados para caracterizar a dos sub-haplogrupos concretos, produjeron una situación de ambigüedad que solo se pudo resolver mediante la consulta de la filogenia mitocondrial humana. En la '*multiplex 2*' (ver sección **II-5.5**), la variante 13708 se incluyó en el diseño para identificar aquellas muestras pertenecientes al sub-clado H2a5. En el análisis de los datos se detectaron 6 muestras con la variante 13708, pero que carecían de la variante 4769, polimorfismo que determina la asignación al linaje parental de H2a5, H2a. Una búsqueda exhaustiva en *PhyloTree* nos mostró que esta mutación (13708) se encontraba también definiendo al haplogrupo H18. Del mismo modo, 6 muestras inicialmente clasificadas como H6a, por su posesión de la mutación 3915, carecían en su secuencia control de los polimorfismos típicos del linaje H6 (16362, 16482, 239). Esta situación, de nuevo, nos llevó a buscar en la filogenia y descubrir a la variante 3915 definiendo también al linaje H17. Estos dos ejemplos

nos muestran cómo, al margen de la región control, el resto del genoma mitocondrial está también caracterizado por un considerable nivel de homoplasia, y que el hecho de poseer un buen nivel de resolución e información tanto de la región control como de la codificante nos permite resolver estos conflictos filogenéticos.

3. Los patrones genéticos en Andalucía: otros enfoques moleculares

La perspectiva matrilineal obtenida en la presente Tesis Doctoral sobre la composición genética de las poblaciones humanas de Andalucía está siendo complementada por análisis de otros marcadores genómicos. Estas aproximaciones han dado lugar a un número apreciable de artículos científicos (Calderón et al., 2006, 2015; Ambrosio et al., 2010a, 2010b, 2012; Fortes-Lima et al., 2014; Reales et al., 2014).

3.1. Linajes paternos en el sur de Iberia

Nuestro equipo ha analizado, paralelamente al estudio del ADNmt y de otros marcadores genéticos, la herencia patrilínea de las mismas poblaciones andaluzas (manuscrito en preparación). Esta información es clave para establecer correspondencias entre las historias materna y paterna de la región más meridional de la Península Ibérica. La **Tabla VI-1** muestra, de manera resumida, los linajes del cromosoma Y (C-Y) presentes en las dos subpoblaciones.

Tabla VI-1. Frecuencia (%) de los haplogrupos del cromosoma Y detectados en Andalucía occidental y oriental (Ambrosio et al., 2010b; Ambrosio, 2011).

Haplogrupo (C-Y)	Huelva (N=167)	Granada (N=249)
R	62,28	65,46
E	11,98	6,83
J	7,19	9,24
G	7,19	4,42
I	4,79	9,64
T	4,79	2,81
L	1,20	0,40
Q	0,00	0,80
F	0,60	0,00
A	0,00	0,40

Al contrario de lo revelado en el análisis de los linajes del ADNmt expuesto en la presente Tesis, el estudio de la diversidad paterna no revela una subestructuración estadísticamente significativa dentro de Andalucía ($F_{ST}= 0,0003$;

$P= 0,3694$) (manuscrito en preparación). Sin embargo, algunas interesantes señales distintivas en cuanto a la aparición de haplogrupos en el extremo occidental u oriental de Andalucía han sido observadas.

Ambas poblaciones andaluzas, Huelva y Granada, presentan un patrón propio del sudoeste europeo en cuanto a su diversidad patrilineal. Así, el clado más frecuentemente detectado es el R, y dentro del mismo, el sub-linaje R1b1a2 (M269), que alcanza en las dos poblaciones ~60%. Este haplogrupo se podría considerar como equivalente al clado mitocondrial H, por su frecuencia y distribución en Europa. También, y de forma similar al linaje materno H, el R-M269 muestra un gradiente de frecuencias desde el este al oeste del continente, oscilando del 14% en Anatolia (Cinnioğlu et al., 2004) hasta más del 80% en población vasca (Alonso et al., 2005). En la actualidad, existe un fuerte debate acerca del origen evolutivo y mecanismo de dispersión por Europa del R-M269 (Rosser et al., 2000; Semino et al., 2000; Balaesque et al., 2010). Algunos autores han propuesto que este clado represente señales de movimientos humanos post-glaciales desde Iberia hacia el resto de Europa (Semino et al., 2000; Cinnioğlu et al., 2004), lo que paraleliza con lo sugerido para los haplogrupos mitocondriales H1, H3 y V (Torroni et al., 1998, 2001; Pereira et al., 2005a).

Los linajes E y J del cromosoma Y son particularmente interesantes para estudiar los procesos de poblamiento del Mediterráneo. Por ello, estos dos haplogrupos se estudian habitualmente de manera conjunta (Semino et al., 2004; Cruciani et al., 2007). La distribución geográfica de sus sub-linajes ha sido explicada principalmente como resultado de migraciones que han tenido lugar desde el Neolítico y, en particular, aquellas que han ocurrido durante la protohistoria e historia hasta tiempos recientes (Ambrosio et al., 2010a, 2010b). Algunos de estos sub-haplogrupos están presentes y de manera frecuente en regiones costeras, de ahí que se haya propuesto que su entrada se realizara por vía marítima (Di Giacomo et al., 2004).

El haplogrupo E registra frecuencias más elevadas en Andalucía occidental (~12%) que en la población oriental (6,8%), y además muestra una diversificación más profunda. En la población autóctona de Huelva se encuentran representados 7 sub-haplogrupos E, frente a 4 en Granada. El linaje E (principalmente su radiación E1b1b1) presenta una fuerte estructura filogeográfica, siendo más frecuente en África y más moderado en otras regiones del Mediterráneo, incluyendo el sur de Europa y el Levante. El sub-clado E1b1a se asocia, sin embargo, con el África subsahariana (Cruciani et al., 2004). En las dos poblaciones andaluzas analizadas por nuestro equipo, esta última radiación no está presente, es decir, todos los individuos pertenecientes al linaje E son a su vez E1b1b1 (E-M35).

Descendiendo a los sub-clados dentro de E1b1b1, el E-M81 es conocido como el '*marcador bereber*', alcanzando frecuencias mayoritarias en dichas poblaciones (Cruciani et al., 2004; Semino et al., 2004). La amplia distribución geográfica, pero no homogénea, de los cromosomas E-M81 en Iberia no parece concordante con las zonas donde la ocupación islámica fue más intensa y prolongada. Por tanto, este hecho reforzaría la hipótesis de que los movimientos migratorios

entre el Magreb e Iberia no hubieran estado restringidos a este período histórico (Ambrosio et al., 2010a). Por otro lado, el E-M78 podría estar relacionado con el incremento en flujo génico relacionado con el comercio, principalmente, aquel dirigido hacia el Reino de *Tartessos* (Ambrosio et al., 2010b). Esta hipótesis vendría apoyada por la ausencia de este marcador en la población andaluza oriental (Huelva: 1,2%; Granada: 0%). Del mismo modo, los niveles de E-V13 en Andalucía (mucho más frecuentes en el oeste que en el este, 4,2% frente a <1%, respectivamente) pueden estar ligados a contactos con poblaciones que hubieran viajado a través del mar desde áreas bajo control griego durante tiempos protohistóricos, cuando el Reino de *Tartessos* ganó relevancia estratégica debido a su riqueza en minerales.

Otro sub-clado patrilineal interesante es el E-M123 (y su radiación derivada E-M34), el cual se encuentra en relativas altas frecuencias entre la población judía. Por tanto, se ha relacionado su presencia con los asentamientos en Andalucía, territorio donde hubo comunidades judías desde al menos la época del Imperio Romano (García Iglesias, 1978).

Por otro lado, el haplogrupo J del C-Y registra un 7,2% en Huelva y un 9,2% en Granada. Este linaje muestra una distribución clinal a través de la Europa mediterránea, con picos de frecuencias en el Oriente Medio y el sudoeste de Asia (30-80%) (Ambrosio et al., 2010b). En cuanto a sus radiaciones, la presencia del sub-clado J1-M267 sería consecuencia de la migración de población árabe durante la expansión islámica, mientras que J2-M172 podría estar relacionado con las colonias griegas y fenicias establecidas en la Península debido a un comercio marítimo ligado a los metales. En ambos casos, estos movimientos fueron altamente direccionales y los migrantes solo se asentaron en determinadas regiones de Iberia (Ambrosio et al., 2010b).

El linaje G2a (P-15) es el clado paterno más frecuente del G en Europa occidental. Varios autores (Semino et al., 2000, 2004) han relacionado la distribución de este haplogrupo en el Mediterráneo con la expansión de la agricultura durante el Neolítico. Otra hipótesis que se maneja para su distribución espacial propone que la dispersión de G2a en Europa estuviese ligada a la expansión del Imperio Romano, lo que está apoyado por altas frecuencias en Italia. El peso de este marcador disminuye según nos alejamos de los límites de lo que fue dicho Imperio (Ambrosio, 2011).

Dentro del haplogrupo I, destacan los sub-clados I1 (I-M253) e I2b1 (I-M223), cuya distribución ha sido relacionada con una dispersión post-glacial desde la región franco-cantábrica (Rootsi et al., 2004). Estas dos sub-ramas presentan valores elevados en el centro y norte de Europa, y su presencia en Andalucía podría explicarse por la llegada de los visigodos.

Por último, algunos autores han relacionado el linaje T-M70 con poblaciones judías (Nogueiro et al., 2010). La considerable frecuencia de este clado detectada en Huelva (4,8%) podría asociarse con los altos valores registrados del haplogrupo mitocondrial K (11,4%), el cual se encuentra especialmente representado entre judíos (Behar et al., 2004).

Los resultados derivados del análisis del C-Y en población andaluza, pese a no ser del todo equiparables a los perfiles mitocondriales, presentan rasgos semejantes. Para ambos sistemas uniparentales, nos encontramos con dos poblaciones determinadas por un *background* génico característico de población europea occidental, con frecuencias mayoritarias de los haplogrupos H (ADNmt) y R-M269 (C-Y) enriquecidas con influencias de diversas áreas continentales. Estos hallazgos nos muestran la existencia de relevantes procesos de flujo génico con poblaciones del área mediterránea y otras regiones geográficamente próximas.

Asimismo, el estudio de una batería de microsatélites del C-Y (Y-STRs) revela altas diversidades haplotípicas para ambas poblaciones, y un patrón equivalente a lo anteriormente expuesto, esto es, cercanía con el resto de la Península Ibérica pero con trazas de haplotipos característicos de poblaciones norteafricanas y de otras regiones mediterráneas más distantes, como el Oriente Próximo o el sudoeste de Asia (Ambrosio et al., 2012).

Cabe destacar que estas influencias poblacionales observadas en el *pool* génico andaluz no son detectables en el análisis de los marcadores culturales de herencia paterna: los apellidos. Nuestro reciente estudio (Calderón et al., 2015) muestra una baja asociación entre apellidos y la diversidad del C-Y, incluso en el caso de los apellidos más minoritarios. Por tanto, un gran número de los descendientes de población que inmigró desde regiones geográficas distantes y se estableció en Andalucía (lo que se infiere de la presencia de linajes del C-Y prevalentes en el mundo mediterráneo, como E-M81, J1-M267, J2-M172, I1-M253 o G2a-P15) adoptaron los apellidos españoles más comunes. Se debe subrayar que el análisis de los apellidos más frecuentes en nuestra muestra andaluza ha mostrado un buen ajuste con respecto a lo observado en Andalucía y en el territorio nacional. Esta información revela, de nuevo, la representatividad de nuestras muestras con respecto a los patrones característicos de Andalucía.

Las principales diferencias entre los perfiles mitocondriales y del C-Y en Andalucía se fundamentan en (1) la continuidad o discontinuidad genética a lo largo del territorio y (2) la asimilación diferencial de genes provenientes de África. La marcada influencia africana detectada en el genoma mitocondrial es una señal de identidad de la población andaluza occidental (Huelva), pero no es tan intensa para el C-Y. La falta de correspondencia entre los patrones genéticos maternos y paternos de las poblaciones humanas es reflejo de diferencias intrínsecas en el comportamiento migratorio y reproductivo entre sexos. Numerosos estudios han detectado una mayor diversificación entre áreas continentales cuando se analiza la diversidad del C-Y, comparado con lo encontrado para el ADNmt (ver Lippold et al., 2014). Estos datos se han interpretado habitualmente como consecuencia de los patrones de migración más frecuentes y recurrentes en el caso de la mujer, en comparación con el varón. Por tanto, las barreras genéticas no son tan intensas en el caso de los linajes maternos (Seielstad et al., 1998). Además de diferencias entre tasas de migración de sexos, existen otros factores que contribuyen a la observación de diferentes patrones entre los sistemas genéticos de herencia uniparental. Entre estos

procesos, se encontrarían las diferencias en el tamaño efectivo poblacional entre hombres y mujeres, relacionadas con la varianza en el éxito reproductor debido a fenómenos como, por ejemplo, la poliginia (Heyer et al., 2012).

3.2. Algunos resultados sobre marcadores autosómicos

Además de la herencia uniparental, nuestro equipo de investigación también ha abordado al estudio de algunos marcadores autosómicos en el sur de la Península Ibérica. La caracterización de alotipos de las inmunoglobulinas humanas en Andalucía occidental y oriental ha revelado interesantes resultados (Calderón et al., 2006; Fortes-Lima et al., 2014).

La presencia del haplotipo subsahariano *GM*1,17 5** en nuestras poblaciones andaluzas es alrededor del doble de lo encontrado en el resto de Iberia. Una excepción a este patrón, lo constituye la población del noroeste de la Península (Galicia), la cual registra valores próximos a los de Andalucía (Calderón et al., 2007). Las similares frecuencias de este marcador africano en dos regiones distantes y caracterizadas por una influencia muy desigual de la dominación musulmana, podrían ser explicadas por la conexión entre el sudoeste y noroeste de Iberia por la '*Vía de la Plata*'. Esta idea ha sido explorada en la presente Tesis para explicar la distribución de linajes mitocondriales africanos en la fachada atlántica de la Península (**Capítulo III**). En un marco geográfico más amplio, Huelva y Granada aparecen próximas al centroide de los mapas genéticos (a pesar de su situación geográfica en un extremo del espacio mediterráneo), lo que se interpreta en base a la larga y compleja historia de este territorio. Los resultados, por tanto, indican que diversas migraciones prehistóricas e históricas hacia el sur de Iberia habrían tenido un gran impacto en el escenario genético ibérico.

Por otro lado, se han encontrado valores similares del haplotipo *GM*1,17 23' 21,28* en el sudoeste de Europa (25%) y en el Magreb (28%), indicando una historia genética compartida entre ambas orillas del Mediterráneo occidental. Esta situación sería correspondiente con algunos componentes del *pool* mitocondrial, en concreto, aquellos referidos al clado H, los cuales tienen un peso equivalente en estos dos territorios (**Capítulo V**). Estos resultados nos estarían indicando que el poblamiento del Magreb tiene que haber ocurrido, al menos en parte, desde la Península Ibérica (Fortes-Lima et al., 2014).

Asimismo, la observación de las frecuencias alélicas del gen *APOE* en nuestras poblaciones indica un enriquecimiento del alelo $\epsilon 4$ en el sur de España (y también en Portugal), con valores más elevados a lo esperado debido al marcado gradiente sur-norte que se observa en Europa. La magnitud de dichas frecuencias tiene un paralelismo con lo encontrado en la franja del noroeste africano. Por tanto, la interpretación de estos patrones se fundamenta, de nuevo, en el efecto de múltiples episodios de flujo génico a través del Estrecho de Gibraltar. Este abordaje supone, además, un acercamiento al análisis de las relaciones entre la composición genética de las poblaciones humanas y el desarrollo de

enfermedades complejas, ya que la apolipoproteína E se encuentra relacionada con desórdenes como, por ejemplo, el Alzheimer (Reales et al., 2014).

Cabe señalar que, del mismo modo que en el caso de la herencia paterna, estos dos marcadores autosómicos no revelan una heterogeneidad genética en el sur de Iberia (Fortes-Lima et al., 2014; Reales et al., 2014). Por tanto, es la historia femenina de la población la que desvela una marcada diferenciación entre el extremo occidental y oriental de Andalucía.

La información mostrada a partir de los marcadores estudiados (tanto autosómicos como uniparentales) nos proporciona claves interesantes acerca de la composición y estructura genética de las poblaciones ibéricas así como de la asimilación de genes provenientes de África. La influencia de linajes africanos en Iberia no muestra además un patrón simple e isotrópico. El peso de estas señales poblacionales varía dependiendo del marcador analizado (Currat et al., 2010), pero en todos los casos nos encontramos con situaciones complejas difícilmente interpretables como el resultado de un evento único de introducción de genes desde el norte de África. Al contrario, ese escenario revela historias genéticas compartidas entre ambas orillas del Mediterráneo occidental. De manera recurrente en todos los marcadores, las trazas tienen tendencia a ser más intensas en la región sudoccidental de Iberia, extendiéndose por el resto de la fachada atlántica de la Península. De este modo, es frecuente observar similitudes genéticas en todo este rango geográfico, desde Huelva (extremo meridional) hasta Galicia, pasando también por la vecina Portugal.

4. Perspectivas de futuro

Los resultados presentados en esta Tesis (basados en el genoma mitocondrial) y los que se derivan de otros marcadores moleculares nos muestran el papel principal que ha tenido Andalucía como cruce de caminos y puerta abierta para numerosas culturas de todo el espacio mediterráneo. Por tanto, su población autóctona debe ser protagonista de nuevas investigaciones a nivel genómico más prospectivo. Además, la representatividad de nuestro *stock* de muestras andaluzas –fundamentada en un cuidado proceso de muestreo y en las evidencias desde los resultados que estamos obteniendo– justifican su uso en otros estudios que puedan seguir aportando avances al conocimiento.

Uno de los enfoques más atractivos y actuales en el marco de la Genética de Poblaciones es relacionar estructura, composición genética y enfermedades. En este sentido, los nuevos avances en nuestra comprensión del genoma humano están convirtiendo al estudio de las relaciones entre genotipos y fenotipos de enfermedades en uno de los objetivos centrales de la Biología y de la Medicina.

La β -talasemia es una hemoglobinopatía de gran interés antropológico, en la medida de que pudiera estar sometida a selección positiva en áreas endémicas para la malaria, (ver sección I-1.2). Para desvelar la estructura genética que

subyace en ese desorden, hemos obtenido muestras de pacientes de Andalucía occidental (Huelva) con la intervención directa del Dr. J.N. Rodríguez (Hospital Juan Ramón Jiménez). Por otro lado, nuestras muestras poblacionales de Huelva están siendo empleadas como población control. Los resultados preliminares indican la existencia de un complejo sustrato genético. Hay varias mutaciones deletéreas (talasémicas) que determinan la manifestación de la enfermedad en pacientes y, en población control, se detecta una heterogeneidad en la aparición de otras variantes en el gen de la β -globina. Dado que la β -talasemia es considerada como un importante marcador antropogenético en el Mediterráneo (Penman et al., 2009), el enfoque poblacional, y no exclusivamente clínico –como suele ser habitual en este tipo de estudios– es fundamental en términos de política sanitaria y para la evaluación de las historias genéticas de poblaciones vinculadas a espacios geográficos concretos.

Por otro lado, nos interesa relacionar variantes genéticas con rasgos fenotípicos poblacionales mediante el análisis de las bases moleculares de una de las adaptaciones bioculturales por excelencia: la persistencia de la lactasa (LP). La capacidad para digerir la lactosa (el principal azúcar de la leche) en la mayoría de las poblaciones disminuye rápidamente después del periodo de lactancia, debido a una reducción progresiva del enzima lactasa en el intestino delgado. La persistencia de la enzima lactasa –la cual hidroliza la lactosa en glucosa y galactosa– se hereda de forma autosómica dominante y permite la producción constante del enzima lactasa durante la vida adulta. Como se ha expuesto en la sección **I-1.2**, la LP en ciertas poblaciones humanas de diferentes continentes se ha explicado mediante un mecanismo de co-evolución entre genes y cultura. Esta hipótesis reflejaría la adaptación a los nuevos patrones de dieta que surgieron en el Neolítico con el desarrollo de la agricultura y la domesticación de animales.

Estas dos aproximaciones serán estudiadas por nuestro grupo de investigación con un enfoque tanto genómico como antropológico y que constituye uno de los ejes centrales de nuestro recientemente concedido proyecto de investigación I+D+i (CGL2014-53985-R). En población ibérica no hay muchos estudios con una visión poblacional sobre mutaciones relacionadas con estos caracteres, por lo que nuestras contribuciones pueden ser muy apreciadas por la comunidad científica.

Además, con el objetivo de continuar avanzando en el conocimiento de la variación del genoma humano, tanto neutral como causada por selección, los trabajos antropogenéticos más novedosos se están orientando hacia una nueva fase experimental. Para ello, se analizan densos paneles de SNPs y la información genética resultante se procesa e interpreta mediante avanzadas y complejas herramientas bioinformáticas (Ragoussis, 2009). Los estudios genómicos a gran escala (GW, *genome-wide*) nos ofrecen una valiosa información acerca de la estructura de las poblaciones humanas, datos que complementan a los revelados a partir del análisis de regiones genómicas concretas: (1) en primer lugar, el elevado nivel de resolución alcanzado con estos análisis masivos de SNPs permite la detección refinada de relaciones de parentesco interpoblacional, así como de historias de mezcla y sub-estructuración poblacional (Novembre &

Ramachandran, 2011); (2) nos permiten reconstruir la ancestría individual o local de las poblaciones, lo que presenta un especial significado para poblaciones geográficamente próximas pero con historias evolutivas, demográficas y socio-culturales distintivas (Rosenberg et al., 2003; Price et al., 2008; Pickrell & Pritchard, 2012); (3) por último, estos enfoques nos conducen a la evaluación de señales de selección en el genoma humano y desvelar el papel de ciertas variantes asociadas con enfermedades (GWAs, *genome-wide association studies*) en poblaciones ligadas históricamente a territorios específicos (Solovieff et al., 2011; Botigué et al., 2013).

Estudios desarrollados en Europa han puesto de manifiesto la apreciable subestructuración de su población, encontrándose significativas diferencias genéticas entre el norte y el sur del Continente (Seldin et al., 2006; Lao et al., 2008; Novembre et al., 2008; Nothnagel et al., 2010). Esta estructura muestra que, a nivel continental, las poblaciones europeas han experimentado diversas historias demográficas y han preservado, también, actitudes culturales distintivas con respecto a los patrones de cruzamientos entre parientes (fenómeno consanguíneo).

El extremo sudoccidental del continente europeo, la Península Ibérica, de nuevo destaca como un área clave respecto a los resultados obtenidos mediante esa nueva metodología experimental. Auton et al. (2009), utilizando un genotipado masivo de marcadores bialélicos, detectaron un gradiente norte-sur en el número de haplotipos diferentes. En concreto, Iberia es la región europea que más destaca en niveles de diversidad. Los autores aportan dos explicaciones para este resultado. En primer lugar, la elevada diversidad del sudoeste europeo podría ser un reflejo de una contribución sustancial procedente de África. El papel de las migraciones desde el norte de África hacia Iberia como fuente de diversidad ha sido explorado por Botigué et al., (2013). Por otro lado, se apunta a que la contracción de las poblaciones humanas en el sur de Europa durante el LGM sería otro escenario fundamental para explicar los patrones genéticos actuales a nivel continental.

En conclusión, el análisis de marcadores genéticos específicos –uniparentales, tales como los empleados en la presente Tesis Doctoral, o autosómicos– y, los barridos genómicos de SNPs a gran escala, revelan de manera paralela la posición central que ocupa la Península Ibérica respecto a sus patrones de diversidad genética en un contexto mediterráneo y europeo.

Capítulo VII

Conclusions

1. Mitochondrial DNA profiles of autochthonous Andalusians have shown internal genetic heterogeneity. This micro-geographic structure is based on distinctive traces from several continental areas. Western and eastern Andalusians would have experienced different population influences and maternal gene assimilation.
2. Mitochondrial diversity evaluated in southern Iberia (global profiles, African signatures and European background) have conducted to the same scenario, that is, genetic discontinuities within Andalusia region.
3. Andalusians register high genetic diversity values because of many gene flow episodes from the prehistory onwards. This dynamic distinguishes southern Iberia from other peninsular areas, in particular, the Cantabrian cornice, where the isolation, population fragmentation and genetic drift effects predominate. Thus, according to the findings drawn in our present research, Andalusians should be assumed as the most suitable candidate population to assess the role of Iberia in the peopling processes of the Mediterranean space.
4. The Mediterranean Sea has been a surmountable barrier to human migrations during both prehistoric and historic times. The geographic proximity between the African and European Mediterranean coasts, mainly at its western end (14 km), has allowed multiple coast-to-coast human movements. Therefore, the Strait of Gibraltar and neighboring eastward areas –the *Alboran* sea– must be considered as a permeable bridge between Europe and Africa.
5. Western Andalusians from Huelva province harbors the highest African maternal signatures of all peninsular Iberia. Comparatively, the African influence in eastern Andalusians from Granada is rather low. Interestingly, the African maternal influence is especially notable in the Atlantic façade of Iberia. This pattern is also observed for other haploid and autosomal African markers into the Atlantic Iberian gene pool.
6. Mitogenome sequencing revealed that several and complex migration episodes underlie beneath the present distribution and diversity of African lineages in Europe. In particular, we have proposed here for the first time, as far as we know, that the bulk of these African traces is due to ancient transcontinental links, and we provide arguments against a historical, simplistic view of the African human migrations towards Iberia, mainly related to the western Islamic expansion. The high sharing of lineages between the North of Africa and Iberia, results from a complex process continued through time, from the early Holocene arrivals to the most recent historic events.
7. Those deep input population movements from North Africa to Iberia allow that some specific radiations, as U6a and L1b, had enough time for local divergence within the Iberian Peninsula. Phylogenetic and phylogeographic analyses supports this ancient scenario. Moreover, we

have shown that Iberia has not acted only as a recipient of migrations, but also as a source population for the expansion of these maternal sub-branches back to North Africa and to other areas of Europe.

