

UNIVERSIDAD DE GRANADA

FACULTAD DE MEDICINA

**DEPARTAMENTO DE MEDICINA LEGAL, TOXICOLOGÍA
Y ANTROPOLOGÍA FÍSICA**



**GENÉTICA DE LOS FLUJOS POBLACIONALES EN EL
TERRITORIO QUE CONFORMABA EL REINO DE
GRANADA**

**GENETICS POPULATION FLOWS IN THE REGION
COMPRISED BY THE FORMER KINGDOM OF
GRANADA**

Programa Oficial de Doctorado en Biomedicina

María Saiz Guinaldo

Granada 2016

Editor: Universidad de Granada. Tesis Doctorales
Autora: María Saiz Guinaldo
ISBN: 978-84-9125-976-3
URI: <http://hdl.handle.net/10481/44019>

ÍNDICE

CONTENTS

ABREVIATURAS	XV
JUSTIFICACIÓN	XIX
OBJETIVOS	XXI

CAPÍTULO 1

<u>ANTECEDENTES</u>	<u>1</u>
ABSTRACT	1
1. GENÉTICA FORENSE	3
1.1 ORÍGENES	3
1.2 MÉTODOS ACTUALES	4
1.2.1. Herramientas para la identificación de individuos indubitados	5
1.2.2. Herramientas para la identificación de personas desconocidas	6
1.2.3. Herramientas para relacionar individuos con acciones criminales	7
1.2.4. Medicina legal y almacenamiento de datos	8
1.3 FUTURO EN GENÉTICA FORENSE	8
2. SECUENCIACIÓN DE NUEVA GENERACIÓN	9
2.1 NGS EN GENÉTICA FORENSE	10
2.2 NGS EN EL ESTUDIO DE LA HISTORIA GENÉTICA HUMANA	11
2.3 NGS EN GENÉTICA DE POBLACIONES	12
3. ANTROPOLOGÍA MOLECULAR	13
3.1 POLIMORFISMOS	13
3.2 MEDIDAS DE DIVERSIDAD	13
3.3 ANTROPOLOGÍA MOLECULAR EN LA PENÍNSULA IBÉRICA	14
4. BIBLIOGRAFÍA	18

CAPÍTULO 2

<u>ESTUDIO DEL CONTEXTO HISTÓRICO-GEOGRÁFICO DE LA POBLACIÓN</u>	<u>25</u>
ABSTRACT	25
1. INTRODUCCIÓN	27
1.1 HISTORIA DEL REINO DE GRANADA	27
1.1.1. Orígenes	27
1.1.2. Conquista musulmana	27
1.1.3. La dinastía zirí	28
1.1.4. Los almorávides	28

1.1.5. Dominación almohade	29
1.1.6. La dinastía nasri o nazarí	29
1.1.7. El reino cristiano de Granada	32
1.2 INFLUENCIAS EN LA CULTURA	34
1.3 INFLUENCIA EN EL ARTE	35
1.4 INFLUENCIAS EN LA LENGUA	36
1.5 INFLUENCIAS EN LA TOPONIMIA	37
1.6 INFLUENCIAS EN LA GASTRONOMÍA	37
2. MATERIALES Y MÉTODOS	39
2.1 TOMA DE MUESTRAS	39
2.2 ESTUDIO DE APELLIDOS	39
3. RESULTADOS Y DISCUSIÓN	41
3.1 IMPORTANCIA DEL ESTUDIO DE APELLIDOS	41
3.2 ESTUDIO DE APELLIDOS	42
4. BIBLIOGRAFÍA	47

CAPÍTULO 3

GENOTIPADO DE STRs EN CROMOSOMAS AUTOSÓMICOS **49**

ABSTRACT	49
1. INTRODUCCIÓN	51
1.1. GENERACIÓN DE DIVERSIDAD EN MICROSATÉLITES	54
1.2. APLICACIONES	55
1.2.1. aSTRs como herramienta en estudios de criminalística	55
1.2.2. aSTRs como herramienta en estudios de identificación genética	55
1.2.3. aSTRs como herramienta en estudios de filiación	56
1.2.4. Estudio de la estructura de las poblaciones	56
1.2.5. Estudio de la evolución de las poblaciones	57
1.2.6. Otras aplicaciones	57
2. MATERIAL Y MÉTODOS	59
2.1 POBLACIÓN DE ESTUDIO	59
2.2 ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA	59
2.2.1. Extracción	59
2.2.2. Cuantificación	59
2.2.3. Amplificación	59
2.2.4. Visualización de los resultados	62
2.3 ANÁLISIS ESTADÍSTICO DE LA MUESTRA	63
3. RESULTADOS Y DISCUSIÓN	66
3.1 ESTUDIO POBLACIONAL	66
3.2 ESTUDIO DE LA ESTRUCTURA DE LA POBLACIÓN	69
3.3 ESTUDIO DE FRECUENCIAS ALÉLICAS	70
3.4 ESTUDIO DE DISTANCIAS GENÉTICAS	71

3.5 ESTUDIO DE LOS COMPONENTES ANCESTRALES	73
4. BIBLIOGRAFÍA	75

CAPÍTULO 4

GENOTIPADO DE INDELS EN CROMOSOMAS AUTOSÓMICOS **81**

ABSTRACT	81
1. INTRODUCCIÓN	83
1.1. ORIGEN Y EVOLUCIÓN DE LOS INDELS	84
1.2. APLICACIONES	86
1.2.1. INDELS como marcadores genéticos forenses	86
1.2.2. INDELS como herramienta complementaria en estudios de filiación	88
1.2.3. Estudio de la estructura de las poblaciones	88
1.2.4. INDELS en el desarrollo de enfermedades	88
2. MATERIAL Y MÉTODOS	90
2.1. POBLACIÓN DE ESTUDIO	90
2.2. ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA	90
2.2.1. Extracción	90
2.2.2. Cuantificación	90
2.2.3. Amplificación	90
2.2.4. Visualización de los resultados	93
2.3. ANÁLISIS ESTADÍSTICO DE LA MUESTRA	94
3. RESULTADOS Y DISCUSIÓN	96
3.1. ESTUDIO POBLACIONAL	96
3.2. ESTUDIO DE LA ESTRUCTURA DE LA POBLACIÓN	97
4. BIBLIOGRAFÍA	100

CAPÍTULO 5

GENOTIPADO DE STRs EN CROMOSOMA Y **105**

ABSTRACT	105
1. INTRODUCCIÓN	107
1.1. ORIGEN Y EVOLUCIÓN DEL CROMOSOMA Y	108
1.2. MARCADORES GENÉTICOS EN EL CROMOSOMA Y	109
1.2.1. STRs de cromosoma Y	109
1.2.2. SNPs de cromosoma Y	113
1.3. HAPLOGRUPOS DE CROMOSOMA Y	113
1.4. APLICACIONES	115
1.4.1. Y-STRs como herramienta en estudios de filiación	116
1.4.2. Y-STRs como herramienta en estudios de criminalística	116
1.4.3. Estudio de la antropología evolutiva y la historia de las poblaciones	116
1.4.4. Estudios de genealogía genética	117
1.4.5. Investigación en biomedicina	117

2. MATERIAL Y MÉTODOS	119
2.1. POBLACIÓN DE ESTUDIO	119
2.2. ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA	119
2.2.1. Extracción	119
2.2.2. Cuantificación	119
2.2.3. Amplificación	119
2.2.4. Visualización de los resultados	121
2.3. ANÁLISIS ESTADÍSTICO DE LA MUESTRA	122
3. RESULTADOS Y DISCUSIÓN	126
3.1 ESTUDIO POBLACIONAL	126
3.2 ESTUDIO DE LA ESTRUCTURA DE LA POBLACIÓN	128
3.3 ESTUDIO DE HAPLOGRUPOS	129
3.3.1. Estudio de todos los haplogrupos en la población GMA	131
3.3.2. Estudio de los haplogrupos mayoritarios en la población GMA	134
3.4 ESTUDIO DE APELLIDOS	135
3.4.1. Haplogrupos y apellidos	135
3.4.2. Haplotipos y apellidos	137
3.5 ESTUDIO DE FRECUENCIAS ALÉLICAS	137
3.6 ESTUDIO DE DISTANCIAS GENÉTICAS	139
3.7 ESTUDIO DE LOS COMPONENTES ANCESTRALES	140
4. BIBLIOGRAFÍA	142

CAPÍTULO 6

POLIMORFISMOS EN LA REGIÓN CONTROL DEL ADN MITOCONDRIAL 153

ABSTRACT	153
1. INTRODUCCIÓN	155
1.1. ORIGEN Y EVOLUCIÓN DEL ADN MITOCONDRIAL	157
1.2. POLIMORFISMOS DE ADN MITOCONDRIAL	158
1.2.1. Nomenclatura	159
1.2.2. Heteroplasmia	159
1.3. HAPLOGRUPOS DE ADN MITOCONDRIAL	161
1.4. APLICACIONES	163
1.4.1. ADNmt como herramienta en estudios de filiación	163
1.4.2. ADNmt como herramienta en estudios de criminalística	163
1.4.3. Estudio de la antropología evolutiva y la historia de las poblaciones	164
1.4.4. Investigación en biomedicina	165
2. MATERIAL Y MÉTODOS	166
2.1. POBLACIÓN DE ESTUDIO	166
2.2. ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA	166
2.2.1 Extracción	166
2.2.2 Cuantificación	166

2.2.3 Amplificación	166
2.2.4 Purificación del producto amplificado	168
2.2.5 Secuenciación	168
2.2.6 Purificación del producto secuenciado	169
2.2.7 Visualización de los resultados	170
2.3. ANÁLISIS ESTADÍSTICO DE LA MUESTRA	170
3. RESULTADOS Y DISCUSIÓN	173
3.1. ESTUDIO POBLACIONAL	173
3.2. ESTUDIO DE LA ESTRUCTURA DE LA POBLACIÓN	179
3.3. ESTUDIO DE HAPLOGRUPOS	181
3.3.1. Estudio de los haplogrupos mayoritarios en la población GMA	184
3.3.2. Estudio de todos los haplogrupos en la población GMA	186
3.4. ESTUDIO DE FRECUENCIAS DE HAPLOGRUPOS	188
3.5. ESTUDIO DE DISTANCIAS GENÉTICAS	189
4. BIBLIOGRAFÍA	191
CHAPTER 7	
GENETIC VARIABILITY IN THE GMA POPULATION AND ITS RELATION WITH THE FORMER INHABITANTS OF NORTH AFRICA	199
ABSTRACT	199
1. INTRODUCTION	201
2. FORENSIC ANALYSIS	202
3. GMA POPULATION STRUCTURE	203
4. HAPLOGROUPS	204
5. ALLELIC AND HAPLOGROUP FREQUENCIES	206
6. GENETIC DISTANCES	207
7. FINAL CONSIDERATIONS	209
8. REFERENCES	211
CONCLUSIONES	213
APÉNDICE I: Protocolos de extracción y cuantificación	217
APÉNDICE II: Fundamentos de la electroforesis capilar	219
APÉNDICE III: Software	225
APÉNDICE IV: Publicaciones	237

ABREVIATURAS

ABBREVIATIONS

A	Adenina
AC	Antes de Cristo
ADN	Ácido Desoxirribonucleico
ADNmt	ADN mitocondrial
AIMs	Ancestry Informative Markers (Marcadores informativos de ancestralidad)
AMOVA	Analysis of molecular variance (Análisis de la Varianza Molecular)
ARN	Ácido Ribonucleico
ARNm	ARN mensajero
ARNmi	micro ARN
ATP	Adenosín trifosfato
C	Citosina
CDP	Combined power of discrimination (Poder de discriminación combinado)
CEP	Combined power of exclusion (Poder de exclusión combinado)
CODIS	Combined DNA Index System (Sistema de comparación de ADN)
DC	Discrimination capacity (Capacidad de discriminación)
ddNTP	Didesoxinucleótido degenerado
Del	Delección
d.f.	Grados de libertad
DMSO	Dimetilsulfóxido
DNA	Desoxyribonucleic Acid
dNTP	Desoxinucleótido
DTR	Dye Terminator Removal (Eliminación de terminadores marcados)
EMPOP	EDNAP Mitochondrial DNA Population Database (Base de datos poblacionales de ADN mitocondrial EDNAP)
G	Guanina

GMA	Granada-Málaga-Almería
GWAS	Genome-wide association studies (Estudios de asociación en genoma completo)
HD	Haplotype diversity (Diversidad haplotípica)
He	Heterozigosidad esperada
Hg	Haplogrupo
Ho	Heterozigosidad observada
HRM	High Resolution Melting (Detección de gran sensibilidad del punto de fusión)
HV	Hypervariable Region (Región hipervariable)
HWE	Hardy-Weinberg Equilibrium (Equilibrio Hardy-Weinberg)
INDEL	Insertion/Deletion polymorphism (Polimorfismos Inserción/Delección)
ISFG	International Society of Forensic Genetics (Sociedad Internacional de Genética Forense)
IUPAC	International Union of Pure and Applied Chemistry (Unión Internacional de Química Pura y Aplicada)
Kb	Kilobases
kV	Kilovoltios
LR	Likelihood Ratio (Índice de probabilidad)
Mb	Megabases
MDP	Mean power of discrimination (Poder de discriminación medio)
MDS	Multidimensional scaling (Escalando multidimensional)
MEP	Mean power of exclusion (Poder de exclusión medio)
MHo	Mean observed heterozygosity (Heterozigosidad observada media)
min	Minutos
MID	Marshfield Diallelic Insertion/Deletion Polymorphism Database (Base de datos de polimorfismos dialélicos de inserción/delección Marshfield)
MMP	Mean matching probability (Probabilidad de coincidencia media)
MP	Matching Probability (Probabilidad de coincidencia)
MPIC	Mean polymorphic informative content (Índice de contenido polimórfico medio)
ng	Nanogramo
NGS	Next Generation Sequencing (Secuenciación de nueva generación)

NRY	Nonrecombining portion of the Y chromosome (Porción no recombinante del cromosoma Y)
OXPHOS	Oxidative phosphorylation (Sistema de fosforilación oxidativa)
p	p-valor
PAR	Pseudoautosomal Regions (Regiones pseudoautosómicas)
pb	Pares de bases
PCR	Polymerase chain reaction (Reacción en cadena de la polimerasa)
PD	Poder de discriminación
PE	Poder de exclusión
pg	picogramos
PIC	Polymorphic informative content (Índice de contenido polimórfico)
POP	Performance Optimized Polymer (Polímero de rendimiento optimizado)
RFLP	Restriction Fragment Length Polymorphisms (Polimorfismos de Fragmentos de Restricción)
RFU	Relative Fluorescent Units (Unidades de Fluorescencia Relativa)
RM-YSTR	Rapidly Mutating Y Short Tandem Repeat (STRs de cromosoma Y rápidamente mutables)
s	Segundos
SNP	Single Nucleotide Polymorphism (Polimorfismos de nucleótido único)
SHOX	Short Stature Homeobox (Homeobox de tamaño corto)
STR	Short Tandem Repeat (Repeticiones cortas en tándem)
T	Timina
TPI	Typical Paternity Index (Índice típico de paternidad)
U	Unidades
YHRD	Y-Chromosome Haplotype Reference Database (Base de datos de haplotipos de referencia de cromosoma Y)
°C	Grados centígrados
μl	Microlitro

JUSTIFICACIÓN

JUSTIFICATION

Desde que en nuestro medio se comenzó a utilizar el ADN como técnica de identificación humana en el marco forense, allá por el año 1990, en el Laboratorio de Identificación Genética de la Universidad de Granada se han analizado miles de muestras, bien en un entorno civil (normalmente casos de filiación) o bien en un contexto de investigación criminal (delitos contra la libertad sexual, homicidios, etc).