8. The high-resolution genealogical analysis of the main mitochondrial European clade (haplogroup H) revealed that eastern Andalusians from Granada province show a typical Iberian H structure, enriched indeed by other less common sub-lineages, which unveil genetic signals from other areas of the European continent and the Mediterranean space. Other signatures from central Europe (e.g. H7 and H11) and the eastern end of the Mediterranean (H13 and H20) are present in the maternal genome of Granada people. The higher diversification of haplogroup H profile in Granada Andalusians could be explained by its geographic position, which is more open to Mediterranean contacts.
9. The parallelism in clade H structure at the both ends of the Strait of Gibraltar together with the coalescence estimates for the main sub-lineages H1 and H3 show that these European mitochondrial sub-clades would have been introduced into northern Africa in post-glacial times by a southward migration from the Iberian Peninsula.
10. Both African haplogroups and clade H analyses have also shown deep rooting links between North of Africa and Iberia. A scenario of presumable bidirectional gene flow between both shores is supported not only by genetic evidences, but also by archaeological data regarding Paleolithic industries and Neolithization process. However, distinctive demographic histories would have affected the assimilation of genetic variants coming from the neighbor continent by the sink population. Thus, the stronger signature of European markers in the Maghreb, when compared to the African influence in Iberia, would be the consequence of the smaller demographic sizes of North Africa during the past millennia. In consequence, those southward and northward migrations would not be comparable in terms of their effect on the present population gene pool.
11. Our data further support the existence of a significant genetic structure along the Mediterranean Basin. The genetic composition of the Mediterranean people is shaped in a longitudinal axis (west to east). The mitochondrial evidence is a reflection of past evolutionary histories of European and Mediterranean human populations including major expansions and events, such as the Paleolithic colonization, the population retreat to southern refugia during last glacial times and the Neolithic transition.
12. The northern fringe of Africa is not genetically homogeneous. It has been explained through cultural, demographic and geographic factors, being the wide physical barrier of the Libyan Desert between the Nile Delta and Tunisia, the main element that have prevented a sustained human gene

flow there. The population from the western and eastern ends of North Africa are notably influenced by their Eurasiatic northern neighbors.

- 13.** The experimental procedures used in the present PhD Thesis have been adapted to the concrete aims and features of the mitochondrial lineages analyzed. A suitable strategy for assessing initially the mtDNA profiles of a population is to focus on control region sequences. The high variability encompassed in this fragment allowed us establishing a general overview of the mitochondrial diversity and a first classification of samples in haplogroups. Afterwards, we could choose between different methodologies to genotype coding SNPs. High-resolution analyses are required to refine our knowledge about the maternal heritage of populations and to further study the phylogeography and phylogeny of mtDNA lineages.
- 14.** The study of Y chromosome variation and autosomal markers on the same Andalusian sample set has not revealed a significant genetic heterogeneity within southern Iberia. However, we found similar patterns to that exposed here. The genetic pool of Andalusians is enriched by several continental influences that disclose the existence of relevant gene flow events with populations from the Mediterranean space.
- 15.** The results presented in this extensive research, based on the maternal legacy, together with other complementary findings drawn from the analysis of other genomic markers, reveal that our sample set could be considered as representative of the Andalusian territory. In this context, a suitable sampling process and sample selection, efforts in data compilation for comparative purposes, and powerful statistical tools are needed to accurately assess the timing of migratory episodes between North Africa and Europe, their actual impact on the European gene pool, the role and permeability of physical barriers, and the most plausible dispersal routes of human migrations.
- 16.** The region of the Mediterranean Basin is among the most demanding admixture zones in the globe. The chosen area for our research work, the western end of the Mediterranean, is complex and it will continue to be of strategic importance for addressing further approaches as genome-wide screenings and analyses of specific human diseases of great interest in terms of population genetics and human health.

Referencias

- Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogväli E-L, Kivisild T, Bandelt H-J, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A. 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, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshiar Kashani B, Battaglia V, Grugni V, Angerhofer N, Rogers MP, Herrera RJ, Woodward SR, Labuda D, Smith DG, Cybulski JS, Semino O, Malhi RS, Torroni A. 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc Natl Acad Sci U S A* 110:14308–14313.
- Acquaro E. 1999. Los fenicios en el Mediterráneo Central en la época de Tarteso. En: Alvar J, Blázquez JM, editores. *Los enigmas de Tarteso*. Madrid: Cátedra.
- Ajbar Machmuâ. 1867. *Crónica anónima del siglo XI*. Colección de obras arábicas de historia y geografía. Madrid: Real Academia de la Historia.
- Al-Abri A, Podgorná E, Rose JI, Pereira L, Mulligan CJ, Silva NM, Bayoumi R, Soares P, Cerný V. 2012. Pleistocene-Holocene boundary in Southern Arabia from the perspective of human mtDNA variation. *Am J Phys Anthr* 149:291–298.
- Alfonso-Sánchez MA, Cardoso S, Martínez-Bouzas C, Peña JA, Herrera RJ, Castro A, Fernández-Fernández I, de Pancorbo MM. 2008. Mitochondrial DNA Haplogroup Diversity in Basques : A reassessment Based on HVI and HVII Polymorphisms. *Am J Hum Biol* 20:154–164.
- Almagro-Basch M, del Amo M, Beltrán A. 1974. *Huelva: Prehistoria y Antigüedad*. Madrid: Editorial Nacional.
- Alonso S, Flores C, Cabrera V, Alonso A, Martín P, Albarrán C, Izagirre N, de la Rúa C, García O. 2005. The place of the Basques in the European Y-chromosome diversity landscape. *Eur J Hum Genet* 13:1293–1302.
- Álvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintáns B, Zarrabeitia MT, Cuscó I, Lareu M V, García O, Pérez-Jurado L, Carracedo A, Salas A. 2009. New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS One* 4:e5112.
- Álvarez L, Santos C, Ramos A, Pratdesaba R, Francalacci P, Aluja MP. 2010. Mitochondrial DNA patterns in the Iberian Northern plateau: population dynamics and substructure of the Zamora province. *Am J Phys Anthr* 142:531–539.
- Alves M, Camacho R, Soares P, Pereira L. 2012. From Networks to Trees. En: Rocha M, Luscombe N, Fernández-Riverola F, Corchado J, editores. *6th International Conference on Practical Applications of Computational Biology & Bioinformatics*. Berlin: Springer. p 129–136.
- Alzualde A, Izagirre N, Alonso S, Alonso A, Albarrán C, Azkarate A, de la Rúa C. 2006. Insights into the “isolation” of the Basques: mtDNA lineages from the historical site of Aldaieta (6th-7th centuries AD). *Am J Phys Anthr* 130:394–404.
- Amar S, Shamir A, Ovadia O, Blanaru M, Reshef A, Kremer I, Rietschel M, Schulze TG, Maier W, Belmaker RH, Ebstein RP, Agam G, Mishmar D. 2007. Mitochondrial DNA

- HV lineage increases the susceptibility to schizophrenia among Israeli Arabs. *Schizophr Res* 94:354–358.
- Ambrosio B, Dugoujon JM, Hernández C, de la Fuente D, González-Martín A, Fortes-Lima CA, Novelletto A, Rodríguez JN, Calderón R. 2010a. The Andalusian population from Huelva reveals a high diversification of Y-DNA paternal lineages from haplogroup E: Identifying human male movements within the Mediterranean space. *Ann Hum Biol* 37:86–107.
- Ambrosio B, Hernández C, Novelletto A, Dugoujon JM, Rodríguez JN, Cuesta P, Fortes-Lima C, Calderón R. 2010b. Searching the Peopling of the Iberian Peninsula from the Perspective of Two Andalusian Subpopulations: A Study Based on Y-chromosome Haplogroups J and E. *Coll Antropol* 34:1215–1228.
- Ambrosio B. 2011. Estructura de la población andaluza de Huelva basada en la variación de su DNA mitocondrial y del cromosoma Y. Aportes al conocimiento sobre la historia humana reciente dentro del espacio mediterráneo. Tesis Doctoral. Universidad Complutense de Madrid.
- Ambrosio B, Novelletto A, Hernández C, Dugoujon JM, Fortes-Lima C, Rodríguez JN, Calderón R. 2012. Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain). *Forensic Sci Int Genet* 6:e66–e71.
- Ammerman AJ, Cavalli-Sforza LL. 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton, New Jersey: Princeton University Press.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.
- Anderung C, Bouwman A, Persson P, Carretero JM, Ortega AI, Elburg R, Smith C, Arsuaga JL, Ellegren H, Götherström A. 2005. Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proc Natl Acad Sci U S A* 102:8431–8435.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
- Antonarakis S. 2010. Human Genome Sequence and Variation. En: Speicher M, Antonarakis S, Motulsky A, editores. *Vogel and Motulsky's Human Genetics. Problems and Approaches*. 4th ed. Berlin Heidelberg: Springer-Verlag.
- Antonarakis S, Cooper D. 2010. Human Gene Mutation: Mechanisms and Consequences. En: Speicher M, Antonarakis S, Motulsky A, editores. *Vogel and Motulsky's Human Genetics. Problems and Approaches*. 4th ed. Berlin Heidelberg: Springer-Verlag.
- Arenas M, François O, Currat M, Ray N, Excoffier L. 2012. Influence of Admixture and Paleolithic Range Contractions on Current European Diversity Gradients. *Mol Biol Evol*:1–5.
- Arié R. 1983. III. España Musulmana (siglos VIII-XV). En: Tuñón de Lara M, editor. *Historia de España*. Barcelona: Labor.

- Arredi B, Poloni ES, Tyler-Smith C. 2006. The peopling of Europe. En: Crawford MH, editor. *Anthropological Genetics. Theory, Methods and Applications*. Cambridge University Press. p 380–408.
- Atkinson QD, Gray RD, Drummond AJ. 2009. Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc Biol Sci* 276:367–373.
- Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, Reynolds A, Indap A, Wright MH, Degenhardt JD, Gutenkunst RN, King KS, Nelson MR, Bustamante CD. 2009. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res* 19:795–803.
- Avise JC. 2000. *Phylogeography: The History and Formation of Species*. Cambridge, Massachusetts: Harvard University Press.
- Badro DA, Douaihy B, Haber M, Youhanna SC, Salloum A, Ghassibe-Sabbagh M, Johnsrud B, Khazen G, Matisoo-Smith E, Soria-Hernanz DF, Wells RS, Tyler-Smith C, Platt DE, Zalloua PA. 2013. Y-Chromosome and mtDNA Genetics Reveal Significant Contrasts in Affinities of Modern Middle Eastern Populations with European and African Populations. *PLoS One* 8:e54616.
- Balaresque P, Bowden GR, Adams SM, Leung H-Y, King TE, Rosser ZH, Goodwin J, Moisan J-P, Richard C, Millward A, Demaine AG, Barbujani G, Previderè C, Wilson IJ, Tyler-Smith C, Jobling MA. 2010. A predominantly neolithic origin for European paternal lineages. *PLoS Biol* 8:e1000285.
- Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol* 13:729–744.
- Bandelt H-J, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
- Bandelt H-J, Kloss-Brandstätter A, Richards MB, Yao Y-G, Logan I. 2014. The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies. *J Hum Genet* 59:66–77.
- Barbieri C, Vicente M, Rocha J, Mpoloka SW, Stoneking M, Pakendorf B. 2013. Ancient substructure in early mtDNA lineages of southern Africa. *Am J Hum Genet* 92:285–292.
- Barbujani G, Bertorelle G. 2001. Genetics and the population history of Europe. *Proc Natl Acad Sci U S A* 98:22–25.
- Barton R, Bouzouggar A, Hogue J, Lee S, Collcutt S, Ditchfield P. 2013. Origins of the Iberomaurusian in NW Africa: New AMS radiocarbon dating of the Middle and Later Stone Age deposits at Taforalt Cave, Morocco. *J Hum Evol* 65:266–281.
- Batzler MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ. 1994. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci U S A* 91:12288–12292.
- Bauchet M, McEvoy B, Pearson LN, Quillen EE, Sarkisian T, Hovhannesian K, Deka R, Bradley DG, Shriver MD. 2007. Measuring European population stratification with microarray genotype data. *Am J Hum Genet* 80:948–56.

- Behar DM, Hammer MF, Garrigan D, Villems R, Bonne-Tamir B, Richards M, Gurwitz D, Rosengarten D, Kaplan M, Della Pergola S, Quintana-Murci L, Skorecki K. 2004. MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. *Eur J Hum Genet* 12:355–364.
- Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S. 2008. The Dawn of Human Matrilineal Diversity. *Am J Hum Genet* 82:1130–1140.
- Behar DM, Harmant C, Manry J, van Oven M, Haak W, Martinez-Cruz B, Salaberria J, Oyharçabal B, Bauduer F, Comas D, Quintana-Murci L. 2012a. The Basque Paradigm: Genetic Evidence of a Maternal Continuity in the Franco-Cantabrian Region since Pre-Neolithic Times. *Am J Hum Genet* 90:486–493.
- Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva NM, Kivisild T, Torroni A, Villems R. 2012b. A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root. *Am J Hum Genet* 90:675–684.
- Bekada A, Fregel R, Cabrera VM, Larruga JM, Pestano J, Benhamamouch S, González AM. 2013. Introducing the Algerian mitochondrial DNA and Y-chromosome profiles into the North African landscape. *PLoS One* 8:e56775.
- Bellwood P, Oxenham M. 2008. The Expansions of Farming Societies and the Role of the Neolithic Demographic Transition. En: Bocquet-Appel J-P, Bar-Yosef O, editores. *The Neolithic Demographic Transition and its Consequences*. Springer. p 13–34.
- Bergstrom CT, Pritchard J. 1998. Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. *Genetics* 149:2135–2146.
- Bernabeu J. 1996. Indigenismo y migracionismo. Aspectos de la neolitización en la fachada oriental de la Península Ibérica. *Trab Prehist* 53:37–54.
- Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D. 1995. Human Mitochondrial DNA Variation and the Origin of Basques. *Ann Hum Genet* 59:63–81.
- Biraben J-N. 1979. Essai sur l'évolution du nombre des hommes. *Population (Paris)* 34:13–25.
- Biraben J-N, Masset C, Thillaud P-L. 1997. El poblamiento prehistórico de Europa. En: Bardet J-P, Dupâquier J, editores. *Historia de las poblaciones de Europa. Volumen I: De los orígenes a las premisas de la revolución demográfica*. Madrid: Editorial Síntesis.
- Bittles A. 2010. Consanguinity, Genetic Drift, and Genetic Diseases in Populations with Reduced Numbers of Founders. En: Speicher M, Antonarakis S, Motulsky A, editores. *Vogel and Motulsky's Human Genetics. Problems and Approaches*. 4th ed. Berlin Heidelberg: Springer-Verlag.
- Bocquet-Appel J-P, Demars P-Y, Noiret L, Dobrowsky D. 2005. Estimates of Upper Palaeolithic meta-population size in Europe from archaeological data. *J Archaeol Sci* 32:1656–1668.
- Bosch E, Calafell F, Pérez-Lezaun A, Comas D, Mateu E, Bertranpetit J. 1997. Population history of north Africa: evidence from classical genetic markers. *Hum Biol* 69:295–311.

- Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J. 2001. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet* 68:1019–1029.
- Botigué LR, Henn BM, Gravel S, Maples BK, Gignoux CR, Corona E, Atzmon G, Burns E, Ostrer H, Flores C, Bertranpetit J, Comas D, Bustamante CD. 2013. Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc Natl Acad Sci U S A* 110:11791–6.
- Brakez Z, Bosch E, Izaabel H, Akhayat O, Comas D, Bertranpetit J, Calafell F. 2001. Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol* 28:295–307.
- Brandstätter A, Parsons TJ, Parson W. 2003. Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Leg Med* 117:291–298.
- Brandstätter A, Zimmermann B, Wagner J, Göbel T, Röck AW, Salas A, Carracedo A, Parson W. 2008. Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East. *BMC Evol Biol* 8:191.
- Brandt G, Szécsényi-Nagy A, Roth C, Alt KW, Haak W. 2015. Human paleogenetics of Europe – The known knowns and the known unknowns. *J Hum Evol* 79:73–92.
- Braudel F. 1966. *La Méditerranée et le monde méditerranéen à l'époque de Philippe II*. Paris: Armand Colin.
- Braudel F. 1987. *El Mediterráneo*. Madrid: Espasa Calpe.
- Brisighelli F, Álvarez-Iglesias V, Fondevila M, Blanco-Verea A, Carracedo A, Pascali VL, Capelli C, Salas A. 2012. Uniparental markers of contemporary Italian population reveals details on its pre-Roman heritage. *PLoS One* 7:e50794.
- Broodbank C. 2006. The Origins and Early Development of Mediterranean Maritime Activity. *J Mediterr Archaeol* 19:199–230.
- Brooks N, Chiapello I, Lernia SD, Drake N, Legrand M, Moulin C, Prospero J. 2005. The climate-environment-society nexus in the Sahara from prehistoric times to the present day. *J North African Stud* 10:253–292.
- Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, Richards SM, Sarkissian CD, Ganslmeier R, Friederich S, Dresely V, van Oven M, Kenyon R, Van der Hoek MB, Korfach J, Luong K, Ho SYW, Quintana-Murci L, Behar DM, Meller H, Alt KW, Cooper A, Adhikarla S, Ganesh Prasad AK, Pitchappan R, Varatharajan Santhakumari A, Balanovska E, Balanovsky O, Bertranpetit J, Comas D, Martínez-Cruz B, Melé M, Clarke AC, Matisoo-Smith EA, Dulik MC, Gaieski JB, Owings AC, Schurr TG, Vilar MG, Hobbs A, Soodyall H, Javed A, Parida L, Platt DE, Royyuru AK, Jin L, Li S, Kaplan ME, Merchant NC, John Mitchell R, Renfrew C, Lacerda DR, Santos FR, Soria Hernanz DF, Spencer Wells R, Swamikrishnan P, Tyler-Smith C, Paulo Vieira P, Ziegler JS. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat Commun* 4:1764.
- Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* 76:1967–1971.

- Budowle B, Allard MW, Wilson MR, Chakraborty R. 2003. Forensics and mitochondrial DNA: applications, debates, and foundations. *Annu Rev Genomics Hum Genet* 4:119–141.
- Cabré AM, Domingo A, Menacho T. 2002. Demografía y crecimiento de la población española durante el siglo XX. *Mediterráneo económico* 1:121–138.
- Calderón R, Carrion M, Perez-Miranda A, Peña JA, Dugoujon JM, Crouau-Roy B. 2003. Allele variation of DYS19 and Y-Alu insertion (YAP) polymorphisms in Basques: an insight into the peopling of Europe and the Mediterranean region. *Hum Biol* 75:117–127.
- Calderón R, Ambrosio B, Guitard E, González-Martín A, Aresti U, Dugoujon JM. 2006. Genetic position of Andalusians from Huelva in relation to other European and North African populations: a study based on GM and KM allotypes. *Hum Biol* 78:663–679.
- Calderón R, Lodeiro R, Varela TA, Fariña J, Ambrosio B, Guitard E, González-Martín A, Dugoujon JM. 2007. GM and KM immunoglobulin allotypes in the Galician population: new insights into the peopling of the Iberian Peninsula. *BMC Genet* 8:37.
- Calderón R, Hernández CL, Cuesta P, Dugoujon J-M. 2015. Surnames and Y-chromosomal markers reveal low relationships in Southern Spain. *PLoS One* 10:e0123098.
- Camps G. 1974. *Les Civilisations Préhistoriques De L’Afrique Du Nord Et Du Sahara*. Paris: Doin.
- Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.
- Cardoso S, Zarrabeitia MT, Valverde L, Odriozola A, Alfonso-Sánchez MÁ, de Pancorbo MM. 2010. Variability of the entire mitochondrial DNA control region in a human isolate from the Pas Valley (northern Spain). *J Forensic Sci* 55:1196–1201.
- Cardoso S, Valverde L, Alfonso-Sánchez MA, Palencia-Madrid L, Elcoroaristizabal X, Algorta J, Catarino S, Arteta D, Herrera RJ, Zarrabeitia MT, Peña JA, de Pancorbo MM. 2013. The Expanded mtDNA Phylogeny of the Franco-Cantabrian Region Upholds the Pre-Neolithic Genetic Substrate of Basques. *PLoS One* 8:e67835.
- Caro Baroja J. 1991. *Los moriscos del Reino de Granada: Ensayo de historia social*. 2nd ed. Madrid: Istmo.
- Carrieri G, Bonafè M, De Luca M, Rose G, Varcasia O, Bruni A, Maletta R, Nacmias B, Sorbi S, Corsonello F, Feraco E, Andreev KF, Yashin AI, Franceschi C, De Benedictis G. 2001. Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer’s disease. *Hum Genet* 108:194–198.
- Carrilero M. 2000. 1. La prehistoria, Tartessos y otros pueblos de la antigüedad. En: Cano G, editor. *Conocer Andalucía*. Gran Enciclopedia Andaluza del siglo XXI. Sevilla: Ediciones Tartessos.
- Casas MJ, Hagelberg E, Fregel R, Larruga JM, González AM. 2006. Human Mitochondrial DNA Diversity in an Archaeological Site in al-Andalus: Genetic Impact of Migrations from North Africa in Medieval Spain. *Am J Phys Anthr* 131:539–551.

- Castañeda IS, Muiltza S, Schefuss E, Lopes dos Santos RA, Sinninghe Damsté JS, Schouten S. 2009. Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. *Proc Natl Acad Sci U S A* 106:20159–63.
- Castro JA, Picornell A, Ramon M. 1998. Mitochondrial DNA: a tool for populational genetics studies. *Int Microbiol* 1:327–332.
- Cerezo M, Achilli A, Olivieri A, Perego UA, Gómez-Carballa A, Brisighelli F, Lancioni H, Woodward SR, López-Soto M, Carracedo A, Capelli C, Torroni A, Salas A. 2012. Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res* 22:821–826.
- Chalmeta P. 1989. Al-Andalus: musulmanes y cristianos: siglos VIII-XII. En: Domínguez Ortiz A, editor. *Historia de España*. 1st ed. Barcelona: Planeta.
- Chapman R. 1989. *Emerging complexity: the later prehistory of south-east Spain, Iberia and the west Mediterranean*. Cambridge: Cambridge University Press.
- Chapman R. 2008. Producing Inequalities: Regional Sequences in Later Prehistoric Southern Spain. *J World Prehist* 21:195–260.
- Cherni L, Loueslati BY, Pereira L, Ennafaâ H, Amorim A, El Gaaied ABA. 2005. Female gene pools of Berber and Arab neighboring communities in central Tunisia: microstructure of mtDNA variation in North Africa. *Hum Biol* 77:61–70.
- Cherni L, Fernandes V, Pereira JB, Costa MD, Goios A, Frigi S, Yacoubi-Loueslati B, Ben Amor M, Slama A, Amorim A, El Gaaied ABA, Pereira L. 2009. Post-last glacial maximum expansion from Iberia to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia. *Am J Phys Anthr* 139:253–260.
- Chinnery PF. 2006. Mitochondrial DNA in Homo Sapiens. En: Bandelt H-J, Macaulay V, Richards M, editores. *Human Mitochondrial DNA and the Evolution of Homo sapiens*. Springer-Verlag.
- Chinnery PF, Mowbray C, Patel SK, Elson JL, Sampson M, Hitman GA, McCarthy MI, Hattersley AT, Walker M. 2007. Mitochondrial DNA haplogroups and type 2 diabetes: a study of 897 cases and 1010 controls. *J Med Genet* 44:e80.
- Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114:127–48.
- Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. 2009. The Last Glacial Maximum. *Science* 325:710–714.
- Cohen JE, Small C. 1998. Hypsographic demography: The distribution of human population by altitude. *Proc Natl Acad Sci U S A* 95:14009–14014.
- Colominas L, Edwards CJ, Beja-Pereira A, Vigne J-D, Silva RM, Castanyer P, Tremoleda J, Saña Seguí M, Pérez-Ripoll M, Goyache F, Howe CJ, Barker G, Bower MA. 2015. Detecting the T1 cattle haplogroup in the Iberian Peninsula from Neolithic to Medieval times: new clues to continuous cattle migration through time. *J Archaeol Sci* 59:110–117.
- Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A. 2000. Alu insertion polymorphisms in NW Africa and the

- Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* 107:312–319.
- Côrte-Real HBSM, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* 60:331–350.
- Cortés Sánchez M, Jiménez Espejo FJ, Simón Vallejo MD, Gibaja Bao JF, Carvalho AF, Martínez-Ruiz F, Gamiz MR, Flores J-A, Paytan A, López Sáez JA, Peña-Chocarro L, Carrión JS, Morales Muñoz A, Roselló Izquierdo E, Riquelme Cantal JA, Dean RM, Salgueiro E, Martínez Sánchez RM, De la Rubia de Gracia JJ, Lozano Francisco MC, Vera Peláez JL, Rodríguez LL, Bicho NF. 2012. The Mesolithic–Neolithic transition in southern Iberia. *Quat Res* 77:221–234.
- Corvisier N, Hinard F, Salmon P. 1997. La Europa grecorromana. En: Bardet J-P, Dupâquier J, editores. *Historia de las poblaciones de Europa. Volumen I: De los orígenes a las premisas de la revolución demográfica*. Madrid: Editorial Síntesis.
- Costa MD, Cherni L, Fernandes V, Freitas F, Ammar El Gaaied AB, Pereira L. 2009. Data from complete mtDNA sequencing of Tunisian centenarians: testing haplogroup association and the “golden mean” to longevity. *Mech Ageing Dev* 130:222–226.
- Costa M, Pereira J, Pala M, Fernandes V, Olivieri A, Achilli A, Perego U, Rychkov S, Naumova O, Hatina J, Woodward S, Eng K, Macaulay V, Carr M, Soares P, Pereira L, Richards M. 2013. A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nat Commun* 4:2543.
- Coudray C, Guitard E, Kandil M, Harich N, Melhaoui M, Baali A, Sevin A, Moral P, Dugoujon JM. 2006. Study of GM immunoglobulin allotypic system in Berbers and Arabs from Morocco. *Am J Hum Biol* 18:23–34.
- Coudray C, Calderón R, Guitard E, Ambrosio B, González-Martín A, Dugoujon J-M. 2007. Allele frequencies of 15 tetrameric short tandem repeats (STRs) in Andalusians from Huelva (Spain). *Forensic Sci Int* 168:e21–4.
- Coudray C, Olivieri A, Achilli A, Pala M, Melhaoui M, Cherkaoui M, El-Chennawi F, Kossmann M, Torroni A, Dugoujon JM. 2009. The complex and diversified mitochondrial gene pool of Berber populations. *Ann Hum Genet* 73:196–214.
- Coulthard TJ, Ramirez JA, Barton N, Rogerson M, Brücher T. 2013. Were rivers flowing across the Sahara during the last interglacial? Implications for human migration through Africa. *PLoS One* 8:e74834.
- Crawford M. 1973. The use of genetic markers of the blood in the study of the evolution of human populations. En: Crawford M, Workman P, editores. *Methods and Theories of Anthropological Genetics*. Albuquerque: University of New Mexico Press. p 19–38.
- Crimi M, Rigolio R. 2008. The mitochondrial genome, a growing interest inside an organelle. *Int J Nanomedicine* 3:51–57.
- Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* 74:1014–22.

- Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon J-M, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradisnik B, Assum G, Brdicka R, Kozlov AI, Efremov GD, Coppa A, Novelletto A, Scozzari R. 2007. Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol Biol Evol* 24:1300–11.
- Cruz Berrocal M. 2012. The Early Neolithic in the Iberian Peninsula and the Western Mediterranean: A Review of the Evidence on Migration. *J World Prehist* 25:123–156.
- Currat M, Poloni ES, Sanchez-Mazas A. 2010. Human genetic differentiation across the Strait of Gibraltar. *BMC Evol Biol* 10:237.
- Currat M. 2012. Consequences of population expansions on European genetic diversity. En: Kaiser E, Burger J, Schier W, editores. *Population Dynamics in Prehistory and Early History: New Approaches Using Stable Isotopes and Genetics*. Berlin: Walter de Gruyter & Co. p 3–16.
- Delgado A. 2013. Chapter 15. Households, merchants and feasting. Socioeconomic dynamics and commoners' agency in the emergence of the Tartessian World (Eleventh to eight centuries B.C.). En: Berrocal M, editor. *The Prehistory of Iberia: debating early social stratification and the state*. New York; London: Routledge.
- Demand N. 2011. Seafaring in the Mesolithic Mediterranean. En: Demand N, editor. *The Mediterranean Context of Early Greek History*. Wiley-Blackwell.
- Demangeon A. 1907. *Dictionnaire-manuel illustré de géographie*. Paris: Armand Colin.
- Dermitzakis E. 2010. Population Genetic Principles and Human Populations. En: Speicher M, Antonarakis S, Motulsky A, editores. *Vogel and Motulsky's Human Genetics. Problems and Approaches*. 4th ed. Berlin Heidelberg: Springer-Verlag.
- Domínguez C. 2010. Historiography and the geo-literary imaginary. The Iberian Peninsula: Between Lebensraum and espace vécu. En: Cabo F, Abuín A, Domínguez C, editores. *A Comparative History of Literatures in the Iberian Peninsula. Volume I*. John Benjamins Publishing Company.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22:1185–92.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
- Dupâquier J. 1997. Des origines aux prémices de la révolution démographique. En: Bardet JP, Dupâquier J, editores. *Histoire des Populations de l'Europe*. Paris: Librairie Arthème Fayard. p 26–38.
- Elson JL, Turnbull DM, 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, Ho SYW. 2008. A Bayesian Evaluation of Human Mitochondrial Substitution Rates. *Am J Hum Genet* 82:895–902.

- Ennafaa H, Cabrera VM, Abu-Amero KK, González AM, Amor MB, Bouhaha R, Dzimiri N, Elgaaied AB, Larruga JM. 2009. Mitochondrial DNA haplogroup H structure in North Africa. *BMC Genet* 10:8.
- Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–91.
- Excoffier L, Yang Z. 1999. Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. *Mol Biol Evol* 16:1357–1368.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–7.
- Fadhlaoui-Zid K, Plaza S, Calafell F, Ben Amor M, Comas D, El Gaaied ABA. 2004. Mitochondrial DNA heterogeneity in Tunisian Berbers. *Ann Hum Genet* 68:222–33.
- Fadhlaoui-Zid K, Rodríguez-Botigué L, Naoui N, Benammar-Elgaaied A, Calafell F, Comas D. 2011. Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. *Am J Phys Anthr* 145:107–17.
- Fadhlaoui-Zid K, Haber M, Martínez-Cruz B, Zalloua P, Ben Ammar El Gaaied A, Comas D. 2013. Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. *PLoS One* 8:e80293.
- Falchi A, Giovannoni L, Calo CM, Piras IS, Moral P, Paoli G, Vona G, Varesi L. 2006. Genetic history of some western Mediterranean human isolates through mtDNA HVR1 polymorphisms. *J Hum Genet* 51:9–14.
- Falconer DS, Mackay TFC. 2006. *Introducción a la genética cuantitativa*. Zaragoza: Acribia.
- Feder J, Blech I, Ovadia O, Amar S, Wainstein J, Raz I, Dadon S, Arking DE, Glaser B, Mishmar D. 2008. Differences in mtDNA haplogroup distribution among 3 Jewish populations alter susceptibility to T2DM complications. *BMC Genomics* 9:198.
- Fernandes V, Alshamali F, Alves M, Costa MD, Pereira JB, Silva NM, Cherni L, Harich N, Cerny V, Soares P, Richards MB, Pereira L. 2012. The Arabian Cradle: Mitochondrial Relicts of the First Steps along the Southern Route out of Africa. *Am J Hum Genet* 90:347–355.
- Fernandes V, Triska P, Pereira JB, Alshamali F, Rito T, Machado A, Fajkošová Z, Cavadas B, Černý V, Soares P, Richards MB, Pereira L. 2015. Genetic stratigraphy of key demographic events in Arabia. *PLoS One* 10:e0118625.
- Fernández J, Gómez M. 2009. Climate change and population dynamics during the late Mesolithic and the Neolithic transition in Iberia. *Doc Praehist* 36:67.
- Fernandez-Jurado J, García-Sanz C, Rufete P. 1997. *De Tartessos a Onuba*. Huelva: Diputación de Huelva.
- Fernández V. 2003. Capítulo 10. La ciudad histórica andaluza: su origen, conformación y características actuales. En: López Ontiveros A, editor. *Geografía de Andalucía*. Barcelona: Ariel.

- Ferre E. 2000. 2. El relieve. Montañas, valles y costas. En: Cano G, editor. Conocer Andalucía. Gran Enciclopedia Andaluza del siglo XXI. Sevilla: Ediciones Tartessos.
- Forster P, Harding R, Torroni A, Bandelt H-J. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59:935–45.
- Fortes-Lima C, Dugoujon J-M, Hernández CL, Reales G, Calderón R. 2014. Immunoglobulin genes in Andalusia (Spain). Genetic diversity in the Mediterranean space. *C R Biol* 337:646–656.
- Fu Q, Rudan P, Pääbo S, Krause J. 2012. Complete Mitochondrial Genomes Reveal Neolithic Expansion into Europe. *PLoS One* 7:e32473.
- Fu Y-X. 1997. Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics* 147:915–25.
- Gamble C, Davies W, Pettitt P, Richards M. 2004. Climate change and evolving human diversity in Europe during the last glacial. *Philos Trans R Soc L B Biol Sci* 359:243–53; discussion 253–4.
- Gamble C, Davies W, Pettitt P, Hazelwood L, Richards M. 2006. The Late Glacial ancestry of Europeans: Combining genetic and archaeological evidence. *Doc Praehist* 33:1–10.
- García de Cortázar F. 2012. Atlas de historia de España. 7th ed. Barcelona: Planeta.
- García Iglesias L. 1978. Los judíos en la España Antigua. Madrid: Ediciones Cristiandad.
- García O, Fregel R, Larruga JM, Álvarez V, Yurrebaso I, Cabrera VM, González AM. 2011. Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge. *Heredity (Edinb)* 106:37–45.
- Gaweda-Walerych K, Maruszak A, Safranow K, Bialecka M, Klodowska-Duda G, Czyzewski K, Slawek J, Rudzinska M, Styczynska M, Opala G, Drozdziak M, Canter JA, Barcikowska M, Zekanowski C. 2008. Mitochondrial DNA haplogroups and subhaplogroups are associated with Parkinson's disease risk in a Polish PD cohort. *J Neural Transm* 115:1521–1526.
- Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, Papola F, Ciavarella G, Cucci F, Di Stasi L, Gavrilu L, Kerimova MG, Kovatchev D, Kozlov AI, Loutradis A, Mandarino V, Mammi' C, Michalodimitrakis EN, Paoli G, Pappa KI, Pedicini G, Terrenato L, Tofanelli S, Malaspina P, Novelletto A. 2004. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet* 115:357–71.
- Giles RE, Blanc H, Cann HM, Wallace DC. 1980. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 77:6715–6719.
- Gilman A, Thornes J, Wise S. 1985. Land-use and prehistory in South-East Spain. (Unwin A&, editor.). London.
- Gómez-Carballa A, Olivieri A, Behar DM, Achilli A, Torroni A, Salas A. 2012. Genetic Continuity in the Franco-Cantabrian Region: New Clues from Autochthonous Mitogenomes. *PLoS One* 7:e32851.

- González AM, Brehm A, Pérez JA, Maca-Meyer N, Flores C, Cabrera VM. 2003. Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am J Phys Anthr* 120:391–404.
- González AM, García O, Larruga JM, Cabrera VM. 2006. The mitochondrial lineage U8a reveals a Paleolithic settlement in the Basque country. *BMC Genomics* 7:124.
- González AM, Larruga JM, Abu-Amero KK, Shi Y, Pestano J, Cabrera VM. 2007. Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics* 8:223.
- Gray MW, Burger G, Lang BF. 1999. Mitochondrial Evolution. *Science* 283:1476–1481.
- von Grunebaum G. 1984. El Islam. II. Desde la caída de Constantinopla hasta nuestros días. *Historia Universal del Siglo XXI*. Vol 15. Madrid: Siglo XXI de España Editores SA.
- Guichard P. 1976. Al-Andalus: Estructura antropológica de una sociedad islámica en Occidente. Barcelona: Barral Editores.
- Guilaine J. 2003. De la vague à la tombe. La conquête néolithique de la Méditerranée. Paris: Seuil.
- Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95 – 98.
- Handley LJJ, Manica A, Goudet J, Balloux F. 2007. Going the distance: human population genetics in a clinal world. *Trends Genet* 23:432–439.
- Harich N, Costa MD, Fernandes V, Kandil M, Pereira JB, Silva NM, Pereira L. 2010. The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. *BMC Evol Biol* 10:138.
- Harman D. 1992. Role of free radicals in aging and disease. *Ann N Y Acad Sci* 673:126–141.
- Harpending H. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 66:591–600.
- Hartl D, Clark A. 1989. *Principles of Population Genetics*. 2nd ed. Sunderland, Massachusetts: Sinauer Associates.
- Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, 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.
- Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. 2014. A genetic atlas of human admixture history. *Science* 343:747–51.
- Henn BM, Gignoux CR, Feldman MW, Mountain JL. 2009. Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol Biol Evol* 26:217–230.
- Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlouzi-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, Bustamante CD, Comas D. 2012. Genomic

- ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet* 8:e1002397.
- Herrnstadt C, Howell N. 2004. An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders. *Mitochondrion* 4:791–798.
- Heyer E, Chaix R, Pavard S, Austerlitz F. 2012. Sex-specific demographic behaviours that shape human genomic variation. *Mol Ecol* 21:597–612.
- Holt IJ, Harding AE, Morgan-Hughes JA. 1988. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* 331:717–719.
- Horai S, Hayasaka K. 1990. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am J Hum Genet* 46:828–842.
- Horden P, Purcell N. 2000. *The Corrupting Sea: A Study of Mediterranean History*. Oxford: Wiley-Blackwell.
- Hublín J-J. 2015. The modern human colonization of western Eurasia: when and where? *Quat Sci Rev* 118:194-210.
- Hudson G, Carelli V, Spruijt L, Gerards M, Mowbray C, Achilli A, Pyle A, Elson J, Howell N, La Morgia C, Valentino ML, Huoponen K, Savontaus M-L, Nikoskelainen E, Sadun AA, Salomao SR, Belfort R, Griffiths P, Man PYW, de Coo RFM, Horvath R, Zeviani M, Smeets HJT, Torroni A, Chinnery PF. 2007. Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background. *Am J Hum Genet* 81:228–233.
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713.
- Irish JD. 2000. The Iberomaurusian enigma: north African progenitor or dead end? *J Hum Evol* 39:393–410.
- Itan Y, Jones BL, Ingram CJE, Swallow DM, Thomas MG. 2010. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol Biol* 10:36.
- Jeong C, Di Rienzo A. 2014. Adaptations to local environments in modern human populations. *Curr Opin Genet Dev* 29C:1–8.
- Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–612.
- Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C. 2013. *Human Evolutionary Genetics*. New York: Garland Science.
- Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL. 1983. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *J Mol Evol* 19:255–271.
- Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 66:979–988.
- Karachanak S, Carossa V, Nesheva D, Olivieri A, Pala M, Hooshiar Kashani B, Grugni V, Battaglia V, Achilli A, Yordanov Y, Galabov AS, Semino O, Toncheva D, Torroni A.

2012. Bulgarians vs the other European populations: a mitochondrial DNA perspective. *Int J Leg Med* 126:497–503.
- Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R. 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9:1331–4.
- Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ. 2006a. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–387.
- Kivisild T, Metspalu M, Bandelt H-J, Richards M, Villems R. 2006b. The World mtDNA Phylogeny. En: Bandelt H-J, Macaulay V, Richards M, editores. *Human Mitochondrial DNA and the Evolution of Homo sapiens*. Springer-Verlag.
- Kivisild T. 2015. Maternal ancestry and population history from whole mitochondrial genomes. *Investig Genet* 6:3.
- Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, Kronenberg F. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* 32:25–32.
- Knapp AB. 2010. Cyprus's Earliest Prehistory: Seafarers, Foragers and Settlers. *J World Prehist* 23:79–120.
- Kraytsberg Y, Schwartz M, Brown TA, Ebraldise K, Kunz WS, Clayton DA, Vissing J, Khrapko K. 2004. Recombination of human mitochondrial DNA. *Science* 304:981.
- Ladero Quesada M. 1988. De la crisis medieval al Renacimiento (siglos XIV-XV). En: Domínguez Ortiz A, editor. *Historia de España*. 1st ed. Barcelona: Planeta.
- Ladero-Quesada M. 1993. Granada después de la conquista: repobladores y mudéjares. Granada: Diputación de Granada.
- Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balasckova M, Bertranpetit J, Bindoff LA, Comas D, Holmlund G, Kouvatsi A, Macek M, Mollet I, Parson W, Palo J, Ploski R, Sajantila A, Tagliabracci A, Gether U, Werge T, Rivadeneira F, Hofman A, Uitterlinden AG, Gieger C, Wichmann H-E, Rütger A, Schreiber S, Becker C, Nürnberg P, Nelson MR, Krawczak M, Kayser M. 2008. Correlation between genetic and geographic structure in Europe. *Curr Biol* 18:1241–8.
- Larruga JM, Díez F, Pinto FM, Flores C, González AM. 2001. Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain. *Eur J Hum Genet* 9:708–716.
- Lebart L, Morineau A, Warwick K. 1984. Multivariate descriptive statistical analysis: correspondence analysis and related techniques for large matrices. En: *Series in Probability and Mathematical Statistics*. New York: John Wiley & Sons Inc.
- Legendre P, Fortin MJ. 1989. Spatial pattern and ecological analysis. *Vegetatio* 80:107–138.

- Legros F, Malka F, Frachon P, Lombès A, Rojo M. 2004. Organization and dynamics of human mitochondrial DNA. *J Cell Sci* 117:2653–2662.
- Liang Y-G, Balbale AK, Devereux A, Gómez-Rivas C. 2013. Unity and Disunity across the Strait of Gibraltar. *Mediev Encount* 19:1–40.
- Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. 2012. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One* 7:e31743.
- Linstädter J, Medved I, Solich M, Weniger G-C. 2012. Neolithisation process within the Alboran territory: Models and possible African impact. *Quatern Int* 274:219–232.
- Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, Schröder R, Stoneking M. 2014. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investig Genet* 5:13.
- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. 2003. *Biología celular y molecular*. 4th ed. Editorial Médica Panamericana.
- Loogväli E-L, Roostalu U, Malyarchuk BA, Derenko M V, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk H-V, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barac L, Peričić M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska E V, Grechanina E, Richard C, Moisan J-P, Chaventré A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustres M, Gölge M, Mikerezi I, Usanga E, Villems R. 2004. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* 21:2012–21.
- López Ontiveros A. 2003a. Capítulo 2. El territorio andaluz: su formación, delimitación e interpretación. En: López Ontiveros A, editor. *Geografía de Andalucía*. Barcelona: Ariel.
- López Ontiveros A. 2003b. Capítulo 21. Unidad y variedad de Andalucía. En: López Ontiveros A, editor. *Geografía de Andalucía*. Barcelona: Ariel.
- Lubell D. 2001. Late Pleistocene Early Holocene Maghreb. En: Peregrine P, Ember M, editores. *Encyclopedia of Prehistory, Volume 1: Africa*. New York: Kluwer Academic/Plenum Publishers. p 129–149.
- Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM. 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13.
- Maca-Meyer N, Sánchez-Velasco P, Flores C, Larruga JM, González AM, Oterino A, Leyva-Cobián F. 2003a. Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Ann Hum Genet* 67:329–339.
- Maca-Meyer N, González AM, Pestano J, Flores C, Larruga JM, Cabrera VM. 2003b. Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genet* 4:15.
- Macaulay V, Richards M, Hickey E, Vega 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.
- Macaulay V, Richards M. 2013. *Mitochondrial Genome Sequences and Their Phylogeographic Interpretation*. eLS:1–8.

- Mackay A. 1977. *La España de la Edad Media: desde la frontera hasta el imperio 1000–1500*. Madrid: Ediciones Cátedra.
- Madrigal L, Barbujani G. 2006. Partitioning of Genetic Variation in Human Populations and the Concept of Race. En: Crawford M, editor. *Anthropological Genetics. Theory, Methods and Applications*. Cambridge University Press. p 19–37.
- Malyarchuk BA, Rogozin IB, Berikov VB, Derenko M V. 2002. Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. *Hum Genet* 111:46–53.
- Malyarchuk BA, Vanecek T, Perkova MA, Derenko M V, Sip M. 2006. Mitochondrial DNA variability in the Czech population, with application to the ethnic history of Slavs. *Hum Biol* 78:681–96.
- Malyarchuk BA, Derenko M, Perkova M, Grzybowski T, Vanecek T, Lazur J. 2008. Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs. *Eur J Hum Genet* 16:1091–6.
- Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecek T, Tsybovsky I. 2010. The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS One* 5:e10285.
- Manen C, Marchand G, Carvalho AF. 2007. Le Néolithique ancien de la péninsule Ibérique: vers une nouvelle évaluation du mirage africain? En: *Actes du XXVIe congrès préhistorique de France, Avignon*. Vol. 3. Paris: Société Préhistorique Française. p 133–151.
- Margulis L. 1970. *Origin of Eukaryotic Cells*. New Haven, CT: Yale University Press.
- Margulis L. 1981. *Symbiosis in Cell Evolution*. San Francisco: Freeman.
- Martínez C, Marín A. 2000. 2. La Bética. En: Cano G, editor. *Conocer Andalucía*. Gran Enciclopedia Andaluza del siglo XXI. Sevilla: Ediciones Tartessos.
- Martínez-Cruz B, Harmant C, Platt DE, Haak W, Manry J, Ramos-Luis E, Soria-Hernanz DF, Bauduer F, Salaberria J, Oyharçabal B, Quintana-Murci L, Comas D. 2012. Evidence of Pre-Roman Tribal Genetic Structure in Basques from Uniparentally Inherited Markers. *Mol Biol Evol* 29:2211–2222.
- Mellars P. 2006. Archeology and the dispersal of modern humans in Europe: Deconstructing the “Aurignacian.” *Evol Anthropol Issues, News, Rev* 15:167–182.
- Mielke J, Fix A. 2006. The Confluence of Anthropological Genetics and Anthropological Demography. En: Crawford M, editor. *Anthropological Genetics. Theory, Methods and Applications*. Cambridge University Press. p 112–140.
- Mielnik-Sikorska M, Daca P, Malyarchuk B, Derenko M, Skonieczna K, Perkova M, Dobosz T, Grzybowski T. 2013. The History of Slavs Inferred from Complete Mitochondrial Genome Sequences. *PLoS One* 8:e54360.
- Milisauskas S. 2011. *Early Neolithic, the First Farmers in Europe, 7000–5500/5000 BC*. En: *European Prehistory. A Survey*. 2nd ed. Springer. p 153–221.

- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC. 2003. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* 100:171–176.
- Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, Atzmon G, Burns E, Ostrer H, Price AL, Reich D. 2011. The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet* 7:e1001373.
- Moreira JM, Ojeda J. 2003. Capítulo 3. El relieve y las costas andaluzas. En: López Ontiveros A, editor. *Geografía de Andalucía*. Barcelona: Ariel.
- Murdock SH, Ellis DR. 1991. *Applied Demography: An Introduction to Basic Concepts, Methods, and Data*. Boulder, CO: Westview Press.
- Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304.
- Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, Stoneking M. 2001. Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur J Hum Genet* 9:267–272.
- Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- Neiman M, Taylor DR. 2009. The causes of mutation accumulation in mitochondrial genomes. *Proc Biol Sci* 276:1201–1209.
- Niemi A-K, Moilanen JS, Tanaka M, Hervonen A, Hurme M, Lehtimäki T, Arai Y, Hirose N, Majamaa K. 2005. A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects. *Eur J Hum Genet* 13:166–170.
- Nogueiro I, Manco L, Gomes V, Amorim A, Gusmão L. 2010. Phylogeographic analysis of paternal lineages in NE Portuguese Jewish communities. *Am J Phys Anthr* 141:373–81.
- Nordborg M. 2003. Coalescent Theory. En: Balding D, Bishop M, Cannings C, editores. *Handbook of Statistical Genetics*. Wiley.
- Nothnagel M, Lu TT, Kayser M, Krawczak M. 2010. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum Mol Genet* 19:2927–35.
- Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. 2008. Genes mirror geography within Europe. *Nature* 456:98–101.
- Novembre J, Ramachandran S. 2011. Perspectives on human population structure at the cusp of the sequencing era. *Annu Rev Genomics Hum Genet* 12:245–74.
- Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon JM, Coudray C, Santachiara-Benerecetti AS, Semino O, Bandelt H-J, Torroni A. 2006. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* 314:1767–70.
- Onrubia-Pintado J. 1991. Canarias (Îles). En: *Encyclopédie berbère*. Éditions Peeters. p 1731–1755.

- Ottoni C, Martínez-Labarga C, Vitelli L, Scano G, Fabrini E, Contini I, Biondi G, Rickards O. 2009. Human mitochondrial DNA variation in Southern Italy. *Ann Hum Biol* 36:785–811.
- Ottoni C, Primativo G, Hooshar Kashani B, Achilli A, Martínez-Labarga C, Biondi G, Torroni A, Rickards O. 2010. Mitochondrial haplogroup H1 in North Africa: an early Holocene arrival from Iberia. *PLoS One* 5:e13378.
- van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386–94.
- Pakendorf B, Stoneking M. 2005. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6:165–183.
- Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, Tamm E, Karmin M, Reisberg T, Hooshar Kashani B, Perego UA, Carossa V, Gandini F, Pereira JB, Soares P, Angerhofer N, Rychkov S, Al-Zahery N, Carelli V, Sanati MH, Houshmand M, Hatina J, Macaulay V, Pereira L, Woodward SR, Davies W, Gamble C, Baird D, Semino O, Villems R, Torroni A, Richards MB. 2012. Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am J Hum Genet* 90:915–924.
- Pardiñas AF, Roca A, García-Vazquez E, López B. 2012. Assessing the genetic influence of ancient sociopolitical structure: micro-differentiation patterns in the population of Asturias (Northern Spain). *PLoS One* 7:e50206.
- Pardiñas AF, Martínez JL, Roca A, García-Vazquez E, López B. 2014. Over the sands and far away: interpreting an Iberian mitochondrial lineage with ancient Western African origins. *Am J Hum Biol* 26:777–783.
- Parson W, Parsons TJ, Scheithauer R, Holland MM. 1998. Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int J Leg Med* 111:124–132.
- Passarino G, Rose G. 2013. Mitochondrial DNA Polymorphisms. eLS.
- Penman BS, Pybus OG, Weatherall DJ, Gupta S. 2009. Epistatic interactions between genetic disorders of hemoglobin can explain why the sickle-cell gene is uncommon in the Mediterranean. *Proc Natl Acad Sci U S A* 106:21242–6.
- Pennarun E, Kivisild T, Metspalu E, Metspalu M, Reisberg T, Behar DM, Jones SC, Villems R. 2012. Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. *BMC Evol Biol* 12:234.
- Pereira L, Cunha C, Amorim A. 2004. Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. *Int J Leg Med* 118:132–6.
- Pereira L, Richards M, Goios A, Alonso A, Albarrán C, Garcia O, Behar DM, Gölge M, Hatina J, Al-Gazali L, Bradley DG, Macaulay V, Amorim A. 2005a. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res* 15:19–24.
- Pereira L, Cunha C, Alves C, Amorim A. 2005b. African female heritage in Iberia: a reassessment of mtDNA lineage distribution in present times. *Hum Biol* 77:213–229.

- Pereira L, Richards M, Goios A, Alonso A, Albarrán C, García O, Behar DM, Gölge M, Hatina J, Al-Gazali L, Bradley DG, Macaulay V, Amorim A. 2006. Evaluating the forensic informativeness of mtDNA haplogroup H sub-typing on a Eurasian scale. *Forensic Sci Int* 159:43–50.
- Pereira L, Freitas F, Fernandes V, Pereira JB, Costa MD, Costa S, Máximo V, Macaulay V, Rocha R, Samuels DC. 2009. The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet* 84:628–40.
- Pereira L, Silva NM, Franco-Duarte R, Fernandes V, Pereira JB, Costa MD, Martins H, Soares P, Behar DM, Richards M, Macaulay V. 2010a. Population expansion in the North African late Pleistocene signalled by mitochondrial DNA haplogroup U6. *BMC Evol Biol* 10:390.
- Pereira L, Cerný V, Cerezo M, Silva NM, Hájek M, Vasíková A, Kujanová M, Brdicka R, Salas A. 2010b. Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. *Eur J Hum Genet* 18:915–23.
- Pérez J. 1988. La crisis del siglo XVII. En: Domínguez Ortiz A, editor. *Historia de España*. Barcelona: Planeta.
- Pérez V, Reher D. 1988. *Demografía histórica en España*. Madrid: El Arquero.
- Pérez V, Rowland R. 1997. La Península Ibérica. En: Bardet J-P, Dupâquier J, editores. *Historia de las poblaciones de Europa. Volumen I: De los orígenes a las premisas de la revolución demográfica*. Madrid: Editorial Síntesis.
- Philibert RA, Zadorozhnyaya O, Beach SRH, Brody GH. 2008. Comparison of the genotyping results using DNA obtained from blood and saliva. *Psychiat Genet* 18:275–81.
- Pickrell JK, Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet* 8:e1002967.
- Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J. 2012. The genetic history of Europeans. *Trends Genet* 28:496–505.
- Plácido D. 1989. La España romana y visigoda (siglos III a. C.-VII d. C.). En: Domínguez Ortiz A, editor. *Historia de España*. 1st ed. Barcelona: Planeta.
- Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, Comas D. 2003. Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann Hum Genet* 67:312–328.
- Podgorná E, Soares P, Pereira L, Černý V. 2013. The Genetic Impact of the Lake Chad Basin Population in North Africa as Documented by Mitochondrial Diversity and Internal Variation of the L3e5 Haplogroup. *Ann Hum Genet* 77:513–523.
- Price AL, Butler J, Patterson N, Capelli C, Pascali VL, Scarnicci F, Ruiz-Linares A, Groop L, Saetta AA, Korkolopoulou P, Seligsohn U, Waliszewska A, Schirmer C, Ardlie K, Ramos A, Nemesi J, Arbeitman L, Goldstein DB, Reich D, Hirschhorn JN. 2008. Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* 4:e236.