Lógicamente, un porcentaje significativo de estos casos provenían y provienen de nuestra área de más directa influencia geográfica, o sea, de la provincia de Granada y alrededores.

Considerando que todos los cálculos estadísticos necesarios para establecer la probabilidad o razón de verosimilitud se desarrollan a partir de los estudios de las frecuencias de los diferentes alelos, SNPs y haplotipos de las poblaciones de referencia, es pues necesario conocer con el mayor detalle posible la estructura de todas las áreas donde se trabaje.

En este estudio de Tesis Doctoral nos hemos centrado en la caracterización de la población de Granada, Málaga y Almería correspondiente al antiguo Reino de Granada, y ello por dos motivos principales.

Primero, el más operativo, porque es de donde se reciben muchas de las muestras que nosotros estudiamos; segundo, porque nos permite conocer mejor la estructura genética de la misma y las posibles influencias que hayan podido afectarla y aún se encuentren presentes, teniendo en cuenta que esta zona geográfica fue el último reducto árabe en España y que existen numerosos datos de su repoblación masiva tras la expulsión -más exactamente "las expulsiones"- de poblaciones árabes tras la toma de la ciudad por los Reyes Católicos en 1492.

La historia de la Península Ibérica, y especialmente por lo mencionado anteriormente la de las provincias de Granada, Málaga y Almería, se ha visto influenciada por un gran número de poblaciones. El paso de visigodos, bereberes, sirios, almorávides (saharauis), almohades (marroquíes), judíos y cristianos por el Reino de Granada ha hecho de las actuales ciudades de Granada, Málaga y Almería un crisol de civilizaciones que queda reflejado en sus monumentos arquitectónicos, sus costumbres y su folklore. Sin embargo, quedaba pendiente lo que hemos pretendido hacer con el presente, que no es sino caracterizar con cierto detalle la influencia genética que estas civilizaciones puedan haber dejado en la población actual de estas tres provincias.

Since DNA began to be used as a technique for human identification in forensic context, back in 1990, the Laboratory of Genetic Identification of the University of Granada has analyzed thousands of samples, either in a civil setting (usually cases of affiliation) or in the context of criminal investigation (crimes against sexual freedom, murder, etc).

Logically, a significant percentage of these cases came and come from our most direct geographical influence area, the province of Granada and surroundings.

Whereas all the necessary statistical calculations to establish the probability or likelihood ratio come from studies of the frequencies of different alleles, SNPs and haplotypes of reference populations, is therefore necessary to know in detail the structure of all areas we work with.

In this Doctoral Thesis, we have focused on the characterization of the population of Granada, Malaga and Almería, region where it was established the former Kingdom of Granada, for two main reasons.

First, the most operational, because it is the region where many of the samples are received from; second, because it gives us a better understanding of its genetic structure and the possible influences that may have affect it and still are present, considering that this geographical area was the last Arab stronghold in Spain and that there are numerous data of its massive repopulation after the expulsion - more exactly "expulsions" - of Arab populations after the occupation of the territory by the Catholic Monarchs in 1492.

The history of the Iberian Peninsula, and especially that of the provinces of Granada, Malaga and Almeria, has been influenced by a large number of populations. The passage of Visigoths, Berbers, Syrians, *Almorávides* (Saharawi), *Almohades* (Moroccans), Jews and Christians through the Kingdom of Granada has made the present cities of Granada, Malaga and Almeria a melting pot of civilizations that is reflected in its architectural monuments, traditions and folklore. However, there was still outstanding what we intend to do in this Doctoral Thesis, to characterize in detail the genetic influence that these civilizations may have left in the current population of these three provinces.

OBJETIVOS

OBJECTIVES

1. Caracterizar la estructura genética de la población de Granada, Málaga y Almería mediante el genotipado de los marcadores genéticos:
 - STRs autosómicos: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, AMELOGENINA, D5S818, FGA.
 - INDELS autosómicos: 38 INDELS incluidos en la reacción multiplex desarrollada por Pereira et al. 2012.
 - STRs en el cromosoma Y: DYS389I, DYS389II, DYS390, DYS456, DYS19, DYS385, DYS458, DYS391, DYS392, DYS393, DYS439, DYS635, DYS437, DYS438, DYS448, GATA H4.1
 - Región control completa del ADN mitocondrial.
2. Calcular el grado de variabilidad de cada marcador, mediante sus frecuencias en la población y los diferentes parámetros de interés forense. Discutir la viabilidad de estos marcadores para su uso en identificación genética y sus posibles aplicaciones en el campo forense.
3. Estudiar los apellidos presentes en la población de estudio. Determinar su origen así como su distribución geográfica y establecer una relación entre ambos. Comparar la distribución de los apellidos con los haplotipos y haplogrupos de cromosoma Y encontrados en la población.
4. Identificar y comparar los diferentes haplogrupos de cromosoma Y así como de ADN mitocondrial en la población de estudio. Estudiar las relaciones entre los haplogrupos encontrados en la población y su posible origen.
5. Comparar la población de estudio con poblaciones que han influido en su origen y en la estructura genética actual de la zona geográfica estudiada.

1. Characterize the genetic structure of the population from Granada, Málaga and Almería by genotyping:
 - Autosomal STRs: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, AMELOGENIN, D5S818, FGA.
 - Autosomal INDELS: 38 INDELS included in the multiplex reaction described by Pereira et al. 2012.
 - Y chromosome STRs: DYS389I, DYS389II, DYS390, DYS456, DYS19, DYS385, DYS458, DYS391, DYS392, DYS393, DYS439, DYS635, DYS437, DYS438, DYS448, GATA H4.1
 - Mitochondrial DNA control region.
2. Calculate the variability of each marker by the analysis of their frequencies and several forensic parameters. Discuss the viability of these markers in genetic identification and their possible applications in the forensic field.
3. Study the surnames in the population, determine their origin and geographic distribution and establish the relationship between them. Compare the distribution of the surnames and the Y chromosome haplotypes and haplogroups found in the studied population.
4. Identify and compare the different Y chromosome and mitochondrial DNA haplogroups in the population. Study the relations between the observed haplogroups and their possible origin.
5. Compare the studied population with those that may have influenced its origin and the contemporary genetic structure of this geographical area.

ABSTRACT

Forensic genetics and molecular anthropology are fairly new sciences. Both of them use the same tools in their development and the latest techniques in molecular biology.

Advances in the mapping of the human genome generate valuable information for forensic genetics and molecular anthropology. Nowadays, a huge number of STR markers have been determined, not only autosomal but also in sexual chromosomes. Furthermore, a large number of mitochondrial DNA polymorphisms are defined through sequencing. All this information had led to the resolution of thousand forensic cases as far as the identification of human populations, their origin and evolution.

However, the continuous evolution of molecular biology techniques and the development of whole genome sequencing technologies generate a huge amount of information valuable for both disciplines. Differences in the sequence of the repetition core of some STR markers have been defined by sequencing increasing the discrimination power. On the other hand, a lot of SNPs have been identified as markers to deduce geographical origin of an individual as far as some phenotypic traits. Finally, whole genome of *Homo neanderthalensis* has been sequenced shedding light on the possible relation between them and modern humans.