- Pugach I, Stoneking M. 2015. Genome-wide insights into the genetic history of human populations. *Investig Genet* 6:6.
- Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu M V, Carracedo A. 2004. Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int* 140:251–257.
- Ragoussis J. 2009. Genotyping Technologies for Genetic Research. *Annu Rev Genomics Hum Genet* 10:117–133.
- Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, 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* 62:531–50.
- Rando JC, Cabrera VM, Larruga JM, Hernández M, González AM, Pinto F, Bandelt H-J. 1999. Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* 63:413–428.
- Reales G, Hernández CL, Dugoujon J-M, Novelletto A, Cuesta P, Fortes-Lima C, Rodríguez JN, Calderón R. 2014. New insights into the distribution of APOE polymorphism in the Iberian Peninsula. The case of Andalusia (Spain). *Ann Hum Biol* 41:443–452.
- Rhouda T, Martínez-Redondo D, Gómez-Durán A, Elmtili N, Idaomar M, Díez-Sánchez C, Montoya J, López-Pérez MJ, Ruiz-Pesini E. 2009. Moroccan mitochondrial genetic background suggests prehistoric human migrations across the Gibraltar Strait. *Mitochondrion* 9:402–7.
- Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B. 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59:185–203.
- Richards MB, Macaulay VA, Bandelt H-J, Sykes BC. 1998. Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62:241–260.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge 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, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, Bandelt H-J. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251–76.
- Di Rienzo A, Wilson AC. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci U S A* 88:1597–1601.
- Rito T, Richards MB, Fernandes V, Alshamali F, Cerny V, Pereira L, Soares P. 2013. The First Modern Human Dispersals across Africa. *PLoS One* 8:e80031.
- Roostalu U, Kutuev I, Loogväli E-L, Metspalu E, Tambets K, Reidla M, Khusnutdinova EK, Usanga E, Kivisild T, Villems R. 2007. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol* 24:436–48.
- Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioglu C, Khusnutdinova E, Rudan P,

- Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan J-P, Chaventre A, Ferak V, Füredi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzić R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 75:128–37.
- Rosa A, Brehm A. 2011. African human mtDNA phylogeography at-a-glance. *J Anthr Sci* 89:1–34.
- Rosenberg MS, Anderson CD. 2011. PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods Ecol Evol* 2:229–232.
- Rosenberg NA, Li LM, Ward R, Pritchard JK. 2003. Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* 73:1402–22.
- Rosser ZH, Zerjal T, Hurler ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar A V, Mitchell RJ, Nafa K, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67:1526–43.
- Rowold DJ, Luis JR, Terreros MC, Herrera RJ. 2007. Mitochondrial DNA gene flow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. *J Hum Genet* 52:436–47.
- Rubicz R, Melton P, Crawford M. 2006. Molecular Markers in Anthropological Genetic Studies. En: Crawford M, editor. *Anthropological Genetics. Theory, Methods and Applications*. Cambridge University Press. p 141–186.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303:223–226.
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67:718–26.
- Salas A, Comas D, Lareu M V, Bertranpetit J, Carracedo A. 1998. mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* 6:365–375.
- Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. *Am J Hum Genet* 71:1082–111.
- Salas A, Quintáns B, Álvarez-Iglesias V. 2005a. SNaPshot Typing of Mitochondrial DNA Coding Region Variants. En: Carracedo A, editor. *Forensic DNA Typing Protocols (Methods in Molecular Biology, vol. 297)*. Vol. 297. New Jersey: Humana Press. p 197–208.

- Salas A, Carracedo A, Macaulay V, Richards M, Bandelt H-J. 2005b. A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. *Biochem Biophys Res Commun* 335:891-9.
- De Santiago E. 2000. 8. El Estado Nazarí de Granada. En: Cano G, editor. *Conocer Andalucía*. Gran Enciclopedia Andaluza del siglo XXI. Sevilla: Ediciones Tartessos.
- Sayas J, García L. 1982. II. Romanismo y germanismo. El despertar de los pueblos hispánicos (siglos IV-X). En: Tuñón de Lara M, editor. *Historia de España*. Barcelona: Labor.
- Schapira AH V. 2006. Mitochondrial disease. *Lancet* 368:70-82.
- Scheffler IE. 2001. Mitochondria make a come back. *Adv Drug Deliv Rev* 49:3-26.
- Schneider S, Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* 152:1079-89.
- Schönberg A, Theunert C, Li M, Stoneking M, Nasidze I. 2011. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur J Hum Genet* 19:988-994.
- Schulten A. 2006. Tartessos. Contribución a la historia más antigua de Occidente. Córdoba: Almuzara.
- Schwartz M, Vissing J. 2002. Paternal inheritance of mitochondrial DNA. *N Engl J Med* 347:576-580.
- Secher B, Fregel R, Larruga JM, Cabrera VM, Endicott P, Pestano JJ, González AM. 2014. The history of the North African mitochondrial DNA haplogroup U6 gene flow into the African, Eurasian and American continents. *BMC Evol Biol* 14:109.
- Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspina A-S, Manica A, Moltke I, Albrechtsen A, Ko A, Margaryan A, Moiseyev V, Goebel T, Westaway M, Lambert D, Khartanovich V, Wall JD, Nigst PR, Foley RA, Lahr MM, Nielsen R, Orlando L, Willerslev E. 2014. Genomic structure in Europeans dating back at least 36,200 years. *Science* 346:1113-1118.
- Seielstad MT, Minch E, Cavalli-Sforza LL. 1998. Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278-80.
- Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, Silva G, Belmont JW, Klareskog L, Gregersen PK. 2006. European population substructure: clustering of northern and southern populations. *PLoS Genet* 2:e143.
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. 2000. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290:1155-9.
- Semino O, Magri C, Benuzzi G, Lin A, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS. 2004. Origin, diffusion, and differentiation of Y-

- chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* 74:1023–34.
- Sigurðardóttir S, Helgason a, Gulcher JR, Stefansson K, Donnelly P. 2000. The mutation rate in the human mtDNA control region. *Am J Hum Genet* 66:1599–1609.
- Sjodin P, François O. 2011. Wave-of-advance models of the diffusion of the Y chromosome haplogroup R1b1b2 in Europe. *PLoS One* 6:e21592.
- Slate J, Gemmell N. 2004. Eve “n” Steve: recombination of human mitochondrial DNA. *Trends Ecol Evol* 19:561–563.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84:740–759.
- Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt H-J, Torroni A, Richards MB. 2010. The archaeogenetics of Europe. *Curr Biol* 20:R174–183.
- Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilová E, Macaulay V, Richards MB, Cerny V, Pereira L. 2012. The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol Biol Evol* 29:915–927.
- Soffer O, Gamble C. 1990. *The world at 18000 BP*. London: Hyman.
- Solovieff N, Hartley SW, Baldwin CT, Klings ES, Gladwin MT, Taylor JG, Kato GJ, Farrer LA, Steinberg MH, Sebastiani P. 2011. Ancestry of African Americans with sickle cell disease. *Blood Cells Mol Dis* 47:41–5.
- Souville G. 1998. Contacts et échanges entre la péninsule Ibérique et le Nord-Ouest de l’Afrique durant les temps préhistoriques et protohistoriques. En: *Comptes-rendus des séances de l’Académie des Inscriptions et Belles-Lettres*. Vol. 142. p 163–177.
- Stevanovitch A, Gilles A, Bouzaid E, Kefi R, Paris F, Gayraud RP, Spadoni JL, El-Chennawi F, Beraud-Colomb E. 2003. Mitochondrial DNA Sequence Diversity in a Sedentary Population from Egypt. *Ann Hum Genet* 68:23–39.
- Stewart JB, Freyer C, Elson JL, Larsson N-G. 2008. Purifying selection of mtDNA and its implications for understanding evolution and mitochondrial disease. *Nat Rev Genet* 9:657–662.
- Strachan T, Read A. 2011. *Human Molecular Genetics*. 4th editio. New York: Garland Science/Taylor & Francis Group.
- Straus LG. 2001. Africa and Iberia in the Pleistocene. *Quatern Int* 75:91–102.
- Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–60.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–95.
- Taylor R, Turnbull DM. 2005. Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 6:389–402.

- Templado J, Calvo M. 2006. Flora y Fauna de la Reserva Marina y Reserva de Pesca de la isla de Alborán. Madrid: . Secretaría General de Pesca Marítima, MAPA.
- Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. 2000. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc Natl Acad Sci U S A* 97:7360–7365.
- Tishkoff SA, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet* 4:293–340.
- Tishkoff SA, Kidd KK. 2004. Implications of biogeography of human populations for “race” and medicine. *Nat Genet* 36:S21–S27.
- Tishkoff SA, Gonder MK. 2006. Human Origins Within and Out of Africa. En: Crawford MH, editor. *Anthropological Genetics. Theory, Methods and Applications*. Cambridge University Press. p 337–379.
- Tomas C, Sanchez JJ, Barbaro A, Brandt-Casadevall C, Hernandez A, Ben Dhiab M, Ramon M, Morling N. 2008. X-chromosome SNP analyses in 11 human Mediterranean populations show a high overall genetic homogeneity except in North-west Africans (Moroccans). *BMC Evol Biol* 8:75.
- Torres-Ortiz M. 1999. *Sociedad y Mundo Funerario en Tartessos*. Madrid: Real Academia de la Historia.
- Torrioni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC. 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 TG, Cabell MF, Brown MD, Neel J V, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53:563–590.
- Torrioni A, Lott MT, Cabell MF, Chen Y-S, Lavergne L, Wallace DC. 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.
- Torrioni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus M-L, Wallace DC. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835–1850.
- Torrioni A, Bandelt H-J, D’Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus M-L, Bonn -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.
- Torrioni A, Bandelt H-J, Macaulay V, Richards M, Cruciani F, Rengo C, Mart nez-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, Koivum ki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O, Scozzari R. 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69:844–852.

- Torróni A, Achilli A, Macaulay V, Richards M, Bandelt H-J. 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22:339–45.
- Tuñón de Lara M, Tarradell M, Mangas J. 1980. I. Introducción, primeras culturas e Hispania romana. En: Tuñón de Lara M, editor. *Historia de España*. Barcelona: Labor.
- Turchi C, Buscemi L, Giacchino E, Onofri V, Fendt L, Parson W, Tagliabracci A. 2009. Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enrichment of forensic mtDNA databases with Northern Africa data. *Forensic Sci Int Genet* 3:166–72.
- Underhill PA, Kivisild T. 2007. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41:539–564.
- Valencia R. 2000. 4. Origen y constitución de Al-Andalus. En: Cano G, editor. *Conocer Andalucía*. Gran Enciclopedia Andaluza del siglo XXI. Sevilla: Ediciones Tartessos.
- Valle B. 2000. 5. Ríos, cuencas hidrográficas y gestión de las aguas. En: Cano G, editor. *Conocer Andalucía*. Gran Enciclopedia Andaluza del siglo XXI. Sevilla: Ediciones Tartessos.
- Verpoorte A. 2009. Limiting factors on early modern human dispersals: The human biogeography of late Pleniglacial Europe. *Quatern Int* 201:77–85.
- Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci U S A* 86:9350–9354.
- Villaverde V, Aura JE, Barton CM. 1998. The Upper Paleolithic in Mediterranean Spain: A Review of Current Evidence. *J World Prehist* 12:121–198.
- Villaverde V, Román D, Ripoll MP, Bergadà MM, Real C. 2012. The end of the Upper Palaeolithic in the Mediterranean Basin of the Iberian Peninsula. *Quatern Int* 272-273:17–32.
- Vincze T, Posfai J, Roberts RJ. 2003. NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res* 31:3688–91.
- Walker M, Johnsen S, Rasmussen SO, Popp T, Steffensen J-P, Gibbard P, Hoek W, Lowe J, Andrews J, Björck S, Cwynar LC, Hughen K, Kershaw P, Kromer B, Litt T, Lowe DJ, Nakagawa T, Newnham R, Schwander J. 2009. Formal definition and dating of the GSSP (Global Stratotype Section and Point) for the base of the Holocene using the Greenland NGRIP ice core, and selected auxiliary records. *J Quat Sci* 24:3–17.
- Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, Elsas LJ, Nikoskelainen EK. 1988. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* 242:1427–1430.
- White DJ, Wolff JN, Pierson M, Gemmell NJ. 2008. Revealing the hidden complexities of mtDNA inheritance. *Mol Ecol* 17:4925–4942.
- Weir B, Cockerham C. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution (N Y)* 38:1358 – 1370.

- Wilder JA, Mobasher Z, Hammer MF. 2004. Genetic evidence for unequal effective population sizes of human females and males. *Mol Biol Evol* 21:2047–2057.
- Wright S. 1965. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution (N Y)* 19:395–420.
- Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348.
- Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR. 1985. Mitochondrial origins. *Proc Natl Acad Sci U S A* 82:4443–4447.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.
- Yang Z. 2006. *Computational Molecular Evolution*. Oxford Ser. Oxford University Press.
- Yang Z, Rannala B. 2012. Molecular phylogenetics: principles and practice. *Nat Rev Genet* 13:303–314.
- Zapata L, Peña-Chocarro L, Pérez-Jordá G, Stika HP. 2004. Early neolithic agriculture in the Iberian Peninsula. *J World Prehist* 18:283–325.
- Zeder MA. 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci U S A* 105:11597–604.
- Zhang J, Asin-Cayueta J, Fish J, Michikawa Y, Bonafe M, Olivieri F, Passarino G, De Benedictis G, Franceschi C, Attardi G. 2003. Strikingly higher frequency in centenarians and twins of mtDNA mutation causing remodeling of replication origin in leukocytes. *Proc Natl Acad Sci U S A* 100:1116–1121.
- Zilhão J, D’Errico F. 1999. The Chronology and Taphonomy of the Earliest Aurignacian and Its Implications for the Understanding of Neandertal Extinction. *J World Prehist* 13:1–68.
- Zilhão J. 2001. Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proc Natl Acad Sci U S A* 98:14180–5.

Anexos

ANEXO I (CAPÍTULO III)	243
Tabla S1. Base de datos para análisis comparativos poblacionales del ADNmt global	243
Tabla S2. Caracterización molecular de la diversidad mitocondrial en Andalucía	245
Figura S1. Genealogía de las secuencias mitocondriales de Andalucía	246
ANEXO II (CAPÍTULO IV)	249
Tabla S3. Diseño de los ensayos de genotipado TaqMan para los linajes U6 y L/M	249
Tabla S4. Base de datos para análisis comparativos poblacionales basados en haplogrupos mitocondriales africanos	250
Tabla S5. Haplotipos mitocondriales completos de las muestras andaluzas y bereberes de Marruecos pertenecientes a linajes africanos	255
Tabla S6. Edades de coalescencia estimadas de los linajes U6, M1 y L	256
Figura S2. Localización geográfica de las poblaciones usadas en mapas de contorno	257
Figura S3. Distribución espacial del haplogrupo U6	258
Figura S4. Distribución espacial del haplogrupo M1	259
Figura S5. Distribución espacial del haplogrupo L2a	260
Figura S6. Distribución espacial del haplogrupo L2b	261
Figura S7. Distribución espacial del haplogrupo L3b	262
Figura S8. Distribución espacial del haplogrupo L3d	263
Figura S9. Distribución espacial del haplogrupo L3f	264
Figura S10. Distribución espacial del haplogrupo L3h1b	265
Figura S11. Genealogías de los haplogrupos U6a y L1b	266
Figura S12. Proporción probabilística de <i>clusters</i> fundadores L en dos eventos migratorios	267
Figura S13. Representación gráfica de la frecuencia de fundadores L en dos eventos migratorios	268

Dataset S1. Base de datos de 2.182 mitogenomas para la construcción de las filogenias de haplogrupos africanos	269
Dataset S2. Filogenia del haplogrupo U6.....	269
Dataset S3. Filogenia del haplogrupo M1.....	269
Dataset S4. Filogenia del haplogrupo L1	269
Dataset S5. Filogenia del haplogrupo L2	269
Dataset S6. Filogenia del haplogrupo L3	269
ANEXO III (CAPÍTULO V)	271
Tabla S7. Diseño de los ensayos de genotipado TaqMan para los linajes H y H1 ..	271
Tabla S8. Diseño de las reacciones <i>multiplex</i> para el genotipado interno del linaje H.....	272
Tabla S9. Base de datos para análisis comparativo basado en el haplogrupo H...274	
Tabla S10. Caracterización molecular de los individuos andaluces pertenecientes al haplogrupo H.....	278
Figura S14. Localización geográfica de las poblaciones usadas en mapas de contorno (<i>Surface Maps</i>).....	279
Figura S15. Genealogía de las secuencias mitocondriales H en Andalucía.....	280
Figura S16. Distribución espacial de los sub-haplogrupos H que presentan amplias variaciones geográficas (H2, H4, H5 y H6)	281
Figura S17. Distribución espacial de los sub-haplogrupos H con una distribución preferencial en Mediterráneo occidental y este de Europa (H17 y H18).....	282
Figura S18. Distribución espacial de los sub-haplogrupos H con una distribución preferencial en la Europa central y oriental (H7, H10 y H11).....	283
Figura S19. Distribución espacial de los sub-haplogrupos H con una distribución preferencial en el Mediterráneo central y oriental (H13 y H20)	284

ANEXO I (CAPÍTULO III)

Tabla S1. Conjunto de poblaciones humanas seleccionadas para el análisis comparativo de los patrones de ADNmt globales. Las regiones geográficas establecidas son: *IB* (Iberia), *EUR* (Europa), *NME* (Oriente Medio y Próximo), *SWA* (sudeste de Asia), *NAF* (norte de África).

Población	Región	N	Referencia	Código
Galicia	<i>IB</i>	135	González et al., 2003	1
País Vasco	<i>IB</i>	167	Larruga et al., 2001	2
León	<i>IB</i>	61	Larruga et al., 2001	3
Zamora	<i>IB</i>	214	Álvarez et al., 2010	4
Andalucía (Huelva)	<i>IB</i>	158	<i>presente estudio</i>	5
Andalucía (Granada)	<i>IB</i>	121	<i>presente estudio</i>	6
Andalucía (Córdoba)	<i>IB</i>	108	Casas et al., 2006	7
Islas Canarias	<i>IB</i>	300	Rando et al., 1999	8
España	<i>IB</i>	312	Álvarez et al., 2007	9
Portugal norte	<i>IB</i>	187	Pereira et al., 2004	10
Portugal centro	<i>IB</i>	239	Pereira et al., 2004	11
Portugal sur	<i>IB</i>	123	Pereira et al., 2004	12
Francia	<i>EUR</i>	210	Dubut et al., 2004	13
Escocia	<i>EUR</i>	874	González et al., 2003	14
Noruega	<i>EUR</i>	323	Helgason et al., 2001	15
Finlandia	<i>EUR</i>	50	Sajantila et al., 1995	16
Saami	<i>EUR</i>	176	Helgason et al., 2001	17
Alemania	<i>EUR</i>	266	González et al., 2003	18
República Checa	<i>EUR</i>	300	Vidrová et al., 2008	19
Italia	<i>EUR</i>	395	Turchi et al., 2008	20
Italia centro	<i>EUR</i>	322	Achilli et al., 2007	21
Basilicata (Italia sur)	<i>EUR</i>	92	Ottoni et al., 2009	22
Calabria (Italia sur)	<i>EUR</i>	48	Ottoni et al., 2009	23
Sicilia 1	<i>EUR</i>	49	Vona et al., 2001	24
Sicilia 2	<i>EUR</i>	75	Ottoni et al., 2009	25
Grecia norte	<i>EUR</i>	319	Irwin et al., 2008	26
Creta	<i>EUR</i>	283	Martínez et al., 2008	27
Chipre	<i>EUR</i>	91	Irwin et al., 2008	28
Adygei	<i>EUR</i>	50	Macaulay et al., 1999	29
Palestina	<i>NME</i>	118	Abu-Amero et al., 2007	30
Jordán	<i>NME</i>	145	Abu-Amero et al., 2007	31
Siria	<i>NME</i>	119	Abu-Amero et al., 2007	32
Turquía	<i>SWA</i>	494	Abu-Amero et al., 2007	33
Irak	<i>SWA</i>	116	Abu-Amero et al., 2007	34
Kurdos	<i>SWA</i>	212	Abu-Amero et al., 2007	35
Irán	<i>SWA</i>	712	Abu-Amero et al., 2007	36
Qatar	<i>SWA</i>	90	Rowold et al., 2007	37
Emiratos Árabes Unidos	<i>SWA</i>	131	Rowold et al., 2007	38
Omán	<i>SWA</i>	105	Rowold et al., 2007	39

Población	Región	N	Referencia	Código
Arabia Saudí	SWA	553	Abu-Amero et al., 2008	40
Yemen	SWA	214	Abu-Amero et al., 2007	41
Berberes Marruecos 1	NAF	53	Coudray et al., 2009	42
Berberes Marruecos 2	NAF	70	Coudray et al., 2009	43
Berberes Marruecos 3	NAF	94	Coudray et al., 2009	44
Saharai	NAF	56	Plaza et al., 2003	45
Argelia	NAF	47	Plaza et al., 2003	46
Árabes Túnez	NAF	50	Cherni et al., 2005	47
Berberes Túnez 1	NAF	53	Fadhlaoui-Zid et al., 2004	48
Berberes Túnez 2	NAF	53	Fadhlaoui-Zid et al., 2004	49
Berberes Túnez 3	NAF	49	Fadhlaoui-Zid et al., 2004	50
Berberes Túnez 4	NAF	47	Cherni et al., 2005	51
Berberes Egipto	NAF	78	Coudray et al., 2009	52
Egipto	NAF	126	Abu-Amero et al., 2007	53

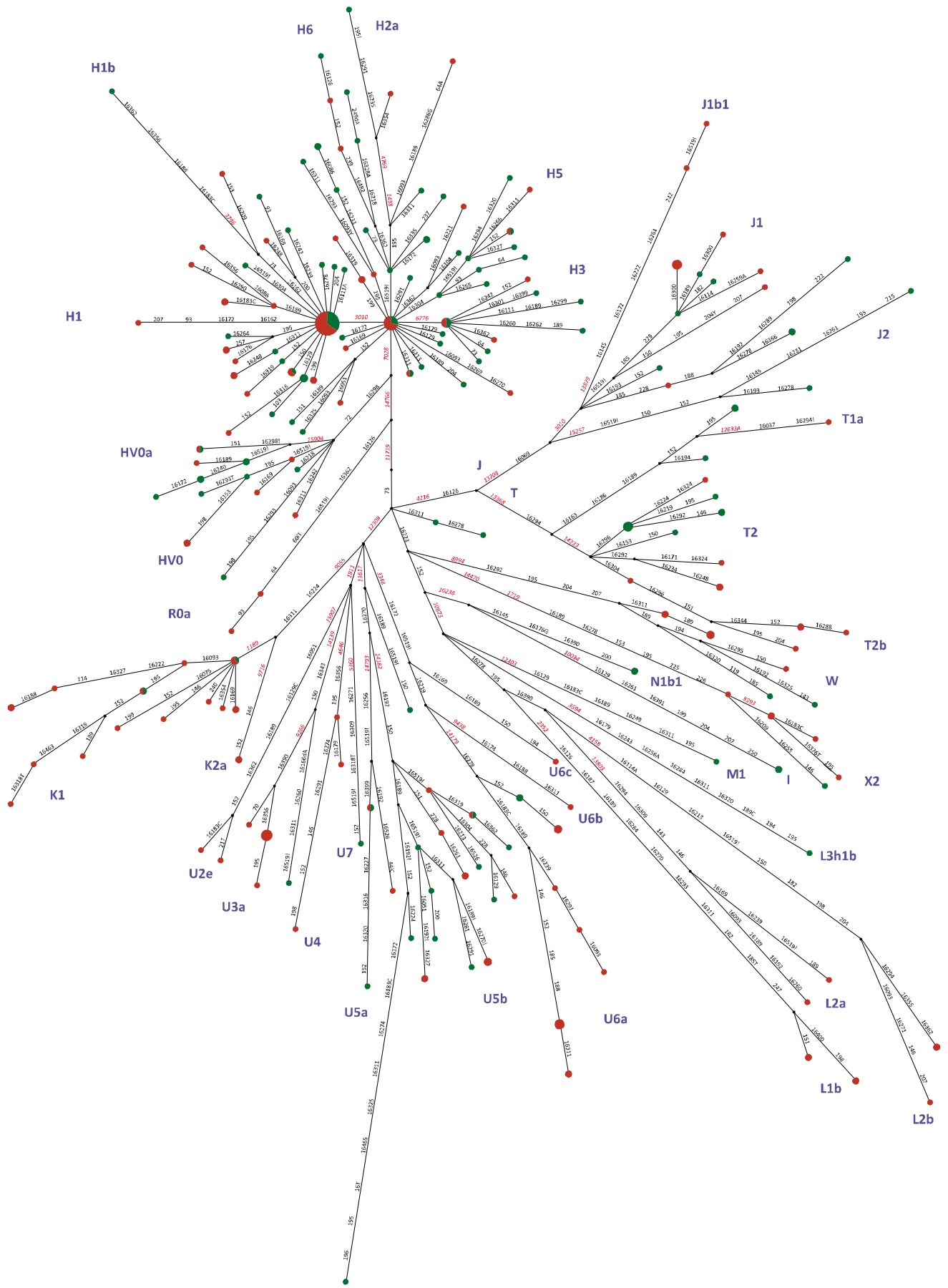
Referencias

- Abu-Amero KK et al. (2007) *BMC Evol Biol* 7:32.
 Abu-Amero KK et al. (2008) *BMC Evol Biol* 8:45.
 Achilli A et al. (2007) *Am J Hum Genet* 80:759–768.
 Álvarez JC et al. (2007) *Leg Med (Tokyo)* 9:293–304.
 Álvarez L et al. (2010) *Am J Phys Anthropol* 142:531–539.
 Casas MJ et al. (2006) *Am J Phys Anthropol* 131:539–551.
 Cherni L et al. (2005) *Hum Biol* 77:61–70.
 Coudray C et al. (2009) *Ann Hum Genet* 73:196–214.
 Dubut V et al. (2004) *Eur J Hum Genet* 12:293–300.
 Fadhlaoui-Zid K et al. (2004) *Ann Hum Genet* 68:222–233.
 González AM et al. (2003) *Am J Phys Anthropol* 120:391–404.
 Helgason A et al. (2001) *Am J Hum Genet* 68:723–737.
 Irwin J et al. (2008) *Int J Legal Med* 122:87–89.
 Larruga JM et al. (2001) *Eur J Hum Genet* 9:708–716.
 Macaulay V et al. (1999) *Am J Hum Genet* 64:232–249.
 Martínez L et al. (2008) *Am J Phys Anthropol* 137:213–223.
 Ottoni C et al. (2009) *Ann Hum Biol* 36:785–811.
 Pereira L et al. (2004) *Int J Legal Med* 118:132–136.
 Plaza S et al. (2003) *Ann Hum Genet* 67:312–328.
 Rando JC et al. (1999) *Ann Hum Genet* 63:413–428.
 Rowold DJ et al. (2007) *J Hum Genet* 52:436–447.
 Sajantila A et al. (1995) *Genome Res* 5:42–52.
 Turchi C et al. (2008) *Int J Legal Med* 122:199–204.
 Vidrová V et al. (2008) *Hum Biol* 80:669–674.
 Vona G et al. (2001) *Am J Hum Biol* 13:576–589.

Tabla S2. Caracterización molecular mitocondrial de las muestras andaluzas analizadas. Las transversiones se indican con el cambio explícito de la base nucleotídica, las deleciones con una 'd' y la heteroplasmia con 'Y' (C/T). Los polimorfismos de la región codificante se muestran en el orden jerárquico en que se han caracterizado. Se incluyen los números de referencia de GenBank para las secuencias de la región control.

-----**VER CD adjunto**-----

Figura S1. Relaciones filogenéticas de haplotipos mitocondriales en Andalucía. La arquitectura de estas conexiones en red ha sido proporcionada a través del uso del programa Network v. 4.612 (<http://www.fluxus-engineering.com/>). El método filogenético usado ha sido el *Median-Joining Network* (MJN). Los haplogrupos se indican en azul. Las posiciones mutadas de las ramas se refieren a la secuencia de referencia rCRS. Todos los polimorfismos son mutaciones de la región control, excepto los que se presentan en rojo (posiciones diagnósticas de haplogrupo de la región codificante). Las transversiones se muestran con el cambio de base tras la mutación. El sufijo ‘i’ indica una inserción, ‘d’ una deleción e ‘Y’ una heteroplasmia. Las retromutaciones se indican con un signo de exclamación. El tamaño de los círculos es proporcional a la frecuencia del haplotipo. Las muestras de Huelva se indican en rojo y las de Granada en verde.



ANEXO II (CAPÍTULO IV)

Tabla S3. Diseño de los ensayos TaqMan para el genotipado de los clados mitocondriales U6 y L/M.

Posición	Haplogrupo		Secuencia (5'-3')	Concentración final (μM)	
3348	U6	<i>primers</i>	3311F	CCATGGCCAACCTCCTACTC	0,9
			3381R	TCGTTCGGTAAGCATTAGGAATG	0,9
		<i>sondas</i>	alelo A	VIC -ATTGTACCCATTCTAATC	0,2
			alelo G	FAM -ATTGTACCCATTCTGAT	0,2
10873	L o M	<i>primers</i>	10837F	CACAACCACCCACAGCCTAATT	0,9
			10981R	GGGTAGGAGTCAGGTAGTTAGTATTAGGA	0,9
		<i>sondas</i>	alelo T	VIC -CATCCCTCTACTATT	0,2
			alelo C	FAM -TCATCCCCTACTATTTT	0,2

Tabla S4. Conjunto de poblaciones humanas seleccionadas para el análisis comparativo de los linajes mitocondriales africanos. Las regiones geográficas establecidas son: *IB* (Iberia), *EUR* (Europa), *CME* (Europa mediterránea central), *EEU* (Europa oriental), *ANA* (Anatolia), *NES* (Oriente Próximo), *ARA* (Arabia), *SWAs* (sudoeste de Asia), *NAF* (norte de África), *WAF* (África occidental), *CAF* (África central), *SWA* (sudoeste de África), *EAF* (África oriental), *SEA* (sudeste de África), *SAF* (Sudáfrica), *MAD* (Madagascar).

Código	Región	Población	Referencia	N
1	<i>IB</i>	Galicia	Richards et al., 2000; González et al., 2003	136
2	<i>IB</i>	León	Larruga et al., 2001	61
3	<i>IB</i>	Cantabria	Maca-Meyer et al., 2003	88
4	<i>IB</i>	Cantabria Pasidegos	Maca-Meyer et al., 2003; Cardoso et al., 2010	143
5	<i>IB</i>	País Vasco	Richards et al., 2000	163
6	<i>IB</i>	Zamora	Álvarez et al., 2010	214
7	<i>IB</i>	Andalucía (Huelva)	<i>presente estudio</i>	280
8	<i>IB</i>	Andalucía (Córdoba)	Casas et al., 2006	108
9	<i>IB</i>	Andalucía (Granada)	<i>presente estudio</i>	470
10	<i>IB</i>	Islas Canarias	Rando et al., 1999	300
11	<i>IB</i>	Ibiza	Picornell et al., 2005	50
12	<i>IB</i>	Mallorca	Picornell et al., 2005	93
13	<i>IB</i>	Portugal (norte)	González et al., 2003; Pereira et al., 2004	271
14	<i>IB</i>	Portugal (centro)	González et al., 2003; Pereira et al., 2004	317
15	<i>IB</i>	Portugal (sur)	González et al., 2003; Pereira et al., 2004	260
16	<i>IB</i>	Azores	Brehm et al., 2003; Santos et al., 2003	235
17	<i>IB</i>	Madeira	Brehm et al., 2003	88
18	<i>EUR</i>	Francia	Dubut et al., 2004	210
19	<i>EUR</i>	Córcega (sur)	Falchi et al., 2006	53
20	<i>EUR</i>	Austria (oeste)	Richards et al., 2000	101
21	<i>EUR</i>	Eslovenia	Malyarchuk et al., 2003	103
22	<i>EUR</i>	Eslovaquia	Malyarchuk et al., 2008	207
23	<i>EUR</i>	Hungría	Irwin et al., 2007	416
24	<i>CME</i>	Italia (Turín)	Turchi et al., 2008	50
25	<i>CME</i>	Italia (Toscana)	Richards et al., 2000; Falchi et al., 2006; Achilli et al., 2007; Turchi et al., 2008	480
26	<i>CME</i>	Italia (Ancona)	Turchi et al., 2008	73
27	<i>CME</i>	Italia (Roma)	Babalini et al., 2005; Turchi et al., 2008	110
28	<i>CME</i>	Italia (sur)	Babalini et al., 2005	103
29	<i>CME</i>	Italia (Basilicata)	Otoni et al., 2009a	92
30	<i>CME</i>	Italia (Calabria)	Otoni et al., 2009a	95
31	<i>CME</i>	Cerdeña (norte)	Falchi et al., 2006	50
32	<i>CME</i>	Cerdeña (centro)	Falchi et al., 2006	52
33	<i>CME</i>	Cerdeña	Richards et al., 2000	69
34	<i>CME</i>	Cerdeña (sur)	Falchi et al., 2006	133

Código	Región	Población	Referencia	N
35	<i>CME</i>	Sicilia	Ottoni et al., 2009a	154
36	<i>EEU</i>	Bosnia	Malyarchuk et al., 2003	144
37	<i>EEU</i>	Grecia (norte)	Irwin et al., 2008	319
38	<i>EEU</i>	Chipre	Irwin et al., 2008	91
39	<i>ANA</i>	Turquía	Richards et al., 2000	271
40	<i>NES</i>	Líbano	Badro et al., 2013	980
41	<i>NES</i>	Siria	Badro et al., 2013	234
42	<i>NES</i>	Druze Galilee	Shlush et al., 2008	185
43	<i>NES</i>	Jordán	González et al., 2009	101
44	<i>NES</i>	Palestina	Badro et al., 2013	120
45	<i>NES</i>	Irak	Al-Zahery et al., 2003	52
46	<i>ARA</i>	Arabia Saudí (oeste)	Abu-Amero et al., 2008	59
47	<i>ARA</i>	Arabia Saudí (sur)	Abu-Amero et al., 2008	88
48	<i>ARA</i>	Yemen	Kivisild et al., 2004; Non et al., 2011	210
49	<i>ARA</i>	Yemen (Tihama)	Cerný et al., 2008	58
50	<i>ARA</i>	Yemen (Ta'izz)	Cerný et al., 2008	68
51	<i>ARA</i>	Arabia Saudí (centro)	Abu-Amero et al., 2008	220
52	<i>ARA</i>	Dubai	Alshamali et al., 2008	249
53	<i>SWAs</i>	Irán	Metspalu et al., 2004	436
54	<i>NAF</i>	Sahara	Plaza et al., 2003	56
55	<i>NAF</i>	Marruecos (Souss)	Brakez et al., 2001	50
56	<i>NAF</i>	Bereber Marruecos (Asni)	Coudray et al., 2009	53
57	<i>NAF</i>	Marruecos (Marrakech)	Falchi et al., 2006	52
58	<i>NAF</i>	Marruecos (El Jadida)	Harich et al., 2010	81
59	<i>NAF</i>	Marruecos	Rando et al., 1998; Turchi et al., 2009	145
60	<i>NAF</i>	Marruecos (Tetouan)	Rhouda et al., 2009	209
61	<i>NAF</i>	Marruecos (Al Hoceima)	Rhouda et al., 2009	79
62	<i>NAF</i>	Bereber Marruecos (Bouhria)	Coudray et al., 2009	70
63	<i>NAF</i>	Marruecos (Oujda)	Rhouda et al., 2009	189
64	<i>NAF</i>	Bereber Marruecos (Figuig)	Coudray et al., 2009	94
65	<i>NAF</i>	Argelia	Bekada et al., 2013	240
66	<i>NAF</i>	Túnez (Testour)	Cherni et al., 2009	50
67	<i>NAF</i>	Túnez	Cherni et al., 2009; Turchi et al., 2009	115
68	<i>NAF</i>	Túnez (Zriba)	Cherni et al., 2005	50
69	<i>NAF</i>	Túnez (Jerba)	Loueslati et al., 2006	59
70	<i>NAF</i>	Libia	Fadhlaoui-Zid et al., 2011	269
71	<i>NAF</i>	Tuaregs Libia	Ottoni et al., 2009b	129
72	<i>NAF</i>	Bereber Egipto (Siwa)	Coudray et al., 2009	78
73	<i>NAF</i>	Egipto (Alexandria)	Saunier et al., 2009	277
74	<i>NAF</i>	Egipto	Krings et al., 1999	66
75	<i>NAF</i>	Egipto (sur)	Krings et al., 1999	80
76	<i>WAF</i>	Cabo Verde	Brehm et al., 2002	292
77	<i>NAF</i>	Mauritania	González et al., 2006	64
78	<i>WAF</i>	Senegal	Watson et al., 1997; Rando et al., 1998; Stefflova et al., 2009	280

Código	Región	Población	Referencia	N
79	<i>WAF</i>	Sierra Leona	Jackson et al., 2005	276
80	<i>WAF</i>	Mali	Ely et al., 2006; González et al., 2006	204
81	<i>WAF</i>	Ghana	Veeramah et al., 2010	238
82	<i>CAF</i>	Nigeria (centro)	Watson et al., 1997	112
83	<i>CAF</i>	Nigeria (sur)	Montano et al., 2013	129
84	<i>CAF</i>	Nigeria (sudeste)	Veeramah et al., 2010	1192
85	<i>CAF</i>	Nigeria (noreste)	Cerný et al., 2007	69
86	<i>CAF</i>	Chad	Cerný et al., 2007	127
87	<i>CAF</i>	Camerún (norte)	Cerný et al., 2007	244
88	<i>CAF</i>	Camerún (noroeste)	Veeramah et al., 2010	256
89	<i>CAF</i>	Camerún (oeste)	Quintana-Murci et al., 2008	199
90	<i>CAF</i>	São Tomé	Mateu et al., 1997; Trovoada et al., 2003	153
91	<i>CAF</i>	Gabón (oeste)	Quintana-Murci et al., 2008	158
92	<i>CAF</i>	Gabón (noroeste)	Quintana-Murci et al., 2008	50
93	<i>CAF</i>	Gabón (norte)	Quintana-Murci et al., 2008	66
94	<i>CAF</i>	Gabón (centro)	Quintana-Murci et al., 2008	146
95	<i>CAF</i>	Gabón (este)	Quintana-Murci et al., 2008	154
96	<i>CAF</i>	Gabón (sudeste)	Quintana-Murci et al., 2008	247
97	<i>CAF</i>	Gabón (sudoeste)	Quintana-Murci et al., 2008	52
98	<i>CAF</i>	Congo	Montano et al., 2013	101
99	<i>SWA</i>	Angola (Cabinda)	Beleza et al., 2005	110
100	<i>SWA</i>	Angola (sudoeste)	Coelho et al., 2009	365
101	<i>EAF</i>	Sudán	Soares et al., 2012	102
102	<i>EAF</i>	Sudán (sur)	Krings et al., 1999	76
103	<i>EAF</i>	Etiopía	Kivisild et al., 2004; Soares et al., 2012	347
104	<i>EAF</i>	Etiopía (este)	Boattini et al., 2013	167
105	<i>EAF</i>	Etiopía (Nyngatom)	Poloni et al., 2009	112
106	<i>EAF</i>	Kenia (Turkana)	Boattini et al., 2013	51
107	<i>EAF</i>	Kenia (norte)	Boattini et al., 2013	69
108	<i>EAF</i>	Kenia (Maasai)	Boattini et al., 2013	81
109	<i>EAF</i>	Kenia	Watson et al., 1997; Brandstätter et al., 2004	329
110	<i>EAF</i>	Somalia	Soares et al., 2012	148
111	<i>EAF</i>	Tanzania (norte)	Tishkoff et al., 2007	278
112	<i>EAF</i>	Tanzania (centro)	Knight et al., 2003	100
113	<i>SEA</i>	Zambia	de Filippo et al., 2010	78
114	<i>SEA</i>	Zimbawe	Castrì et al., 2009	59
115	<i>SEA</i>	Mozambique	Pereira et al., 2001; Salas et al., 2002	416
116	<i>SAF</i>	Sudáfrica (Schmidtsdrift)	Chen et al., 2000	74
117	<i>SAF</i>	Sudáfrica	Quintana-Murci et al., 2010	445
118	<i>MAD</i>	Madagascar	Tofanelli et al., 2009	133

Referencias

- Abu-Amero KK et al. (2008) *BMC Evol Biol* 8:45
- Achilli A et al. (2007) *Am J Hum Genet* 80:759–768.
- Alshamali F et al. (2008) *Forensic Sci Int Genet* 2:e9–e10
- Álvarez L et al. (2010) *Am J Phys Anthr* 142:531–539.
- Al-Zahery N et al. (2003) *Mol Phylogenet Evol* 28:458–472.
- Babalini C et al. (2005) *Eur J Hum Genet* 13:902–912.
- Badro DA et al. (2013) *PLoS One* 8:e54616.
- Bekada A et al. (2013) *PLoS One* 8:e56775.
- Beleza S et al. (2005) *Hum Genet* 117:366–375.
- Boattini A et al. (2013) *Am J Phys Anthr* 150:375–385.
- Brakez Z et al. (2001) *Ann Hum Biol* 28:295–307.
- Brandstätter A et al. (2004) *Int J Leg Med* 118:294–306.
- Brehm A et al. (2002) *Ann Hum Genet* 66:49–60.
- Brehm A et al. (2003) *Hum Genet* 114:77–86.
- Cardoso S et al. (2010) *J Forensic Sci* 55:1196–1201.
- Casas MJ et al. (2006) *Am J Phys Anthr* 131:539–551.
- Castrì L et al. (2009) *Am J Phys Anthr* 140:302–311.
- Cerný V et al. (2007) *Ann Hum Genet* 71:433–452.
- Cerný V et al. (2008) *Am J Phys Anthr* 136:128–137.
- Chen YS et al. (2000) *Am J Hum Genet* 66:1362–1383.
- Cherni L et al. (2005) *Hum Biol* 77:61–70.
- Cherni L et al. (2009) *Am J Phys Anthr* 139:253–260.
- Coelho M et al. (2009) *BMC Evol Biol* 9:80.
- Coudray C et al. (2009) *Ann Hum Genet* 73:196–214.
- Dubut V et al. (2004) *Eur J Hum Genet* 12:293–300.
- Ely B et al. (2006) *BMC Biol* 4:34.
- Fadhlaoui-Zid K et al. (2011) *Am J Phys Anthr* 145:107–117.
- Falchi A et al. (2006) *J Hum Genet* 51:9–14.
- De Filippo C et al. (2010) *Am J Phys Anthr* 141:382–394.
- González AM et al. (2003) *Am J Phys Anthr* 120:391–404.
- González AM et al. (2006) *Ann Hum Genet* 70:631–657.
- González AM et al. (2009) *Ann Hum Biol* 35:212–231.
- Harich N et al. (2010) *BMC Evol Biol* 10:138.
- Irwin J et al. (2007) *Int J Leg Med* 121:377–383.
- Irwin J et al. (2008) *Int J Leg Med* 122:87–89.
- Jackson BA et al. (2005) *Am J Phys Anthr* 128:156–163.
- Kivisild T et al. (2004) *Am J Hum Genet* 75:752–770.
- Knight A et al. (2003) *Curr Biol* 13:464–473.
- Krings M et al. (1999) *Am J Hum Genet* 64:1166–1176.
- Larruga JM et al. (2001) *Eur J Hum Genet* 9:708–716.
- Loueslati BY et al. (2006) *Am J Hum Biol* 18:149–153.
- Maca-Meyer N et al. (2003) *Ann Hum Genet* 67:329–339.
- Malyarchuk BA et al. (2003) *Ann Hum Genet* 67:412–425.
- Malyarchuk BA et al. (2008) *Ann Hum Genet* 72:228–240.
- Mateu E et al. (1997) *Ann Hum Genet* 61:507–518.

Metspalu M et al. (2004) *BMC Genet* 5:26.
Montano V et al. (2013) *BMC Evol Biol* 13:24.
Non AL et al. (2011) *Am J Phys Anthr* 144:1–10.
Ottoni C et al. (2009a) *Ann Hum Biol* 36:785–811.
Ottoni C et al. (2009b) *Ann Hum Genet* 73:438–448.
Pereira L et al. (2001) *Ann Hum Genet* 65:439–458.
Pereira L et al. (2004) *Int J Leg Med* 118:132–136.
Picornell A et al. (2005) *Am J Phys Anthr* 128:119–130.
Plaza S et al. (2003) *Ann Hum Genet* 67:312–328.
Poloni ES et al. (2009) *Ann Hum Genet* 73:582–600.
Quintana-Murci L et al. (2008) *Proc Natl Acad Sci U S A* 105:1596–1601.
Quintana-Murci L et al. (2010) *Am J Hum Genet* 86:611–620.
Rando JC et al. (1998) *Ann Hum Genet* 62:531–550.
Rando JC et al. (1999) *Ann Hum Genet* 63:413–428.
Rhouda T et al. (2009) *Mitochondrion* 9:402–407.
Richards M et al. (2000) *Am J Hum Genet* 67:1251–1276.
Salas A et al. (2002) *Am J Hum Genet* 71:1082–1111.
Santos C et al. (2003) *Ann Hum Genet* 67:433–456.
Saunier JL et al. (2009) *Forensic Sci Int Genet* 3:e97–e103.
Shlush LI et al. (2008) *PLoS One* 3:e2105.
Soares P et al. (2012) *Mol Biol Evol* 29:915–927.
Stefflova K et al. (2009) *PLoS One* 4:e7842.
Tishkoff SA et al. (2007) *Mol Biol Evol* 24:2180–2195.
Tofanelli S et al. (2009) *Mol Biol Evol* 26:2109–2124.
Trovoada MJ et al. (2003) *Ann Hum Genet* 68:40–54.
Turchi C et al. (2008) *Int J Leg Med* 122:199–204.
Turchi C et al. (2009) *Forensic Sci Int Genet* 3:166–172.
Veeramah KR et al. (2010) *BMC Evol Biol* 10:92.
Watson E et al. (1997) *Am J Hum Genet* 61:691–704.

Tabla S5. Haplotipos mitocondriales completos de las muestras de Andalucía y Marruecos pertenecientes a linajes africanos. Los haplogrupos se indican tal y como se muestran en sus respectivas filogenias completas (**Datasets S2-S6**). El cambio de base nucleotídica solo se muestra en transversiones e *indels*. También se presenta el número de referencia de GenBank.

-----**VER CD adjunto**-----

Tabla S6. Estimaciones de edades de coalescencia de los sub-linajes más comunes dentro de los clados U6, M1 y L.

Haplogrupo	Estima de edad por ρ (genoma completo)	Estima de edad por ρ (sustituciones sinónimas)	Estima de edad por Máxima Verosimilitud
U6	35352 [24682 - 46435]	34529 [20161 - 48897]	35454 [23982 - 47404]
U6a	25944 [20350 - 31670]	32121 [21535 - 42707]	25392 [20213 - 30685]
U6b	12472 [7416 - 17670]	7884 [3430 - 12337]	11328 [7724 - 15005]
U6d	13206 [7035 - 19586]	14376 [3431 - 25322]	12034 [6940 - 17272]
U6c	10820 [5286 - 16531]	9096 [0 - 18380]	9924 [5013 - 14976]
M1	26280 [17939 - 34916]	26902 [12069 - 41735]	26137 [18651 - 33861]
M1a	21224 [15123 - 27497]	19493 [10209 - 28777]	20996 [16503 - 25582]
M1b	20259 [12644 - 28150]	17139 [8225 - 26052]	22416 [14810 - 30287]
L1	122789 [98301 - 147841]	124149 [89583 - 158715]	138413 [103499 - 174245]
L1b	29278 [15961 - 43323]	37084 [11807 - 62360]	35757 [17744 - 54974]
L1c	91163 [73724 - 109057]	93524 [70066 - 116982]	87211 [70058 - 104831]
L2	77931 [57873 - 98728]	85997 [51183 - 120810]	96869 [77126 - 117149]
L2a	58685 [37826 - 80637]	80119 [39501 - 120738]	79306 [59869 - 99423]
L2b	29383 [19627 - 39522]	29839 [14528 - 45150]	26271 [20284 - 32409]
L2c	17261 [13993 - 20582]	15768 [11911 - 19624]	20291 [16429 - 24223]
L2d	18261 [10105 - 26748]	23237 [5831 - 40642]	14347 [7984 - 20926]
L2e	35805 [26454 - 45467]	45333 [28298 - 62367]	35929 [26074 - 46131]
L3	54012 [45043 - 63190]	58241 [44289 - 72192]	68179 [56232 - 80425]
L3a	48295 [34860 - 62255]	57816 [34780 - 80851]	56442 [42270 - 71127]
L3b	19373 [12383 - 26599]	28289 [9423 - 47156]	23719 [14005 - 33858]
L3d	30640 [22589 - 38944]	36553 [24576 - 48529]	33296 [24372 - 42518]
L3c	9245 [2343 - 16433]	15768 [315 - 31220]	10108 [2128 - 18467]
L3f	48466 [33507 - 64079]	43234 [22990 - 63478]	53297 [40936 - 66065]
L3e	33669 [25738 - 41832]	35551 [22994 - 48108]	39970 [29738 - 50550]
L3i	33538 [21464 - 46161]	46091 [23412 - 68769]	45326 [30929 - 60361]
L3k	27266 [13741 - 41573]	30222 [8520 - 51923]	27260 [15724 - 39359]
L3x	36534 [26272 - 47170]	33037 [18769 - 47306]	38823 [28227 - 49801]
L3h	72146 [54007 - 90944]	76496 [48708 - 104284]	64029 [52049 - 76330]

Estimaciones de edad evolutiva expresadas en años con intervalos de confianza (95%).

Figura S2. Localización geográfica de las muestras poblacionales que han sido usadas en los mapas de contorno (*Surface Maps*) para observar variaciones espaciales de los linajes mitocondriales africanos. Los códigos y referencias se detallan en la **Tabla S4**.



Figura S3. Distribución espacial del haplogrupo U6 a gran escala (A) y en la cuenca Mediterránea (B).

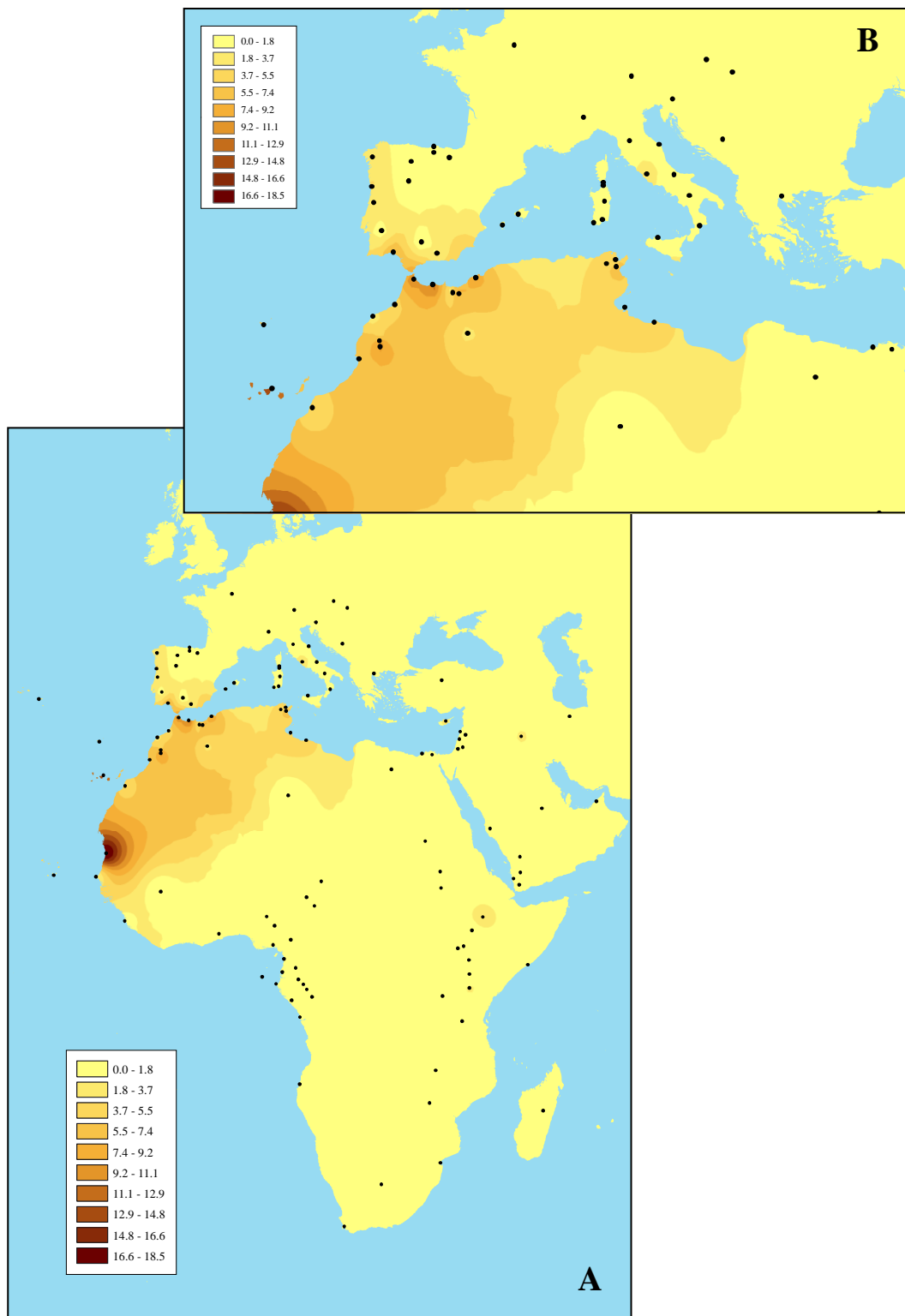


Figura S4. Distribución espacial del haplogrupo M1 a gran escala (A) y en la cuenca Mediterránea (B).