Tanto la genética forense como la antropología molecular son dos ciencias relativamente nuevas. Ambas coinciden en las herramientas que facilitan su desarrollo y las nuevas técnicas de biología molecular.

1. GENÉTICA FORENSE

La ciencia forense es una disciplina enfocada a ayudar a la justicia a resolver casos criminales y civiles. Ésta abarca un gran número de disciplinas; biología, química, física, matemáticas... Sin embargo, en los últimos 20 años, una de las herramientas que ha revolucionado las investigaciones forenses ha sido el análisis del ADN.

1.1 Orígenes

En 1984, Alec Jeffreys descubrió unos elementos extremadamente variables y heredables en las regiones de ADN repetido, analizados mediante sondas multilocus; conocido como "*DNA fingerprinting*" [1]. Este nuevo descubrimiento se aplicó a un gran número de disciplinas tanto en estudios de diversidad y conservación de especies como en estudios clínicos y antropológicos. La genética de poblaciones aprovechó estos marcadores para poder explicar fenómenos de variabilidad genética. La ausencia de selección natural en estas zonas podía resolver y sustentar muchas teorías relacionadas con la herencia. Poder estudiar las distintas poblaciones mediante caracteres que no sufrían selección y cuya herencia era bien conocida, permitió comenzar a respaldar teorías evolutivas, migratorias y de mestizaje [2].

Esta técnica alcanzó gran relevancia social al empezar a aplicarse en casos criminales y civiles, permitiendo comparar el ADN de un individuo con el identificado en restos biológicos encontrados en la escena del crimen, o con el de otra persona, para incluirla como sospechosa o excluirla [3]. La posibilidad de que exista otra persona que tenga el mismo perfil genético va disminuyendo a medida que se aumentan los estudios sobre las muestras comparadas, por lo que las conclusiones obtenidas a través el estudio del ADN permiten identificar a un individuo con una probabilidad muy próxima al 100%. La probabilidad de que dos individuos no emparentados escogidos al azar posean el mismo perfil de ADN tras estudiar 5 loci hipervariables, puede llegar a ser de uno entre un billón dependiendo de la combinación de alelos y de sus frecuencias en una determinada

población. Actualmente la mayoría de los estudios tienen 16 loci, lo que convierte al análisis genético en una poderosa herramienta para la identificación humana.

Con el desarrollo de la PCR (Reacción en Cadena de la Polimerasa) en 1988, se dio un gran giro en el uso de estos marcadores ya que se produjo un incremento en la sensibilidad, permitiendo el análisis de pequeñas cantidades de ADN degradado. Actualmente cimienta las bases para todo el genotipado forense de ADN [4].

1.2 Métodos actuales

En la actualidad, los estudios de identificación humana se realizan a través de multiplex de STRs (Short Tandem Repeats) autosómicos desarrollados por casas comerciales. Otras fuentes de variación genética que presentan usos más especializados son los SNPs (Single Nucleotide Polimorphisms) autosómicos, STRs de cromosomas sexuales así como el análisis del ADN mitocondrial (ADNmt) (Figura 1.1) [4].

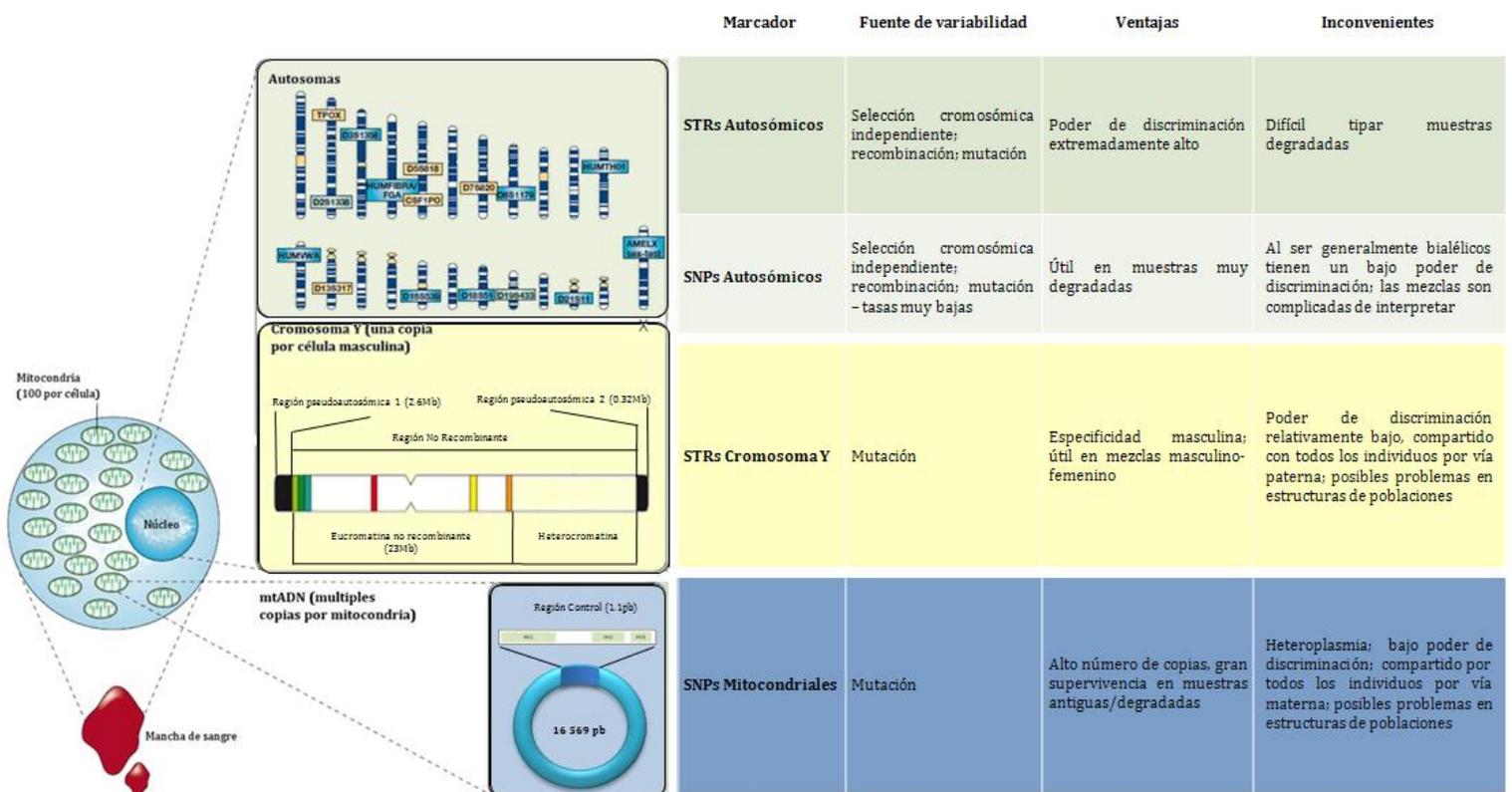


Figura 1.1: Fuentes de variabilidad humana. Imagen basada y modificada de Jobling et al. 2004 [4].

- **STRs autosómicos:** Se han desarrollado reacciones multiplex que determinan hasta 23 marcadores STRs autosómicos con una alta

variabilidad y heterocigosidad, dando lugar a probabilidades de coincidencia menores de 10^{-13} [4].