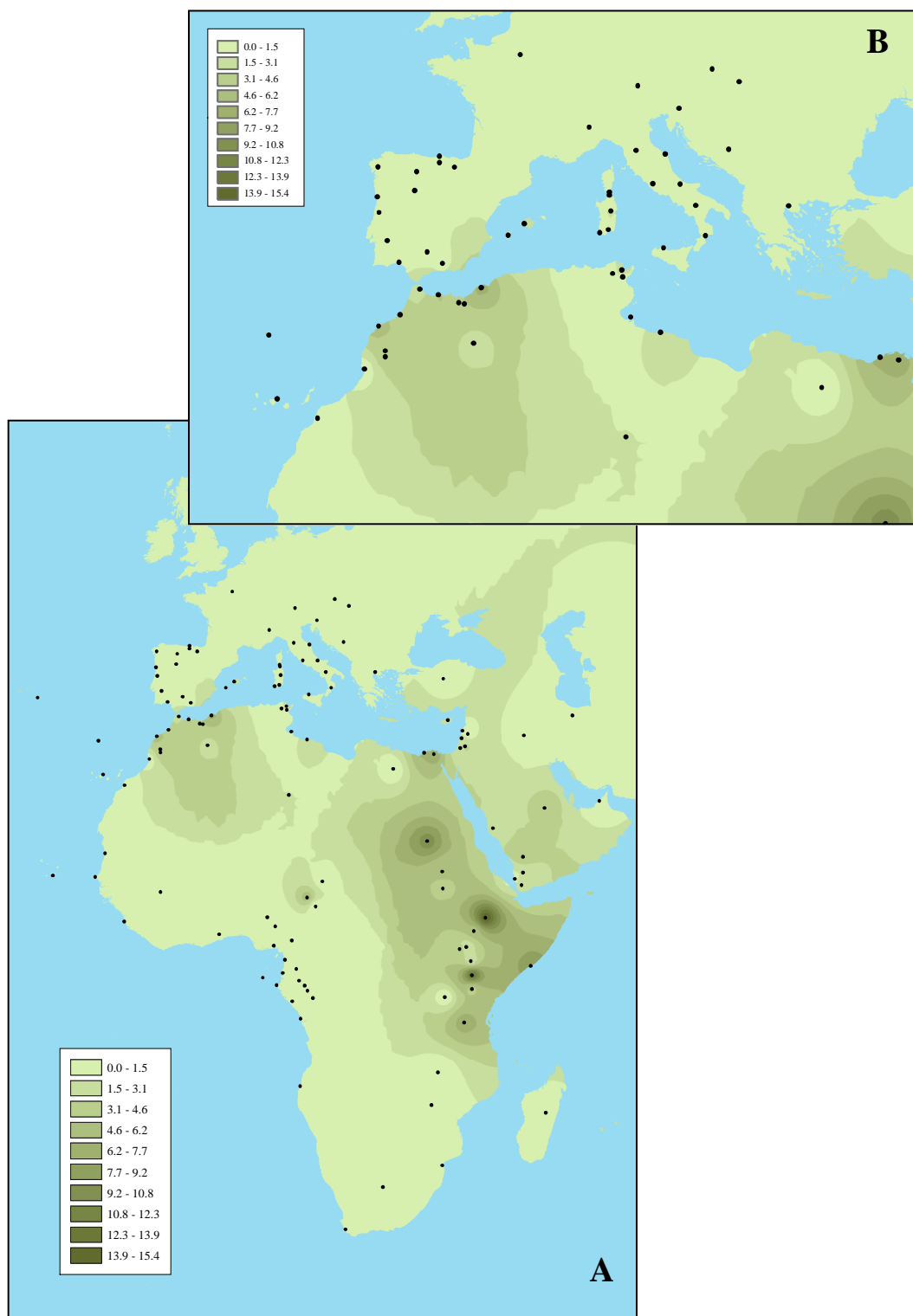


Figura S5. Distribución espacial del haplogrupo L2a a gran escala (**A**) y en la cuenca Mediterránea (**B**).

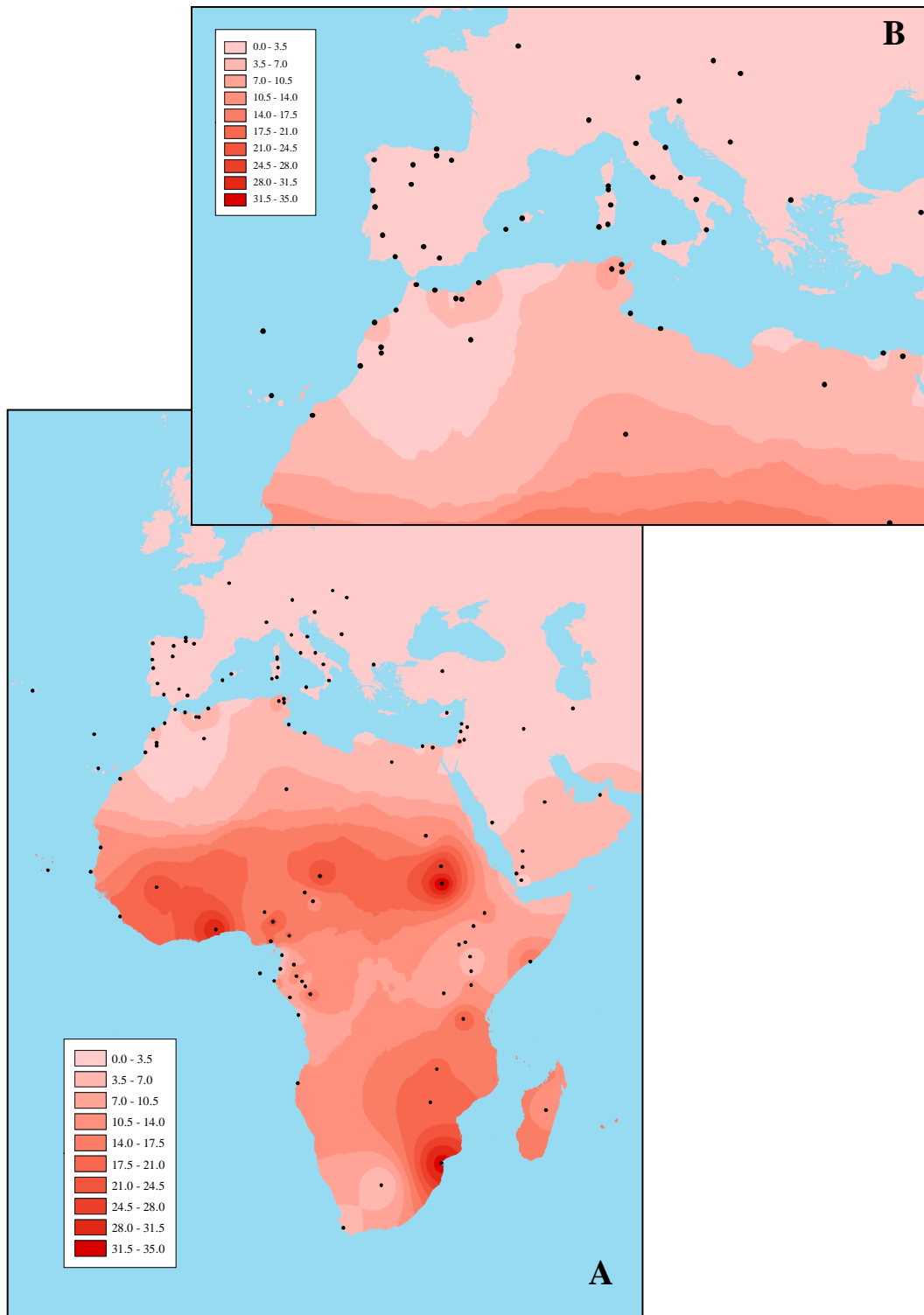


Figura S6. Distribución espacial del haplogrupo L2b a gran escala (A) y en la cuenca Mediterránea (B).

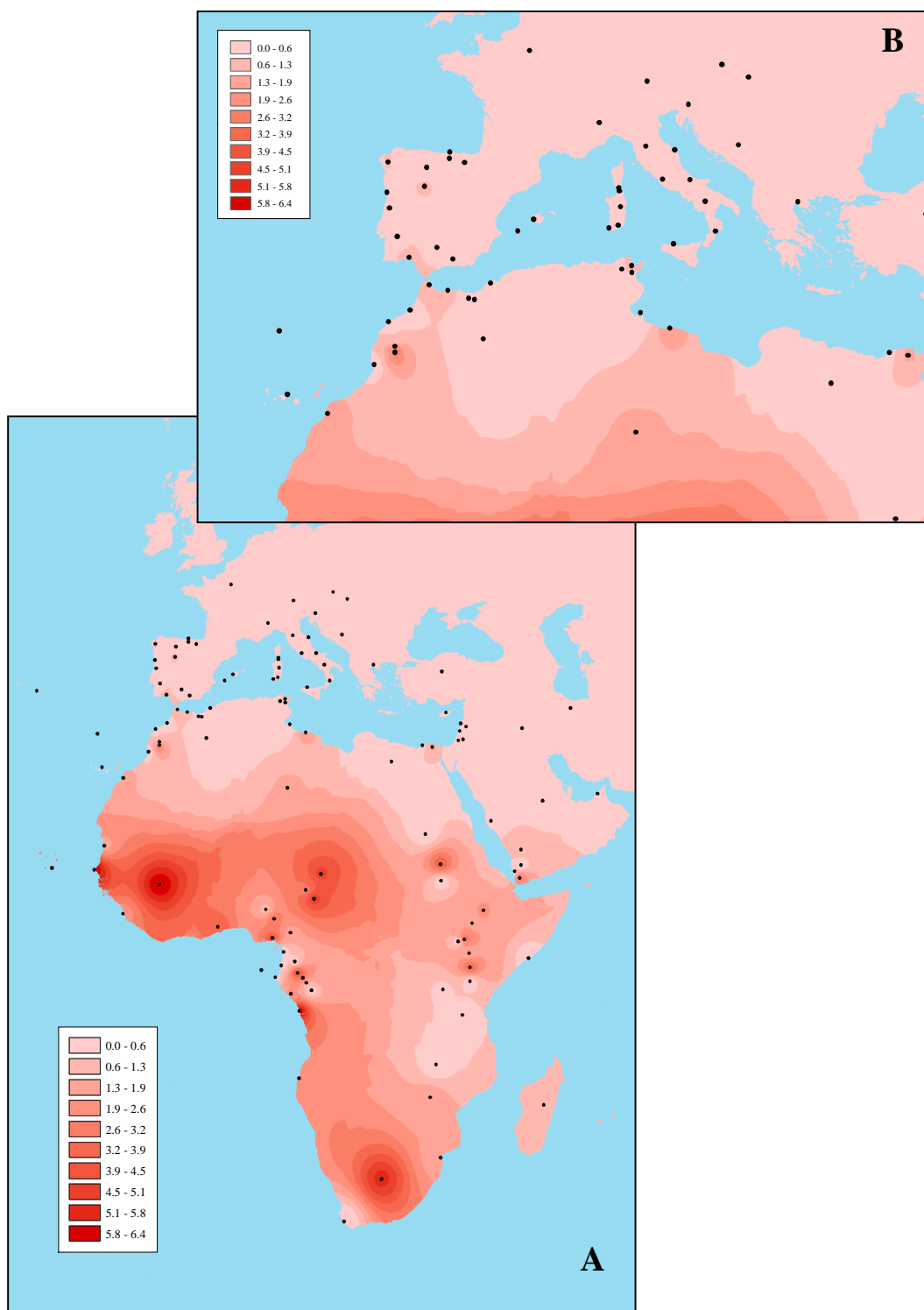


Figura S7. Distribución espacial del haplogrupo L3b a gran escala (A) y en la cuenca Mediterránea (B).

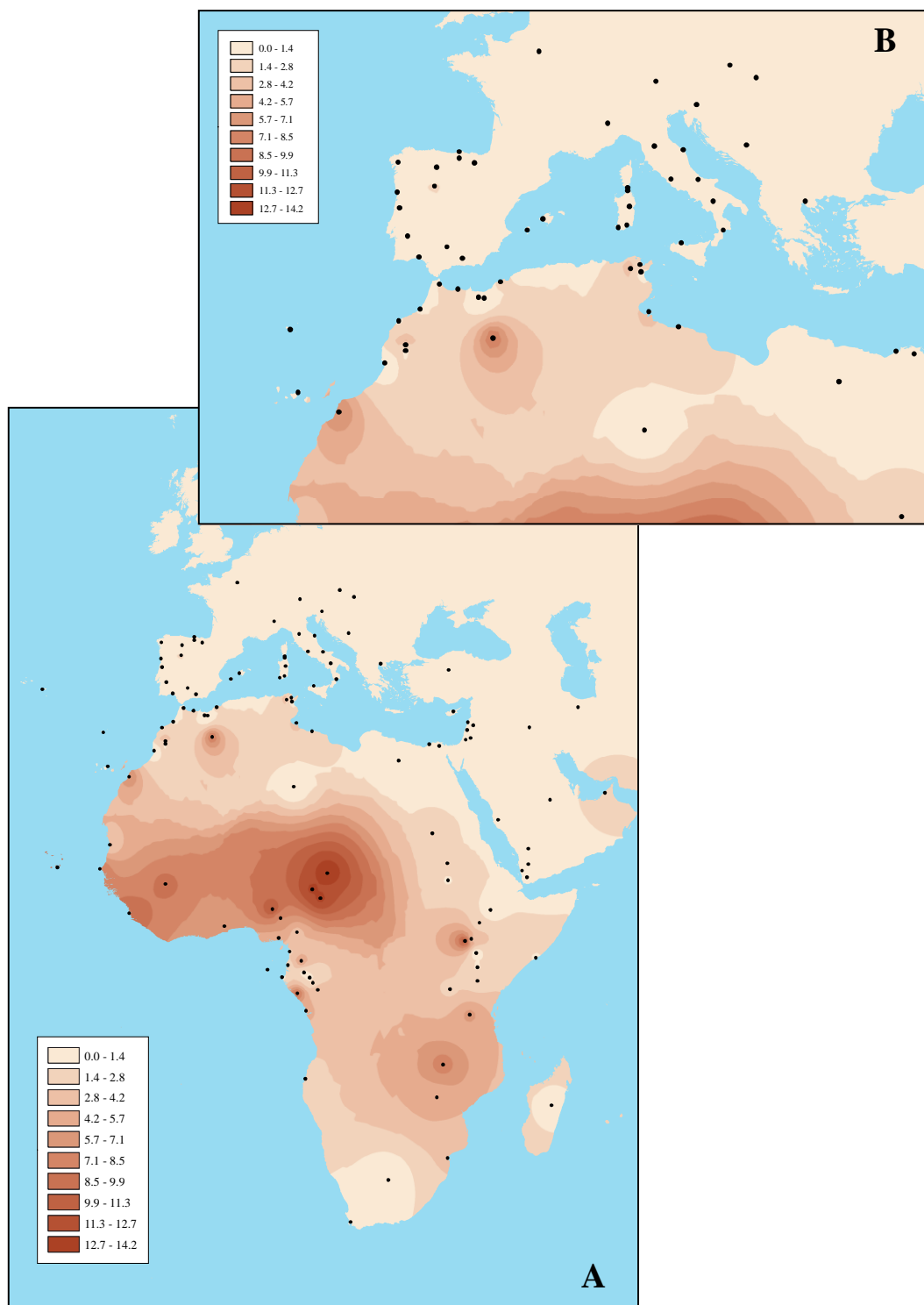


Figura S8. Distribución espacial del haplogrupo L3d a gran escala (A) y en la cuenca Mediterránea (B).

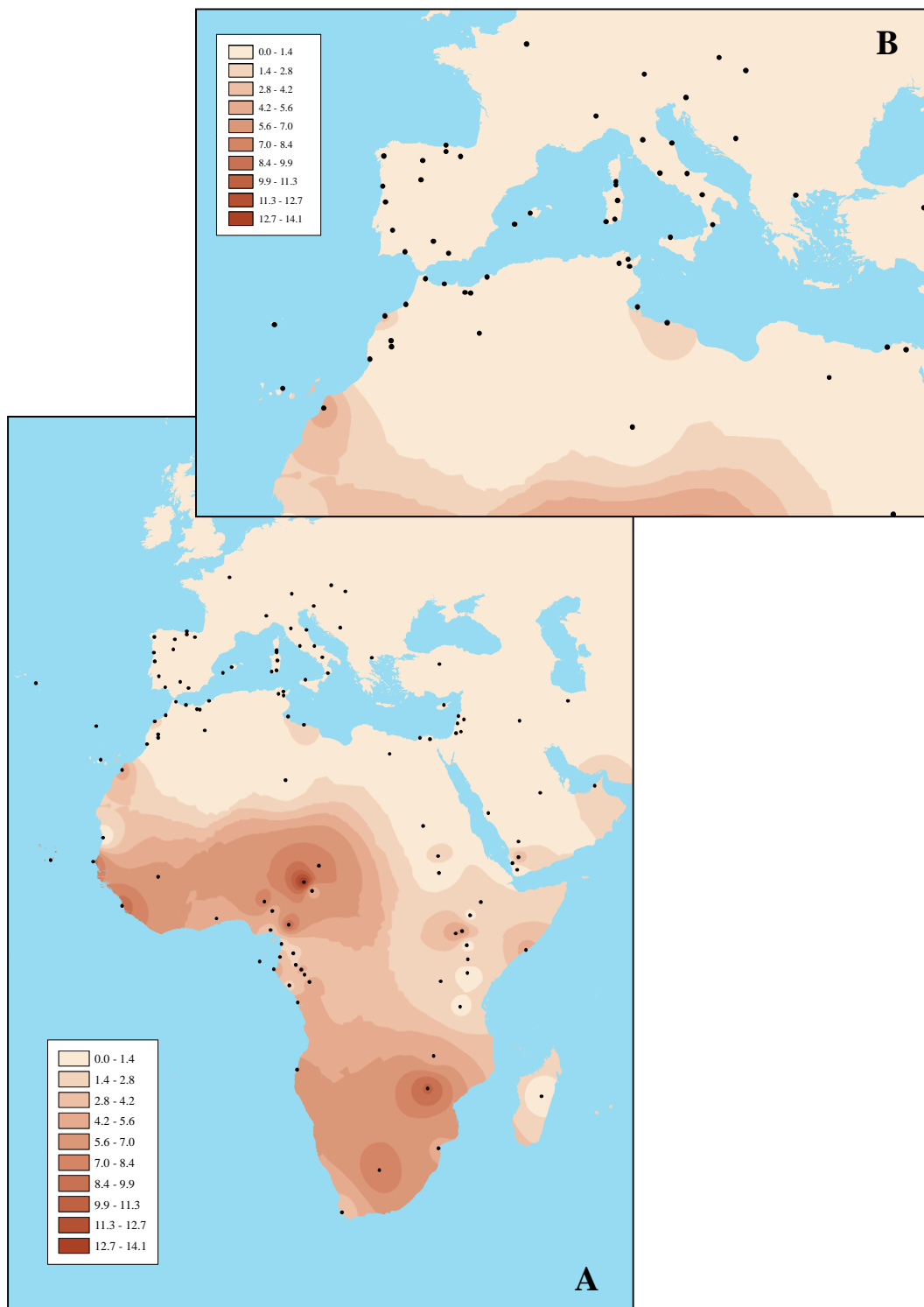


Figura S9. Distribución espacial del haplogrupo L3f a gran escala (A) y en la cuenca Mediterránea (B).

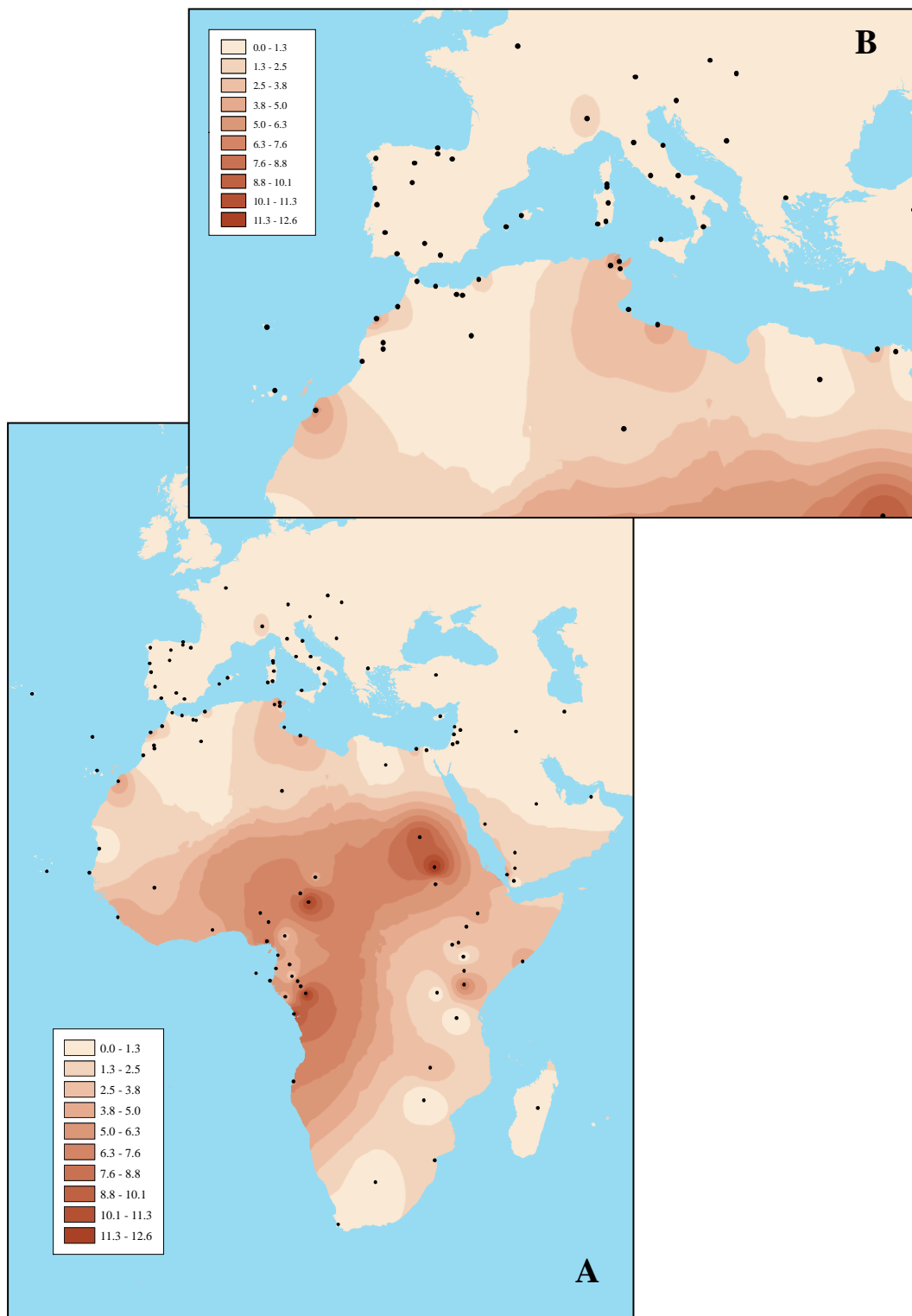


Figura S10. Distribución espacial del haplogrupo L3h1b a gran escala (**A**) y en la cuenca Mediterránea (**B**).

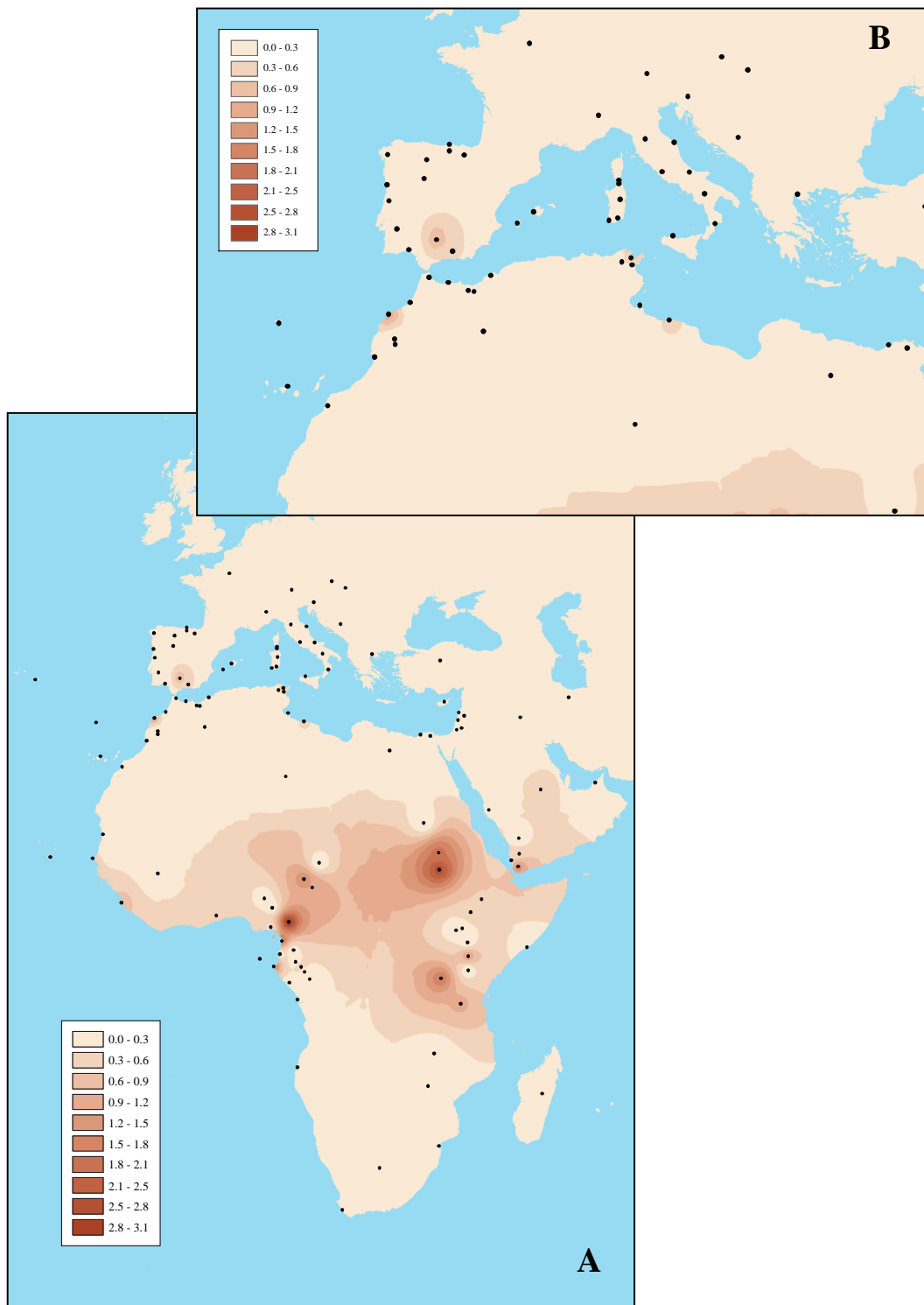


Figura S11. Relaciones filogenéticas de los linajes africanos U6a (A) y L1b (B) construidos a partir de 209 y 175 secuencias HVS-I, respectivamente. Los asteriscos (*) indican los nodos basales con las siguientes mutaciones con respecto a la rCRS: 16172-16219-16278 (U6a) y 16126-16187-16189-16223-16264-16270-16278-16293-16311 (L1b). El tamaño de los círculos es proporcional a la frecuencia del haplotipo. Los códigos y referencias están registrados en la **Tabla S4**. El método filogenético usado ha sido el *Median-Joining Network* (MJN).

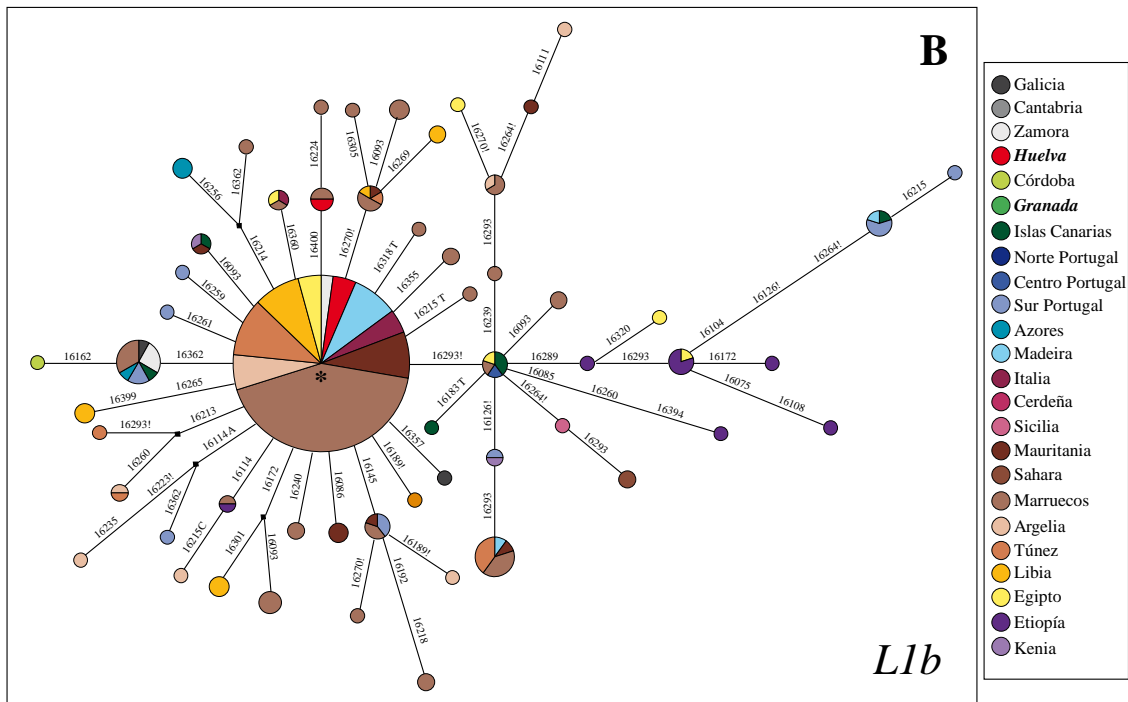
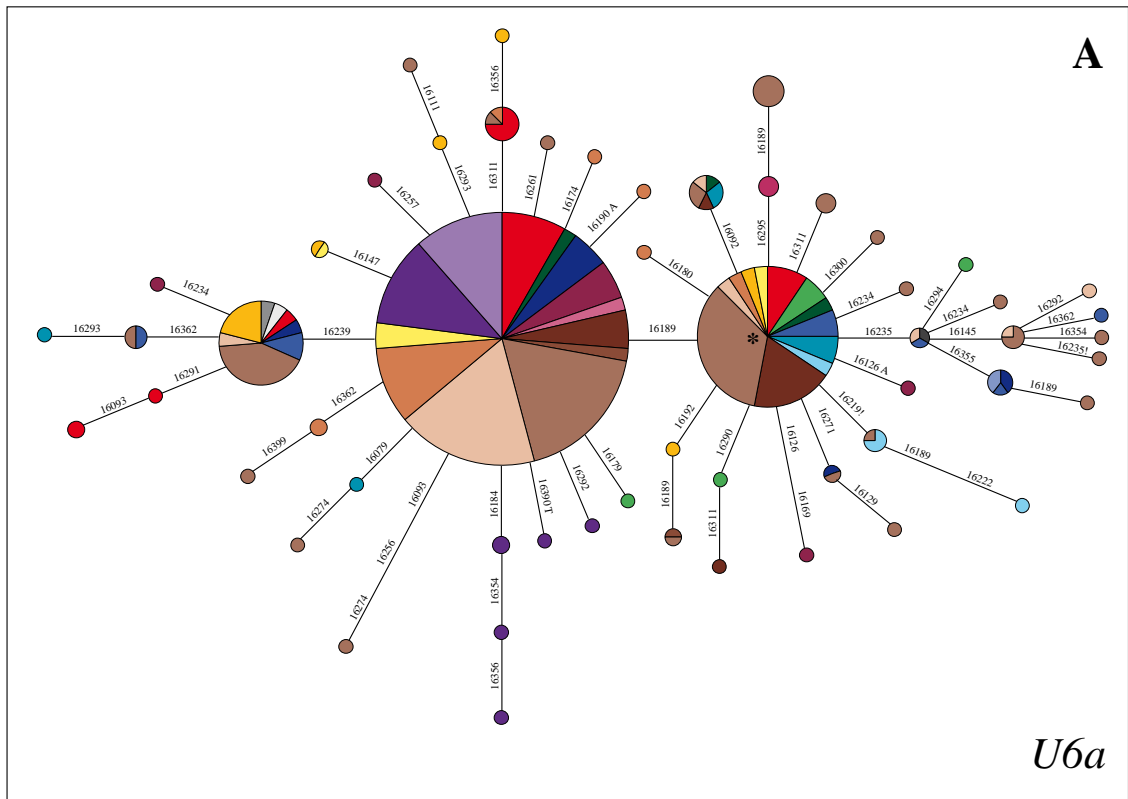
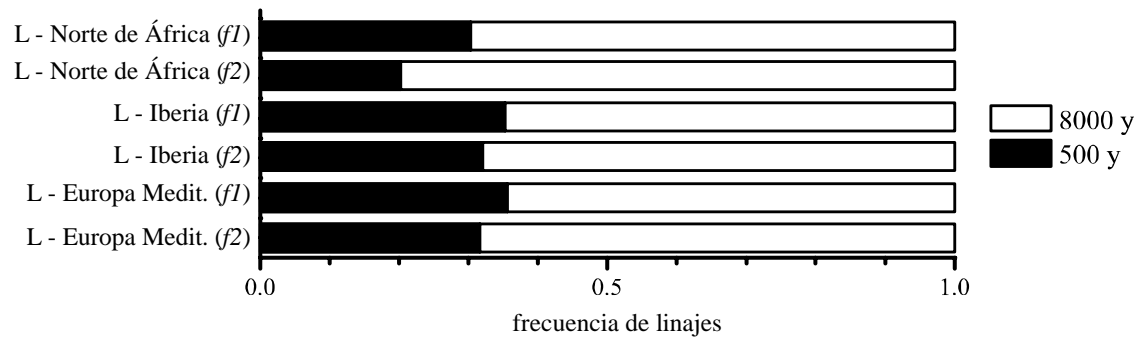


Figura S12. Proporción probabilística de los *clusters* fundadores considerando dos eventos migratorios (a 0,5 y 8,0 ky) –usando los criterios *f1* y *f2*– para la migración de los linajes L hacia el norte de África, Iberia y Europa mediterránea.



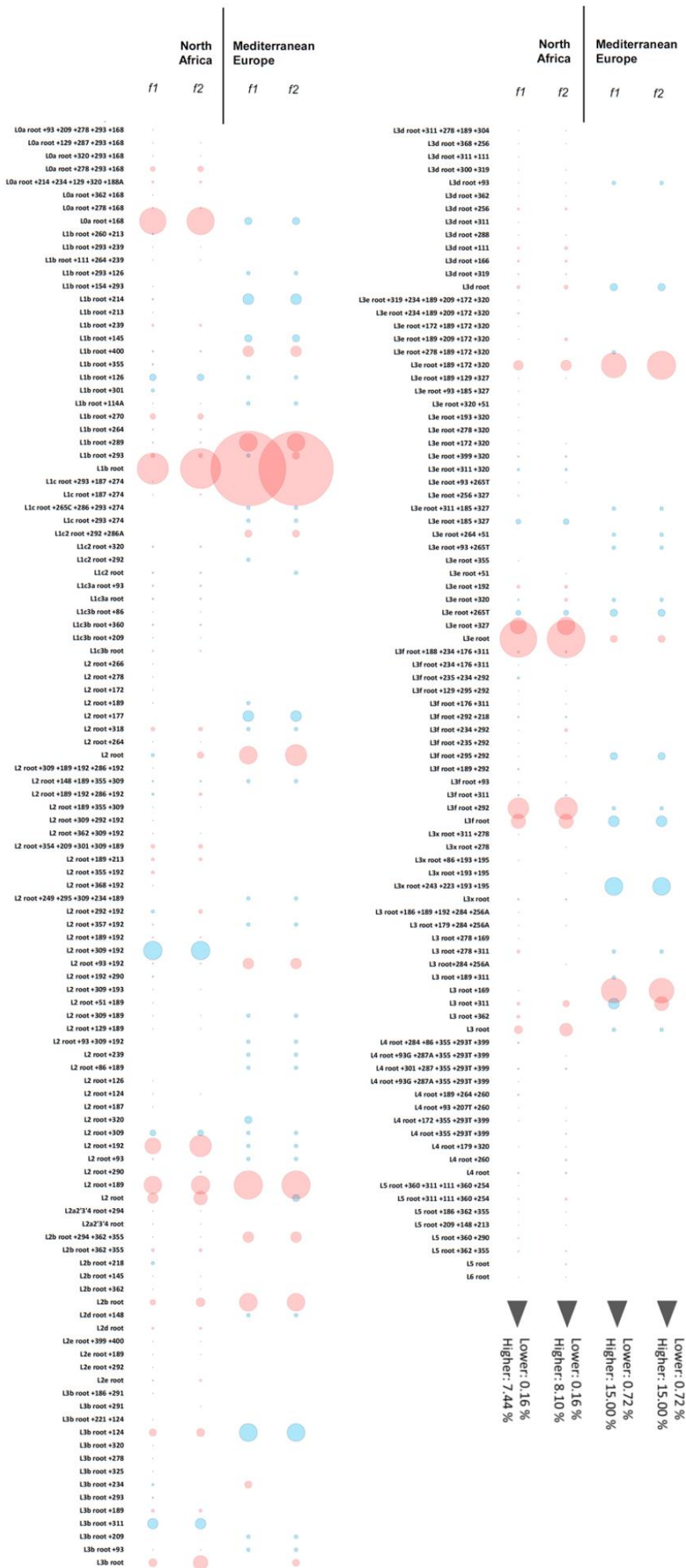


Figura S13. Fundadores del haplogrupo L introducidos en el norte de África y Europa mediterránea. Los círculos rojos indican aquellos clados probablemente introducidos durante el Holoceno temprano, y los círculos azules, los que llegaron durante tiempos históricos. El tamaño de los círculos es proporcional a la frecuencia de cada fundador dentro del total de fundadores L, en los análisis para las dos regiones, con los dos criterios de selección de fundadores (**f1* y *f2*). En la Figura se indican las frecuencias superiores e inferiores para cada análisis.

* Los fundadores tienen que tener al menos una (*f1*) o dos (*f2*) ramas derivadas en la población fuente.

Dataset S1. Base de datos de 2.182 mitogenomas para la construcción de las filogenias de haplogrupos africanos.

Dataset S2. Filogenia del haplogrupo U6 construida usando 266 mitogenomas.

Dataset S3. Filogenia del haplogrupo M1 construida usando 114 mitogenomas.

Dataset S4. Filogenia del haplogrupo L1 construida usando 422 mitogenomas.

Dataset S5. Filogenia del haplogrupo L2 construida usando 706 mitogenomas.

Dataset S6. Filogenia del haplogrupo L3 construida usando 674 mitogenomas.

-----**VER CD adjunto**-----

ANEXO III (CAPÍTULO V)

Tabla S7. Diseño de los ensayos TaqMan para el genotipado de los linajes H y H1.

Posición	Haplogrupo		Secuencia (5'-3')	Concentración final (μ M)	
7028	H	<i>primers</i>	6988F	CACTAGACATCGTACTACACGACACGTA	0,9
			7070R	GATGGCAAATACAGCTCCTATTGA	0,9
		<i>sondas</i>	alelo C	VIC -TACGTTGTAGCCCACTT	0,3
			alelo T	FAM -CGTTGTAGCTCACTTC	0,1
3010	H1	<i>primers</i>	2981F	ACGACCTCGATGTTGGATCAG	0,9
			3062R	ACGTAGGACTTTAATCGTTGAACAAA	0,9
		<i>sondas</i>	alelo A	VIC -CATCCCAATGGTGC	0,3
			alelo G	FAM -ACATCCCGATGGTG	0,1

Tabla S8. Diseño de las reacciones *multiplex* por SNaPshot para el genotipado de 21 SNPs internos al haplogrupo H. Las condiciones y las secuencias de los *primers* y sondas de minisequenciación se tomaron de Álvarez-Iglesias et al., 2009.

SNP	Cambio de base	Sub-Hg	Primers (F y R)	Tamaño (pb)	Conc. primer (μ M)	Sonda de minisequenciación	Tamaño (pb)	Cadena	Conc. primer (μ M)
6365	T/C	H1a1	TGACTCTTACCTCCCTCTCTCC	189	0,2	(gact) ₇ gaGATGGCCCTAAGATAGAGGAGAC	54	H	0,2
			GATGAAATTGATGGCCCTA						
8271T	A/T	H1a2	TAGGCCCTATTAGCCCTAT	110	0,2	(gact) ₁₅ GGGCCCGTATTTACCCCTATAGC	82	L	0,3
			AAGAGGTGTGGTCTCTTAATCTTT						
9150	A/G	H1c1	CCTACTATGCACCTAATTGGA	155	0,3	(gact) ₁₂ gCCTAGAAATCGGTGTCGCCCTT	70	L	0,2
			GGCTTACTAGAAGTGA AACCGTA						
12858	C/T	H1c2	CAACACAGCAGCCATTCAAG	157	0,2	(gact) ₈ gacGCAGCCATTCAAGCAATCCTATA	58	L	0,3
			GAGGCTAGTAGTGGGTGA						
8473	T/C	H1c3	CCCAACTAAAAATATAAACACACAAACT	193	0,2	(gact) ₆ gAAAAATATTAACACAAACTACCACCTACC	56	L	0,2
			GGAGTGGGATCAATAGAG						
9066	A/G	H1f	CCTACTATGCACCTAATTGGA	155	0,3	(gact) ₈ gaGTGTAGAGGGAAGTTAATGGTTGATAT	62	H	0,3
			GGCTTACTAGAAGTGA AACCGTA						
8602	T/C	H1g1	CCCAACTAAAAATATAAACACACAAACT	193	0,2	(gact) ₁₃ CGGCCACTACTGATCATTTCTA	74	L	0,4
			GGAGTGGGATCAATAGAG						

MULTIPLEX 1 (H1)

SNP	Cambio de base	Sub-Hg	Primers (F y R)	Tamaño (pb)	Conc. primer (µM)	Sonda de minisequenciación	Tamaño (pb)	Cadena	Conc. primer (µM)
4769	G/A	H2a	TCCTTCTAATAGTATCCTCTTCAACA TGGGTAACCTCTGGGACTCA	154	0,4	(gact) ₈ gaACCAATCAATCACTCATCAATTAAATCATAAT	50	L	0,2
951	G/A	H2a1	TCACACGATTAACCCAAAGTCA ACTCAGGTGAGTTTATGCTTTATTG	87	0,2	(gact) ₁₁ gacCTTTTATTTGGGGAGGGGTGAT	21	H	0,2
750	G/A	H2a2	GGCTCACATCACCCCATAAA CGTTTTGAGCTGCATTGCT	161	0,2	(gact) ₁₅ CTCTAAATCACCCAGCATCAAAAGG*	84*	L	0,2
13708	G/A	H2a5	AACGAAATAAACCCACCCTA GTTGTTTTGAAAGGGGATG	113	0,2	(gact) ₁₃ gaCTACTAAACCCCATTTAAAGGCCCTG	78	L	0,3
6776	T/C	H3	GCTTCCTAGGTTTTATCGTGTG GAGTGTGGCAGTCAAGTAAA	140	0,4	(gact) ₆ gaCCTGTCTACGTCTATTTCTACTCTGTAATPAT	58	H	0,3
12957	T/C	H3c	CAACACAGCAGCCATTCAAG GAGGCTAGTAGTGGGGTGA	157	0,2	(gact) ₉ CAACAAATAGCCCTTCTAAACGCCTAA	62	L	0,3
4336	T/C	H5a	AGCATTCCTCCTCAACCTA TTTTGGATTCTCAGGGATGG	127	0,4	(gact) ₁₂ gacAGGGATGGGTTTCGATTCTCAT*	72*	H	0,2
3915	G/A	H6a	TAGCAGAGACCAACCGAACCC GAAGATTGTAGTGTGAGGGTGT	158	0,6	(gact) ₁₆ GAAGCCTGAGACTAGTTCGGGACTC	32	H	0,2
3992	C/T	H4	TAGCAGAGACCAACCGAACCC GAAGATTGTAGTGTGAGGGTGT	158	0,6	(gact) ₁₆ gaCCCTATTCTTCATAGCCGAATACA	38	L	0,3
14365	C/T	H4a	CCACCCATCATACTCTTTCA TAGGGGAATGATGTTGTCTC	159	0,3	(gact) ₆ GTTAGCGATGGAGGTAGGATTGGT*	48*	H	0,3
4793	A/G	H7	TCCTTCTAATAGTATCCTCTTCAACA TGGGTAACCTCTGGGACTCA	154	0,4	(gact) ₈ gATAATCATAAATAGCTATAGCAATAAAACTAGGAAT	56	L	0,4
13101C	A/C	H8	CAGCCCTACTCCACTCAAGC TGGGCTATTTCTGCTAGGG	83	0,2	(gact) ₁₄ gaGAAGCGGATGAGTAAGAAGATTCC*	82*	H	0,2
13759	G/A	H11	AACGAAATAAACCCACCCTA GTTGTTTTGAAAGGGGATG	113	0,2	(gact) ₁₀ TTCTCATTTACTAACCAACATTTCCCCCC	66	L	0,1
14872	C/T	H13	CAACATCTCCGCATGATGAA AGGCTCTGGTGTGAGTAGTGC	104	0,2	(gact) ₁₃ gaCTCCTTGGCCGCTGCCTGAT*	74*	L	0,4

*sonda modificada con respecto a Alvarez-Iglesias et al. (2009) PLoS One 4:e5112.

MULTIPLEX 2 (H*)

MULTIPLEX 3 (H*)

Tabla S9. Conjunto de poblaciones humanas seleccionadas para el análisis comparativo del linaje mitocondrial H. Las regiones geográficas establecidas son: *WME* (Europa mediterránea occidental), *WEU* (Europa occidental), *CEU* (Europa central), *NEU* (noreste de Europa), *EEU* (Europa oriental), *CME* (Europa mediterránea central), *CAU* (Cáucaso), *ANA* (Anatolia), *NES* (Oriente Próximo), *ARA* (Arabia), *NAF* (norte de África), *WAF* (África occidental). Se destacan subrayados aquellos estudios que alcanzan un alto nivel de resolución dentro del clado H (a través de la definición de sub-clados por genotipado de SNPs). Se indica en tamaño total de la población (N) y el número de individuos clasificados como H (N_H).

Código	Región	Población	Referencia	N	N _H
<u>1</u>	<i>WME</i>	<u>España (Galicia)</u>	Álvarez-Iglesias et al., 2009	282	124
<u>2</u>	<i>WME</i>	<u>España (Asturias)</u>	García et al., 2011	76	39
<u>3</u>	<i>WME</i>	<u>España (León)</u>	Larruga et al., 2001; Ennafaa et al., 2009	61	36
<u>4</u>	<i>WME</i>	<u>España (Cantabria)</u>	Álvarez-Iglesias et al., 2009	135	52
<u>5</u>	<i>WME</i>	<u>España País Vasco (Vizcaya)</u>	García et al., 2011; Behar et al., 2012	150	71
<u>6</u>	<i>WME</i>	<u>España País Vasco (Guipúzcoa)</u>	García et al., 2011; Behar et al., 2012	169	93
<u>7</u>	<i>WME</i>	<u>España País Vasco (SW Guipúzcoa)</u>	Behar et al., 2012	63	34
<u>8</u>	<i>WME</i>	<u>España País Vasco</u>	Pereira et al., 2005; Behar et al., 2012	257	128
<u>9</u>	<i>WME</i>	<u>España (N Navarra)</u>	Cardoso et al., 2011	110	38
<u>10</u>	<i>WME</i>	<u>España (NW Navarra)</u>	Behar et al., 2012	53	27
<u>11</u>	<i>WME</i>	<u>España (CW Navarra)</u>	Behar et al., 2012	64	31
<u>12</u>	<i>WME</i>	<u>España (NE Navarra)</u>	Behar et al., 2012	55	20
<u>13</u>	<i>WME</i>	<u>España (La Rioja)</u>	Behar et al., 2012	52	30
<u>14</u>	<i>WME</i>	<u>España (Zamora)</u>	Álvarez et al., 2010	214	36
<u>15</u>	<i>WME</i>	<u>España (Madrid)</u>	Pereira et al., 2005	119	58
<u>16</u>	<i>WME</i>	<u>España (Cataluña)</u>	Álvarez-Iglesias et al., 2009	101	39
<u>17</u>	<i>WME</i>	<u>España Andalucía (Huelva)</u>	<i>presente estudio</i>	280	110
<u>18</u>	<i>WME</i>	<u>España Andalucía (Córdoba)</u>	Casas et al., 2006	108	35
<u>19</u>	<i>WME</i>	<u>España Andalucía (Granada)</u>	<i>presente estudio</i>	470	227
<u>20</u>	<i>WME</i>	<u>España (Islas Canarias)</u>	Santos et al. 2010	503	173
<u>21</u>	<i>WME</i>	<u>España (Ibiza)</u>	Picornell et al., 2005	50	19
<u>22</u>	<i>WME</i>	<u>España (Mallorca)</u>	Picornell et al., 2005	93	23
<u>23</u>	<i>WME</i>	<u>Portugal</u>	Pereira et al., 2004, 2005	549	225
<u>24</u>	<i>WME</i>	<u>Portugal (Azores)</u>	Santos et al., 2010	120	45
<u>25</u>	<i>WME</i>	<u>Portugal (Madeira)</u>	Brehm et al., 2003	155	56
<u>26</u>	<i>WEU</i>	<u>Francia (SW, Lapurdi/Baztan)</u>	Behar et al., 2012	58	25
<u>27</u>	<i>WEU</i>	<u>Francia (SW, Chalosse)</u>	Behar et al., 2012	60	26
<u>28</u>	<i>WEU</i>	<u>Francia (SW, Lapurdi Navarra)</u>	Behar et al., 2012	73	35
<u>29</u>	<i>WEU</i>	<u>Francia (SW, Zuberoa)</u>	Behar et al., 2012	62	22
<u>30</u>	<i>WEU</i>	<u>Francia (SW, Bearn)</u>	Behar et al., 2012	56	20
<u>31</u>	<i>WEU</i>	<u>Francia</u>	Richards et al., 2000; Pereira et al., 2005	72	26
<u>32</u>	<i>WEU</i>	<u>Irlanda</u>	Richards et al., 2000; Pereira et al., 2005	102	43
<u>33</u>	<i>CEU</i>	<u>Alemania</u>	Brandstätter et al., 2008	100	28
<u>34</u>	<i>CEU</i>	<u>Republica Checa</u>	Richards et al., 2000; Pereira et al., 2005; Malyarchuk et al., 2006	268	114