- **SNPs autosómicos:** Comparados con los STRs, los SNPs tienen valores de heterocigosidad mucho menores y se necesitan unos 50 SNPs para alcanzar los valores más bajos de probabilidad de coincidencia de un perfil de STRs. Se ha demostrado que el uso combinado de STRs y SNPs, en casos complejos de paternidad, multiplica por cuatro los valores del índice de paternidad [5].
- **Marcadores de cromosomas sexuales:** Debido a que la mutación es la única fuente de variabilidad en los haplotipos de cromosoma Y, es mucho menos diverso que los perfiles de marcadores autosómicos. Sin embargo, el cromosoma Y tiene una propiedad crucial para su uso en forense; su exclusividad en varones. Ésta hace que el análisis de STRs de cromosoma Y proporcione información específica sobre el componente masculino. Se ha desarrollado una base de datos poblacional que contiene más de 136 000 perfiles de 917 poblaciones [6]. Además, se ha comprobado que la variabilidad encontrada en el cromosoma Y muestra una fuerte estructura poblacional por lo que son esenciales este tipo de bases de datos. Por otro lado, el análisis STRs en el cromosoma X es de gran utilidad en estudios de paternidad, sobretodo en casos padre-hija, casos con familiares paternos cercanos, pruebas de maternidad y en casos de violación o incesto [7,8]. Existen múltiples estudios que determinan la capacidad forense de estos marcadores [9,10], así como su idoneidad en estudios poblacionales [9,11,12].
- **ADN mitocondrial:** El ADN mitocondrial comparte características similares con el cromosoma Y ya que a causa de la ausencia de recombinación su herencia es uniparental en bloque. Sin embargo, el alto número de copias presentes en el ADNmt, hace de éste una herramienta de gran ayuda en casos en los que la muestra es antigua o está severamente dañada.

1.2.1. Herramientas para la identificación de individuos indubitados

- **SNPs autosómicos y discriminación biogeográfica:** Muchas de las dificultades técnicas que presenta el uso de STRs no existen para los SNPs. El uso de amplicones pequeños es esencial en el trabajo con muestras altamente degradadas y se hace prácticamente imposible con STRs altamente polimórficos. Sin embargo, ya que los SNPs reflejan un cambio en una única base, los

amplicones con los que se trabaja son mucho más pequeños. Además, al no ser secuencias repetitivas, con el uso de SNPs se evitan los artefactos *stutter* (fenómenos de tartamudeo de la polimerasa). Es muy importante desarrollar kits que incluyan SNPs que sirvan para identificar individuos, sea cual sea su origen biogeográfico, de forma fiable [13–15]. Para ello, los SNPs han de tener niveles altos de diversidad alélica intrapoblacional pero manteniendo distancias similares entre las poblaciones. También, deben presentar bajos niveles de diversidad interpoblacional [16].

- **RM-YSTRs para la identificación de individuos masculinos:** Los STRs de cromosoma Y empleados de rutina en genética forense permiten la identificación de grupos de varones relacionados por vía paterna en la gran mayoría de las poblaciones con un alto nivel de resolución. Por tanto, con los Y-STRs habituales no es posible la identificación de un único individuo ya que los parientes masculinos cercanos no se pueden excluir. No obstante, estudios recientes han descrito un gran número de STR de cromosoma Y con una gran tasa de mutación (RM-YSTRs), que permiten distinguir entre varones emparentados muy cercanos (1-5 generaciones) [17–19].

1.2.2. Herramientas para la identificación de personas desconocidas

Debido a que actualmente no existe una base de datos de perfiles genéticos universal, el genotipado forense de ADN guía a las autoridades en la identificación de personas desconocidas [16].

- **Inferencia del origen biogeográfico a través del ADN:** Aunque la información proporcionada a través del análisis de ADN sobre el origen biogeográfico de los individuos no está del todo clara, los marcadores de linaje se han identificado como fuentes valiosas de información geográfica. Características genéticas poblacionales como un tamaño efectivo poblacional (tamaño teórico) reducido y la falta de recombinación hacen que los marcadores de linaje sean especialmente propensos a presentar fenómenos de deriva genética, así como los efectos potenciales de rasgos culturales como patrones residenciales. Sin embargo, no se puede excluir a priori que la muestra dubitada provenga de un individuo procedente de una población mestiza. Por ello, la información obtenida con el análisis de los marcadores de ADN mitocondrial y cromosoma Y se ha de combinar con la proporcionada por marcadores autosómicos, para una mejor inferencia del origen de la muestra [16].

- **Estimación de rasgos fenotípicos a través de ADN:** La introducción de los estudios de asociación en genoma completo (GWAS) para encontrar genes involucrados en caracteres complejos, proporciona información sobre características externamente visibles como el color de ojos, cabello y piel [20–22]; estatura [23] y morfología del cabello [24–26]. Ya se ha desarrollado un kit aceptado en la comunidad forense que contiene 6 SNPs de 6 genes de pigmentación, IrisPlex (Figura 1.2) [27].

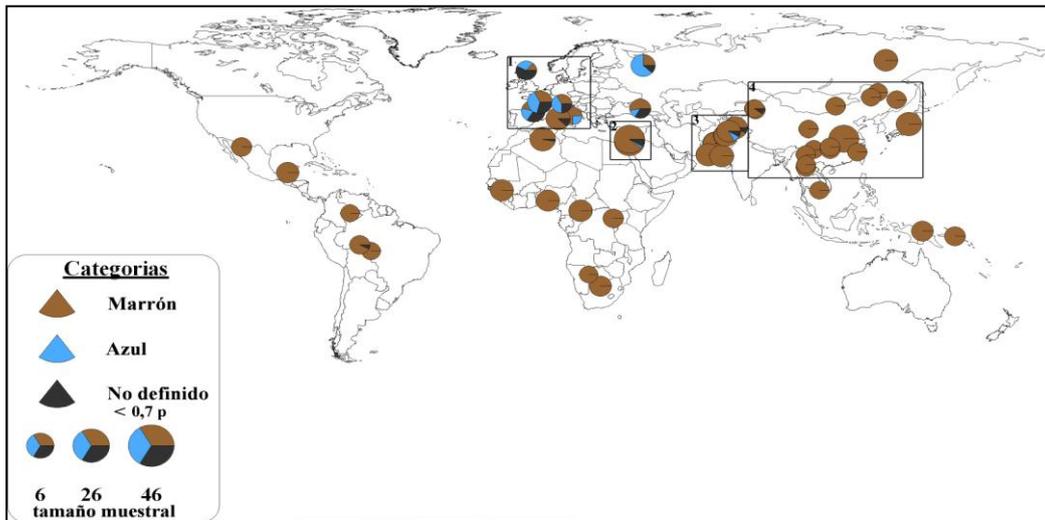


Figura 1.2: Distribución de la predicción del color de ojos basado en los marcadores incluidos en *IrisPlex*. Imagen tomada y modificada de *Walsh et al. 2011* [27].

1.2.3. Herramientas para relacionar individuos con acciones criminales

La información que establece un vínculo entre individuos que han sido identificados por el genotipado de ADN de una muestra y los actos criminales es crucial. No obstante, los intentos de establecer este tipo de relaciones directamente de muestras obtenidas del escenario de un crimen han sido limitados.

- **Determinación del origen celular:** Se han realizado estudios con ARNm (ARN mensajero) y ARNmi (micro ARN) [28,29] para la diferenciación entre los diferentes tejidos humanos. Además, se ha demostrado que sangre, semen, saliva y piel tienen perfiles epigenéticos diferentes (metilación de ADN) [30].
- **Estimación de la edad y el tiempo de depósito de la muestra:** Existen estudios que analizan la degradación dependiente del tiempo en biomarcadores (incluyendo el ARN) para determinar cuántos días, semanas o meses hace que la muestra se depositó [31,32]. Además, es importante determinar en qué

momento del día se depositó una muestra para determinar aspectos cronológicos en la escena del crimen. Para ello se han realizado estudios en muestras de sangre frescas y antiguas determinando hormonas con ciclos circadianos como la melatonina y el cortisol [33,34]. Sin embargo, se necesitan identificar más biomarcadores circadianos que presenten diferentes picos a lo largo del día y que no se vean influenciados por agentes externos [16].