Código	Región	Población	Referencia	N	N_H
<u>35</u>	<i>CEU</i>	<u>Austria</u>	Brandstätter et al., 2008	2487	964
<u>36</u>	<i>CEU</i>	<u>Hungría (judíos Ashkenazi)</u>	Brandstätter et al., 2008	173	55
<u>37</u>	<i>NEU</i>	<u>Finlandia</u>	Finnilä et al., 2001; Pereira et al., 2005	480	188
<u>38</u>	<i>NEU</i>	<u>Rusia</u>	Richards et al., 2000; Pereira et al., 2005	57	18
<u>39</u>	<i>EEU</i>	<u>Rumania (Transilvania)</u>	Brandstätter et al., 2008	360	100
<u>40</u>	<i>EEU</i>	<u>Rumania</u>	Richards et al., 2000; Pereira et al., 2005	94	32
<u>41</u>	<i>EEU</i>	<u>Bulgaria</u>	Richards et al., 2000; Pereira et al., 2005; Karachanak et al., 2012	966	403
<u>42</u>	<i>EEU</i>	<u>Macedonia</u>	Brandstätter et al., 2008	200	88
<u>43</u>	<i>CME</i>	Italia (Torino)	Turchi et al., 2008	50	21
<u>44</u>	<i>CME</i>	<u>Italia (NE)</u>	Boattini et al., 2013	108	42
<u>45</u>	<i>CME</i>	<u>Italia (Udine)</u>	Brisighelli et al., 2012	51	27
<u>46</u>	<i>CME</i>	<u>Italia (NW)</u>	Boattini et al., 2013	162	73
<u>47</u>	<i>CME</i>	<u>Italia (Liguri, La Spezia)</u>	Brisighelli et al., 2012	50	22
<u>48</u>	<i>CME</i>	<u>Italia (CW)</u>	Boattini et al., 2013	52	29
<u>49</u>	<i>CME</i>	<u>Italia (Piceni)</u>	Brisighelli et al., 2012	53	26
<u>50</u>	<i>CME</i>	<u>Italia (C)</u>	Boattini et al., 2013	110	50
<u>51</u>	<i>CME</i>	<u>Italia</u>	Richards et al., 2000; Achilli et al., 2004; Pereira et al., 2005	157	54
<u>52</u>	<i>CME</i>	<u>Italia (Lucera, Foggia)</u>	Brisighelli et al., 2012	60	25
<u>53</u>	<i>CME</i>	<u>Italia (S. Benevento)</u>	Brisighelli et al., 2012; Boattini et al., 2013	260	104
<u>54</u>	<i>CME</i>	Italia (Basilicata)	Ottoni et al., 2009	91	33
<u>55</u>	<i>CME</i>	<u>Italia (Lecce, Salento)</u>	Brisighelli et al., 2012	53	13
<u>56</u>	<i>CME</i>	<u>Italia (Belvedere, Calabria)</u>	Brisighelli et al., 2012	50	14
<u>57</u>	<i>CME</i>	Italia (Calabria)	Ottoni et al., 2009	99	33
<u>58</u>	<i>CME</i>	<u>Italia (Cerdeña)</u>	Boattini et al., 2013	70	32
<u>59</u>	<i>CME</i>	<u>Italia (Sicilia)</u>	Ottoni et al., 2009; Boattini et al., 2013	269	108
<u>60</u>	<i>CAU</i>	<u>Karatchaians-Balkarians</u>	Roostalu et al., 2007	203	50
<u>61</u>	<i>CAU</i>	<u>Cáucaso (N)</u>	Richards et al., 2000; Pereira et al., 2005	213	47
<u>62</u>	<i>CAU</i>	<u>Cáucaso (NW)</u>	Roostalu et al., 2007	234	69
<u>63</u>	<i>CAU</i>	<u>Osetia</u>	Roostalu et al., 2007	296	45
<u>64</u>	<i>CAU</i>	<u>Daguestán</u>	Roostalu et al., 2007	269	60
<u>65</u>	<i>CAU</i>	<u>Georgia</u>	Roostalu et al., 2007	192	30
<u>66</u>	<i>CAU</i>	<u>Armenia</u>	Roostalu et al., 2007	175	54
<u>67</u>	<i>ANA</i>	<u>Turquía</u>	Roostalu et al., 2007	360	90
<u>68</u>	<i>ANA</i>	<u>Turquía (E)</u>	Richards et al., 2000; Pereira et al., 2005	79	11
<u>69</u>	<i>NES</i>	<u>Líbano</u>	Roostalu et al., 2007	167	34
<u>70</u>	<i>NES</i>	<u>Siria</u>	Roostalu et al., 2007	159	28
<u>71</u>	<i>NES</i>	<u>Jordán</u>	Roostalu et al., 2007; González et al., 2008; Ennafaa et al., 2009	276	62
<u>72</u>	<i>NES</i>	<u>Palestina</u>	Richards et al., 2000; Pereira et al., 2005	110	29
<u>73</u>	<i>ARA</i>	<u>Arabia</u>	Richards et al., 2000; Pereira et al., 2005; Roostalu et al., 2007; Ennafaa et al., 2009	1132	102
<u>74</u>	<i>ARA</i>	<u>Dubai</u>	Brandstätter et al., 2008	249	24
<u>75</u>	<i>ARA</i>	Yemen	Cerny et al., 2008	185	9
<u>76</u>	<i>ARA</i>	Yemen (Al-Mahra)	Al-Abri et al., 2012	70	9
<u>77</u>	<i>ARA</i>	Omán (Dhofar)	Al-Abri et al., 2012	196	34

Código	Región	Población	Referencia	N	N _H
<u>78</u>	NAF	<u>Sahara</u>	Ennafaa et al., 2009	103	25
79	NAF	Bereber Marruecos (Asni)	Coudray et al., 2009	53	20
80	NAF	Marruecos (El Jadida)	Harich et al., 2010	81	14
<u>81</u>	NAF	Marruecos	Rando et al., 1998; Ennafaa et al., 2009	302	80
82	NAF	Marruecos (Tetouan)	Rhouda et al., 2009	281	61
83	NAF	Marruecos (Al Hoceima)	Rhouda et al., 2009	80	22
84	NAF	Bereber Marruecos (Bouhria)	Coudray et al., 2009	70	26
85	NAF	Marruecos (Oujda)	Rhouda et al., 2009	193	52
86	NAF	Bereber Marruecos (Figuig)	Coudray et al., 2009	94	23
<u>87</u>	NAF	<u>Argelia</u>	Bekada et al., 2013	240	69
<u>88</u>	NAF	<u>Túnez (Testour)</u>	Cherni et al., 2009	50	16
<u>89</u>	NAF	<u>Túnez</u>	Cherni et al., 2009; Ennafaa et al., 2009	338	85
90	NAF	Túnez (Yerba)	Loueslati et al., 2006	59	12
<u>91</u>	NAF	<u>Libia</u>	Fadhlaoui-Zid et al., 2011	269	46
92	NAF	Bereber Egipto (Siwa)	Coudray et al., 2009	78	1
<u>93</u>	WAF	<u>Mauritania</u>	Rando et al., 1998; González et al., 2006; Ennafaa et al., 2009	64	10

Referencias

- Achilli A et al. (2004) *Am J Hum Genet* 75:910–918.
- Al-Abri A et al. (2012) *Am J Phys Anthr* 149:291–298.
- Álvarez L et al. (2010) *Am J Phys Anthr* 142:531–539.
- Álvarez-Iglesias V et al. (2009) *PLoS One* 4:e5112.
- Behar DM et al. (2012) *Am J Hum Genet* 90:486–493.
- Bekada A et al. (2013) *PLoS One* 8:e56775.
- Boattini A et al. (2013) *PLoS One* 8:e65441.
- Brandstätter A et al. (2008) *BMC Evol Biol* 8:191.
- Brehm A et al. (2003) *Hum Genet* 114:77–86.
- Brisighelli F et al. (2012) *PLoS One* 7:e50794.
- Cardoso S et al. (2011) *Am J Phys Anthr* 145:480–48.
- Casas MJ et al. (2006) *Am J Phys Anthr* 131:539–551.
- Cerný V et al. (2008) *Am J Phys Anthr* 136:128–137.
- Cherni L et al. (2009) *Am J Phys Anthr* 139:253–260.
- Coudray C et al. (2009) *Ann Hum Genet* 73:196–214.
- Ennafaâ H et al. (2009) *BMC Genet* 10:8.
- Fadhlaoui-Zid K et al. (2011) *Am J Phys Anthr* 145:107–117.
- Finnilä S et al. (2001) *Am J Hum Genet* 68:1475–1484.
- García O et al. (2011) *Heredity* 106:37–45.
- González AM et al. (2006) *Ann Hum Genet* 70:631–657.
- González AM et al. (2008) *Ann Hum Biol.* 35:212–231
- Harich N et al. (2010) *BMC Evol Biol* 10:138.
- Karachanak S et al. (2012) *Int J Leg Med* 126:497–503.
- Larruga JM et al. (2001) *Eur J Hum Genet* 9:708–716.
- Loueslati BY et al. (2006) *Am J Hum Biol* 18:149–153.

- Malyarchuk BA et al. (2006) *Hum Biol* 78:681–696.
- Ottoni C et al. (2009) *Ann Hum Biol* 36:785–811.
- Pereira L et al. (2004) *Int J Leg Med* 118:132–136.
- Pereira L et al. (2005) *Genome Res* 15:19–24.
- Picornell A et al. (2005) *Am J Phys Anthr* 128:119–130.
- Rando JC et al. (1998) *Ann Hum Genet* 62:531–550.
- Rhouda T et al. (2009) *Mitochondrion* 9:402–407.
- Richards M et al. (2000) *Am J Hum Genet* 67:1251–1276.
- Roostalu U et al. (2007) *Mol Biol Evol* 24:436–448.
- Santos C et al. (2010) *Am J Phys Anthr* 141:610–619.
- Turchi C et al. (2008) *Int J Leg Med* 122:199–204.

Tabla S10. Caracterización molecular de las 337 muestras autóctonas andaluzas pertenecientes al clado mitocondrial H.

-----**VER *CD* adjunto**-----

Figura S14. Localización geográfica de las 93 poblaciones utilizadas en los mapas de contorno del clado H y de sus principales sub-linajes. Se muestra un marco geográfico a gran escala (A) y una visión detallada de la Península Ibérica (B). Los códigos y referencias están registrados en la **Tabla S9**.

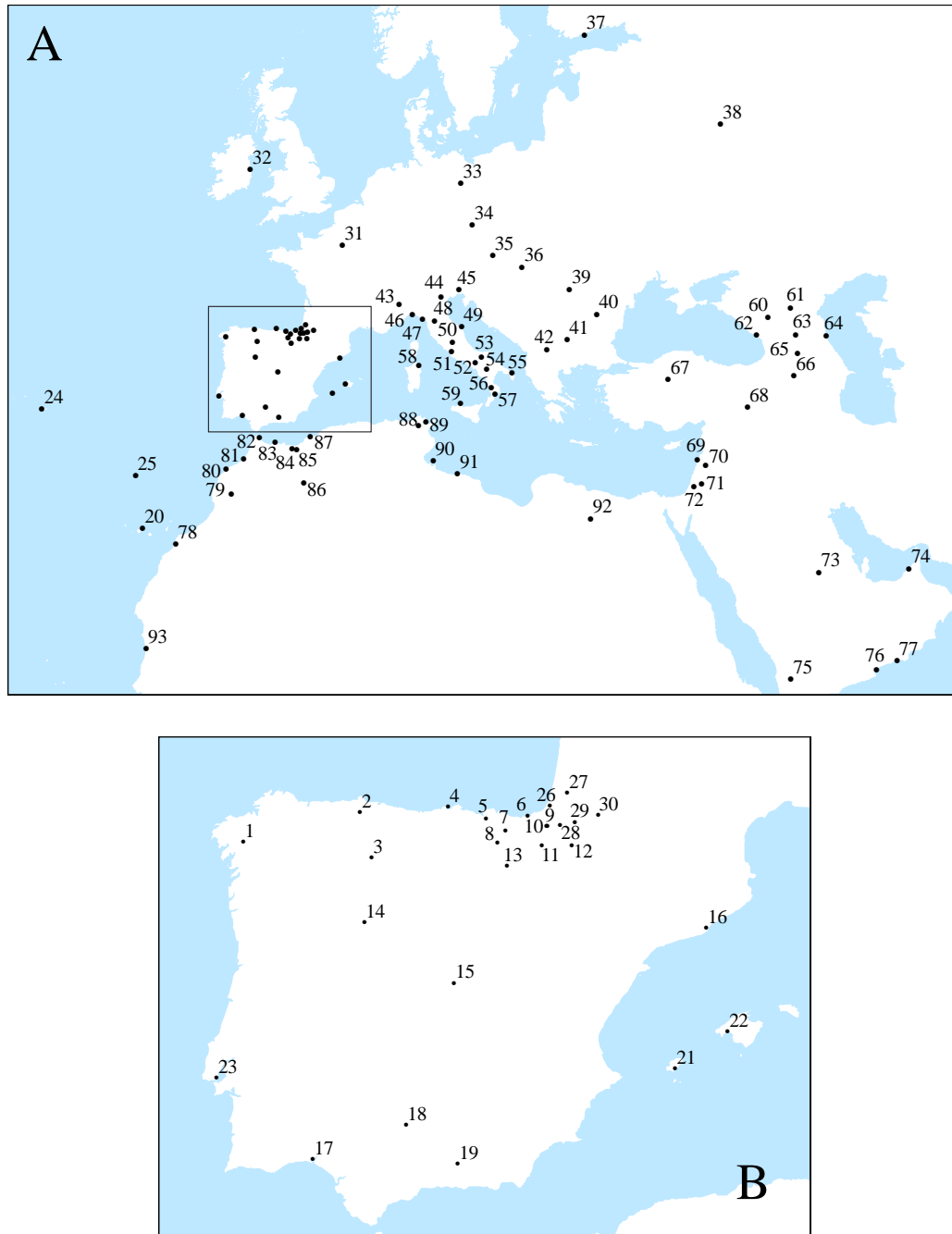


Figura S15. *Median-Joining Network* que muestra las relaciones filogenéticas de las 337 muestras andaluzas pertenecientes al haplogrupo H. Los polimorfismos mostrados son mutaciones de la región control, excepto los indicados en rojo (posiciones diagnósticas de pertenencia a sub-haplogrupo de la región codificante). Las transversiones se muestran con el cambio de base tras la mutación y el sufijo 'd' indica una deleción. Los asteriscos muestran los nodos basales (nps. 7028 para H*, nps. 3010 para H1 y 6776 para H3). El tamaño de los círculos es proporcional a la frecuencia del haplotipo. Las muestras de Huelva se indican en rojo y las de Granada en verde.

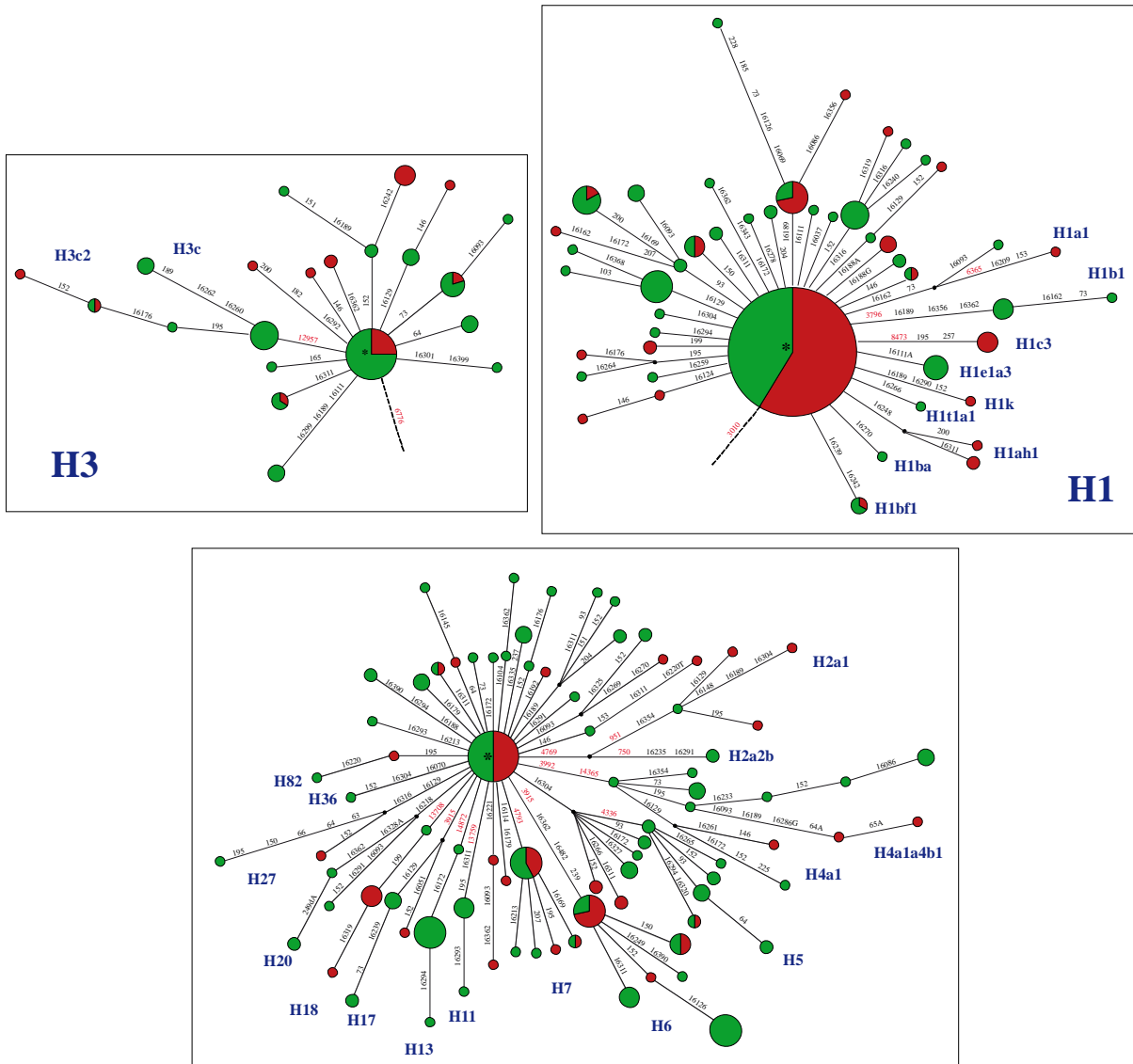


Figura S16. Distribución espacial de los sub-haplogrupos H que poseen una amplia variación geográfica (H2, H4, H5 y H6).

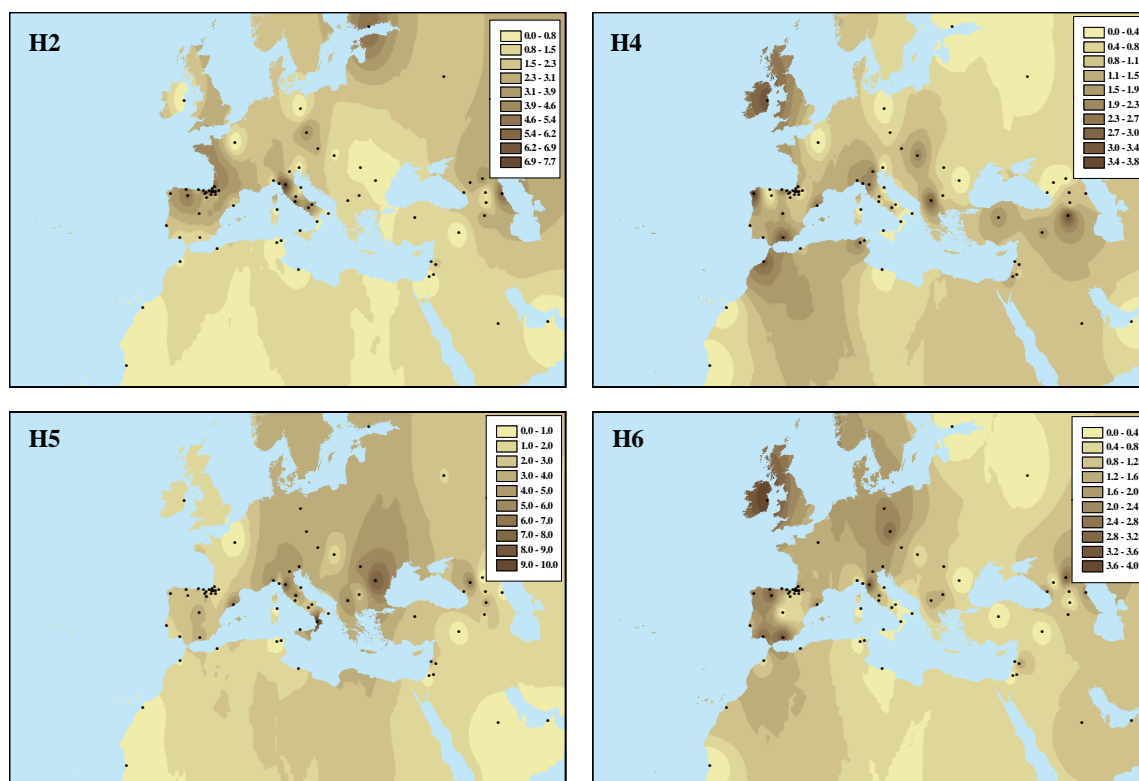


Figura S17. Distribución espacial de los sub-haplogrupos H que poseen una distribución preferencial en el Mediterráneo occidental y el este de Europa (H17). Las frecuencias más altas del haplogrupo H18 se concentran en el oeste de la Península Ibérica.

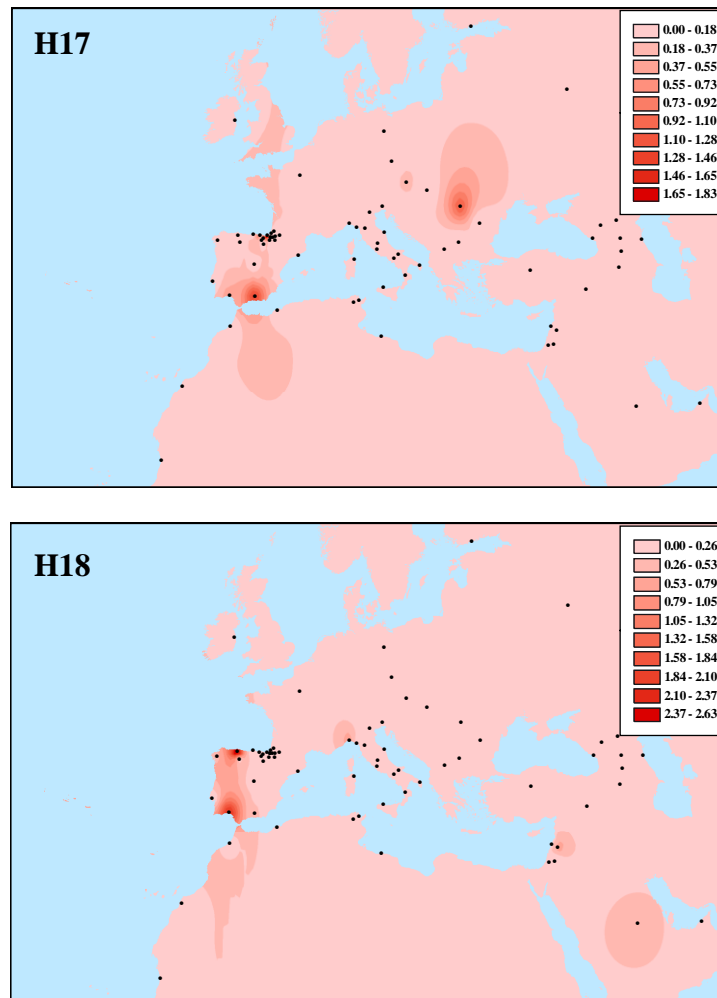


Figura S18. Distribución espacial de los sub-haplogrupos H que poseen una distribución preferencial en la Europa central y oriental (H7, H10 y H11).

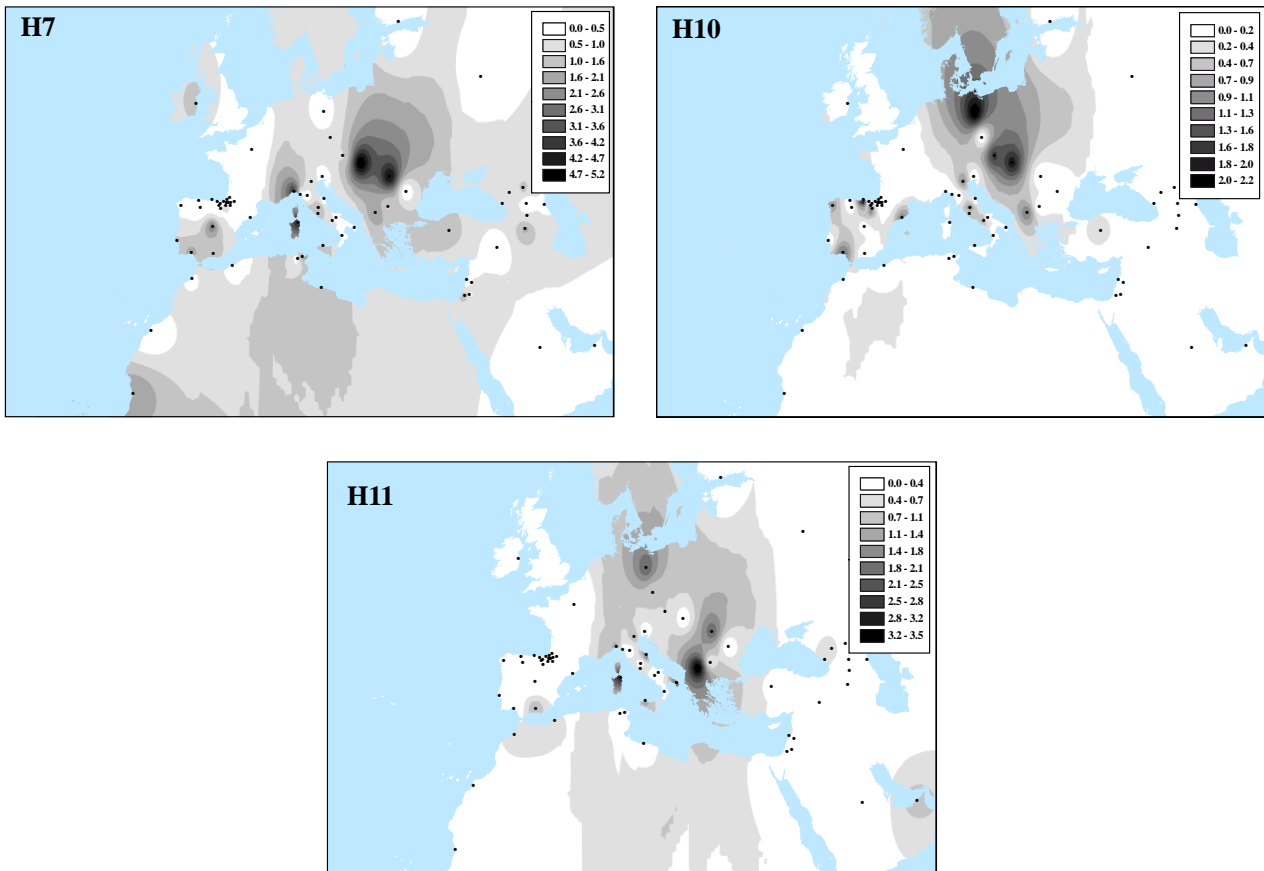


Figura S19. Distribución espacial de los sub-haplogrupos H que presentan una distribución preferencial en el Mediterráneo central y oriental (H13 y H20).