Los métodos empleados en genética forense han de cumplir una serie de requisitos antes de poder ser aplicados en casos criminales. Las técnicas deben estar optimizadas para el trabajo con muestras comprometidas y en pequeñas cantidades. Además, todas tienen que estar validadas demostrando su robustez para poder ser admitidas en un juicio. Por último, se han de implementar unos sistemas de control de calidad una vez se hayan incluido en el trabajo de rutina [4].

1.2.4. Medicina legal y almacenamiento de datos

Los avances científicos y técnicos generan una cantidad de información extraordinaria que debe almacenarse de modo racional y ordenado para su posterior uso. Las bases de datos de identificación genética contienen números y letras asociados al código de identificación de una persona. El acceso a los datos debe de estar perfectamente controlado y las conclusiones que se pueden obtener de los mismos van a depender de los programas informáticos que se autoricen por las diferentes legislaciones. En España, La LEY ORGÁNICA 10/2007, vigente desde el 10 de Octubre 2007, regula la base de datos policial sobre identificadores obtenidos a partir de ADN.

1.3 Futuro en genética forense

Anteriormente se han descrito avances que mejoran el futuro en las investigaciones forenses. Aun así, todavía se han de llevar a cabo muchos cambios ya que existen necesidades no cubiertas. Hay una gran necesidad de desarrollar métodos que permitan identificar en una muestra con mezcla de perfiles los posibles contribuidores a la misma, sobre todo cuando los donantes son del mismo sexo. En este caso, el uso de las nuevas tecnologías de secuenciación de ADN [35] dará lugar a grandes progresos. Además, posibilitará el análisis de pequeñas

cantidades de ADN degradado, muy importante en muchos casos forenses así como en la identificación de víctimas de grandes catástrofes [16].

Otro gran paso importante a alcanzar es aumentar el grado de detalle con el que se puede inferir el origen biogeográfico y los rasgos fenotípicos a través del estudio de materiales biológicos. Para ello, se necesitan nuevas tecnologías para el genotipado en paralelo de un gran número de SNPs [16].

Además de la reconstrucción del crimen mediante biomarcadores, se ha de ir más allá estimando la predisposición genética a determinados rasgos de comportamiento de los perpetradores de los delitos, así como de las condiciones fisiológicas de éstos y de las víctimas en el momento del acto criminal [16].

Existen múltiples avances en la microfabricación de matrices de capilares electroforéticos, así como plataformas que integran el proceso de extracción de las muestras, amplificado y secuenciación, pero aún se necesitará un tiempo para que puedan validarse para su uso entre la comunidad forense. Las técnicas de amplificación de genoma completo tienen un potencial valor forense cuando la cantidad de ADN es muy baja. Sin embargo, la pérdida de heterocigosidad y el desbalance en la amplificación se observa cuando se trabaja con menos de 50pg de ADN y se han de realizar estudios para determinar si existen ventajas sobre los métodos de trabajo convencionales con muestras mínimas [4].

2. SECUENCIACIÓN DE NUEVA GENERACIÓN

La secuenciación de nueva generación (NGS) se refiere a una tecnología de secuenciación de alto rendimiento no basada en el método Sanger. Millones o billones de moléculas de ADN pueden ser secuenciadas al mismo tiempo. Este tipo de secuenciación incluye la secuenciación de segunda generación, basada en el análisis de un gran número de muestras simultáneamente, así como la secuenciación de tercera generación, basada en la determinación de la secuencia de ADN de una única molécula [36].

Las técnicas de secuenciación de segunda generación incluyen tres grandes mejoras sobre las técnicas convencionales. La primera ventaja es que no requieren clonado bacteriano de los fragmentos de ADN. La segunda es que se pueden realizar de miles a millones de reacciones de secuenciación a la vez. Y por último, los resultados se detectan directamente sin necesidad de electroforesis. No obstante, dependiendo de la tecnología los fragmentos obtenidos pueden ser relativamente cortos, por lo que podrían darse dificultades en el ensamblado y

anotación de las secuencias, así como el análisis bioinformático. Por otro lado, se necesita realizar una PCR para amplificar los fragmentos genómicos durante la preparación de las librerías [37].

Las técnicas de tercera generación no solo se centran en la detección de moléculas únicas sino que también permiten la secuenciación en tiempo real [36], alcanzando tamaños de lectura de hasta 1300 pb [38].

2.1 NGS en genética forense

Actualmente el método estándar en los análisis forenses es el genotipado de STRs y la secuenciación se ha limitado a la región control del ADN mitocondrial para establecer el origen y las relaciones entre muestras comprometidas. Mientras que la secuenciación completa del genoma mitocondrial mediante secuenciación de Sanger es una tarea laboriosa, se puede llevar a cabo fácilmente con tecnología NGS.

Un reciente estudio empleó secuenciación de alta cobertura para identificar heteroplasmias en ADN mitocondrial de células normales y en células tumorales. El análisis también reveló diferencias en el ADN mitocondrial de órganos diferentes de un mismo individuo [39]. Este hallazgo tiene grandes implicaciones forenses ya que una muestra dubitada y una muestra de referencia pueden proceder de diferentes tejidos [39]. Sin embargo, las altas tasas de error con estas tecnologías, de hasta un 2%, hacen que no se cumplan los estándares de calidad requeridos para el trabajo forense.

Aunque los STRs de rutina empleados tienen suficiente poder de discriminación para la gran mayoría de sus aplicaciones, el uso de la secuenciación a gran escala ofrece múltiples mejoras en los análisis forenses. La secuenciación puede identificar SNPs dentro y alrededor de los STRs que resultaría en un incremento del índice de parentesco [40,41].

Una opción interesante a desarrollar en un futuro es la sustitución de los STRs por marcadores más informativos derivados de la secuenciación de gran escala. Sin embargo, las bases de datos forenses ya existentes se basan en perfiles de STRs. Las NGS dan resultados bastante pobres en el análisis de STRs por lo que sustituir el genotipado de STRs por NGS necesitaría un gran trabajo de secuenciación para actualizar las bases de datos [40], y ya se están empezando a generar propuestas para la sustitución de una técnica por otra [42].

Por otro lado, la cantidad de información obtenida, la simplificación de los procedimientos de laboratorio y los beneficios al analizar muestras degradadas presente en pequeñas cantidades favorecen al cambio de STRs hacia tecnología NGS en genética forense. Recientes estudios de secuenciación del genoma de dos cromosomas Y separados entre ellos por 13 generaciones identifican al menos una mutación por generación, sugiriendo que cada individuo puede ser discriminado mediante secuenciación [43]. Estudios de asociación han identificado SNPs asociados a la etnia y a rasgos físicos. Esta información es muy importante en forense, ya que ayuda a la identificación de un criminal sin la necesidad de una muestra de referencia (Figura 1.3) [40].

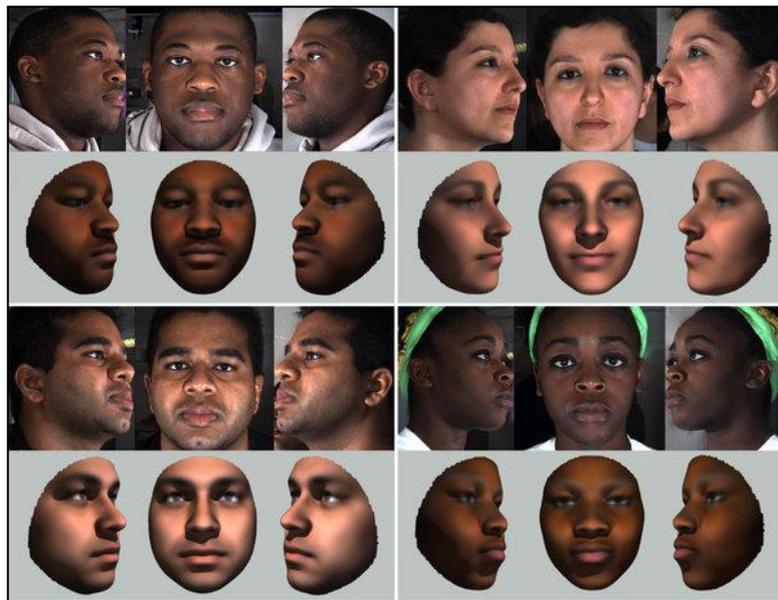


Figura 1.3: Predicciones de la apariencia de las personas a través del análisis de ADN comparadas con fotos reales de los mismos. Imagen tomada de The New York Times de Mark D. Shriver/Penn State University; 2015.

2.2 NGS en el estudio de la historia genética humana

Las técnicas de NGS han permitido el aislamiento y análisis del ADN de diversos animales extinguidos, incluyendo el mamut, el oso de las cavernas, el perezoso terrestre y el moa; que han ayudado en el conocimiento del modo de vida de estas especies [44–46].

El análisis del genoma mitocondrial completo del Neandertal sugiere de 1% a 4% de contribución de los mismos a las poblaciones actuales no africanas [47]. El análisis del ADN mitocondrial [48] y del genoma completo [49] de un homínido

desconocido de 30 000 a 50 000 años sugirió una nueva población, hermana de los neandertales, que se denominaron *denisovanos*.

El estudio de los datos de secuenciación de 31 regiones autosómicas no codificantes en tres poblaciones africanas históricamente aisladas indica un intercambio genético entre los humanos modernos y formas arcaicas de homínidos, que se dieron incluso antes de la salida de África [50].

Los estudios de secuenciación de genoma completo arrojarán nueva luz a este tipo de estudios. Un estudio basado en 379 genomas humanos secuenciados por técnicas de NGS procedentes de 125 poblaciones ha permitido aumentar el conocimiento disponible sobre los patrones de heterocigosidad en el continente africano, los flujos genéticos tanto a corta como a larga distancia, los fenómenos de mestizaje arcaicos y los cambios en el tamaño efectivo de las poblaciones [51].

2.3 NGS en genética de poblaciones

La información genética ofrece grandes conocimientos sobre el origen y la migración de las poblaciones y estudios genotípicos han revelado una subestructura geográfica distinta entre las poblaciones humanas [52]. Las tecnologías NGS permitirán el estudio genético de las poblaciones con una resolución sustancialmente mejor y la detección de las variaciones en una escala mucho menor [40,53].

En 2007 se lanzó el proyecto “1000 Genomes”, el primer intento de aplicar las NGS a la genética de poblaciones [54]. El proyecto “1000 Genomes” completo incluye: genoma secuenciado con baja cobertura, secuenciación del exoma y genotipado de 2500 individuos de 27 poblaciones de Europa, África, Asia y América. Actualmente, se está desarrollando un proyecto para la secuenciación de 10 000 genomas para la detección de genes relacionados con enfermedades raras [55]. Gracias a este tipo de análisis, con un elevado número de muestras y poblaciones, se encontrarán multitud de variables que podrán ser evaluadas para su uso como marcadores de ancestralidad (AIMs) y forense.

Recientes estudios con datos derivados de la secuenciación del genoma completo [56,57] han mostrado que las poblaciones europeas y chinas tienen trayectorias evolutivas similares, y que ambas poblaciones experimentaron un cuello de botella mucho más severo que las poblaciones africanas hace de 10 000 a 60 000 años [58]. Además, se determinó que Eurasia divergió de los africanos hace 38 000 a 64 000 años y que la población San del Sur de África divergió de otras poblaciones humanas mucho antes, 108 000 a 157 000 años [59]. A medida que

estén disponibles, cada vez, más secuencias genómicas de las diversas poblaciones humanas, será posible descifrar la historia humana con mucho más detalle [40,60].

3. ANTROPOLOGÍA MOLECULAR

La secuenciación del ADN mitocondrial por Anderson *y cols.* en 1981 [61], permitió encontrar diferencias entre poblaciones humanas. De forma simultánea, muchos autores comenzaron a trabajar en medidas de distancia y diversidad genética que permitieron establecer diferencias entre grupos humanos y la relación entre estos [62–65].

3.1 Polimorfismos

La antropología utiliza los distintos caracteres encontrados en el hombre como instrumento para describir cuantitativamente las relaciones entre individuos de una misma población y entre poblaciones diversas. Un polimorfismo es una diferencia genética encontrada en una posición del genoma al comparar los individuos de una población, de tal manera que la más rara de ellas no se puede mantener simplemente a través de la mutación periódica. Para que verdaderamente pueda considerarse un polimorfismo, la variación debe aparecer al menos en el 1% de la población [66]. Además, han de tener un origen único, ya que son ideales para reconstrucciones filogenéticas, y no estar bajo la influencia de fenómenos de selección [67].

3.2 Medidas de diversidad

La distancia genética define el grado en que dos poblaciones difieren en sus frecuencias alélicas. De esta forma, dos poblaciones que comparten un mismo origen pero han sufrido distinto desarrollo histórico, se irán diferenciando. Cuanto más tiempo sufran esta divergencia, mayor será la diferencia entre sus frecuencias génicas [62,68]. El cálculo de las distancias genéticas permite entender las relaciones evolutivas entre poblaciones [69,70].

Por otro lado, los índices de diversidad nos proporcionan la medida de la cantidad de variación en una población, siendo útiles para realizar comparaciones entre poblaciones a *grosso* modo. Los más empleados en el estudio de la variabilidad del ADNmt y del cromosoma Y en las poblaciones actuales son el

número de posiciones variables, el número de haplotipos diferentes, la diversidad genética (o haplotípica) de Nei [62], Reynolds [63] y Slatkin [64], la diversidad nucleotídica y el número medio de diferencias por parejas. La distribución de las diferencias entre parejas o *mismatch distribution* proporciona información sobre la diversidad de una muestra (Figura 1.4). Además, estas distribuciones permiten realizar inferencias acerca de la historia demográfica de las poblaciones. Se ha confirmado que episodios de crecimiento demográfico de declive o los cuellos de botella sufridos por una población dejan huellas características en este parámetro [71,72].

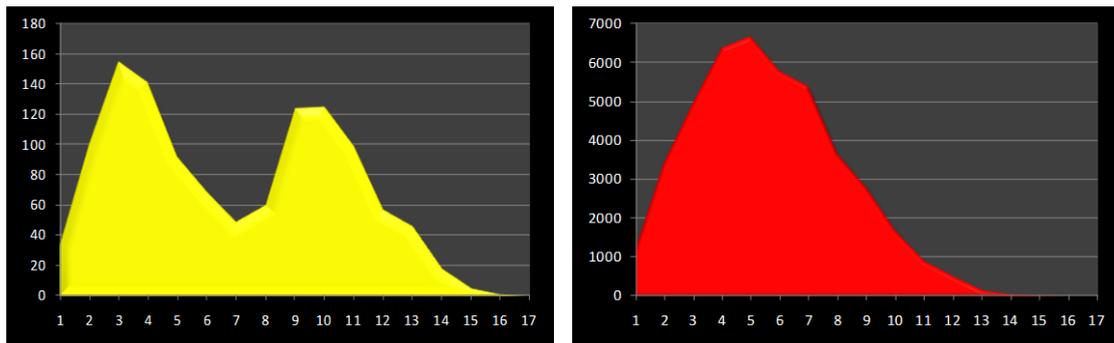


Figura 1.4: Distribución de las diferencias entre parejas. A la izquierda una población en equilibrio demográfico bimodal (Kaqchiquel de Guatemala) y a la derecha una población en expansión unimodal (población española). Imagen tomada de Martínez-González, 2011 (datos no publicados).

3.3 Antropología molecular en la Península Ibérica

Se han realizado múltiples estudios para determinar la procedencia y las relaciones de las diferentes poblaciones de la Península Ibérica con todos aquellos pueblos que influyeron en su nacimiento y formación.

De acuerdo con datos arqueológicos, muchos estudios apoyan la teoría de que el legado genético de Europa ha tenido dos grandes influencias. La primera tuvo lugar durante el Paleolítico Superior, cuando los hombres modernos entraron en el continente y la segunda, más reciente, ocurrió durante el Neolítico con la entrada de la agricultura. Las primeras culturas que llegaron a la Península a través de los Pirineos se establecieron en el norte de Cantabria, en Cataluña y en zonas mediterráneas. Posteriormente, estas culturas paleolíticas también alcanzaron la vertiente Atlántica.

A través del estudio de marcadores clásicos [2,73], estudios de ADNmt [74,75] y estudios de variación de cromosoma Y [76,77], se ha puesto en evidencia el

componente paleolítico presente en la población. Múltiples estudios muestran como la población del País Vasco diverge de otras poblaciones Europeas ya que tiene una gran proporción de alelos ancestrales procedentes del Paleolítico y una baja diversidad genética [78,79]. Se ha observado que tanto vascos como gallegos comparten altos valores de haplogrupo H mitocondrial, especialmente los subhaplogrupos H1 y H3 [80]

La geografía abrupta de la Península ha facilitado el aislamiento de las poblaciones humanas. Por ello se cree que la expansión de los pequeños núcleos aislados durante el Mesolítico también pudo haber afectado al origen genético de Europa. La cornisa cantábrica junto con el este de Andalucía y el levante español tuvieron un papel crucial como áreas de refugio durante el último periodo glacial (hace 20 000 años), así como una fuente de repoblación hacia Europa tras la mejora de las condiciones climáticas [80]. Los datos de haplogrupos de ADNmt apoyan esta teoría ya que se observa un gradiente sudoeste-noreste en las frecuencias de haplogrupos con tiempos de coalescencia postglacial como el H1, H3 y V [81–83]. Las frecuencias más altas para estos haplogrupos se encuentran en el norte de la Península Ibérica y el sudoeste de Francia [81,84]. Además, estudios de ADNmt establecen frecuencias similares de estos tres haplogrupos en el noreste africano y la Península Ibérica, estableciendo esta última como origen poblacional de los mismos en el norte de África. Estos datos, así como evidencias arqueológicas [85], sugieren la existencia de contactos prehistóricos entre las costas norte y sur del estrecho de Gibraltar establecidos tras el último periodo glacial [83].

Durante el periodo Neolítico, dos grandes olas migratorias afectaron a la Península; la primera, conocida como el Neolítico Andaluz, que se expandió desde Málaga hasta el sur de Portugal durante el sexto milenio AC, y una segunda conocida como el Neolítico Cardial, originado en el este del Mediterráneo, que llegó al levante ibérico sobre el 4700 AC [86]. Todos estos procesos migratorios han tenido gran influencia en la generación del gradiente geográfico de frecuencias observado en toda Europa. Por este motivo, la Península Ibérica, situada en el sudoeste del continente europeo, es considerada como la región con menor influencia Neolítica.

Durante la Edad de Cobre se llevaron a cabo grandes innovaciones tecnológicas que, junto a un aumento en la complejidad y estratificación social, dieron lugar a las primeras civilizaciones así como a extensos contactos poblacionales [86]. Con el comienzo de la Edad de Bronce, la tecnología del bronce se extiende por el sudeste de la Península hacia otras áreas y, a finales del periodo, se observan contactos con

otras civilizaciones, no solamente entre ellas sino con otras culturas como Gran Bretaña y Francia, entre otras.

La Edad de Hierro, que finaliza con la expansión del Imperio Romano hacia Hispania, se caracteriza por grandes cambios demográficos y sociales en la Península, ya que coincide con la fundación de colonias de poblaciones mediterráneas (fenicios, griegos y cartagineses). A finales de este periodo los grupos celtas se establecieron principalmente en el centro, el norte y el oeste de la Península mientras que la zona mediterránea estaba habitada por los íberos [86].

Posteriormente, la invasión celta y la influencia de su cultura en el Norte de la Península hicieron que los pueblos del norte quedaran aislados. Gracias al estudio del cromosoma Y se ha podido determinar la subestructura poblacional del Norte de España. Se ha comprobado que tanto la población de Galicia como la de Cantabria se encuentran genéticamente muy alejadas de otras poblaciones peninsulares. Todo esto se debe al gran aislamiento tanto geográfico como cultural, debido a sus raíces celtas, que han experimentado estas poblaciones dentro del conjunto de la Península Ibérica, así como a la baja tasa de inmigración presente en estas regiones [76,87]. Por otro lado, estudios de ADNmt consideran Galicia como un “borde genético” de la variación genética Europea [86,88]

Con la invasión de los pueblos del Norte de África en la Península en el siglo VIII se cierra el ciclo de influencias genéticas. Durante los casi 800 años que duró la invasión árabe en la Península se diseminaron por todo el territorio, exceptuando el País Vasco, Galicia, Cantabria, Asturias y la mayoría de los Pirineos; pero fue en el sur de la Península donde tuvieron una mayor influencia [76]. A través del estudio del ADNmt y del cromosoma Y, se han identificado regiones geográficas con una influencia genética del Norte de África del 8-10% [89-91]. Sin embargo, el estudio de marcadores autosómicos arroja valores menores de influencia norafricana (2,1-9%) (Figura 1.5) [89]. Estudios de mestizaje basados en la caracterización de secuencias ALU encuentran vestigios de genes subsaharianos en poblaciones del norte del Mediterráneo, lo que sugiere contactos continuos entre ambas costas [92]. El hecho de que se hayan detectado trazas genéticas, así como determinados haplotipos específicos a lo largo de toda la costa norte del Mediterráneo, reafirma la hipótesis de que este flujo genético está ligado a las primeras navegaciones transmediterráneas y se mantuvo homogéneo mientras perduró el comercio de esclavos hasta finales del S. XVII; más que tratarse del reflejo de la expansión islámica de los S. VII a S. XV [92]. No obstante, se observa un grado de heterogeneidad significativo a lo largo de todo el Mediterráneo, siendo relativamente más elevado en el sur [92]. A partir del cromosoma Y se ha descrito

la estructura poblacional de la Península Ibérica, calculando los distintos aportes por migración de musulmanes y judíos a la población actual de la Península [90], sobre todo centrados en el estudio del haplogrupo E3b2, común en el norte de África y presente en un 5,6% en la Península [93]. Se han detectado altos niveles de ascendencia patrilineal del norte de África y de judíos sefardíes en la población actual de la Península Ibérica [90]. Sin embargo, al contrario de lo que se podría esperar según los datos históricos que favorecería un gradiente sur-norte de influencia genética del norte de África; la mayor influencia genética del norte de África se encuentra en Galicia y el norte de Castilla (>20%). El principal gradiente de frecuencias genéticas norafricanas se observa entre el oeste y el este, siendo el último el que presenta menores proporciones. Tras la Reconquista, los moriscos se distribuyeron homogéneamente en la Península, pero su expulsión final en 1609 fue mucho más efectiva en algunas regiones de España, Valencia y oeste de Andalucía, mientras que en Galicia y Extremadura la población se dispersó y se integró en la sociedad [90].

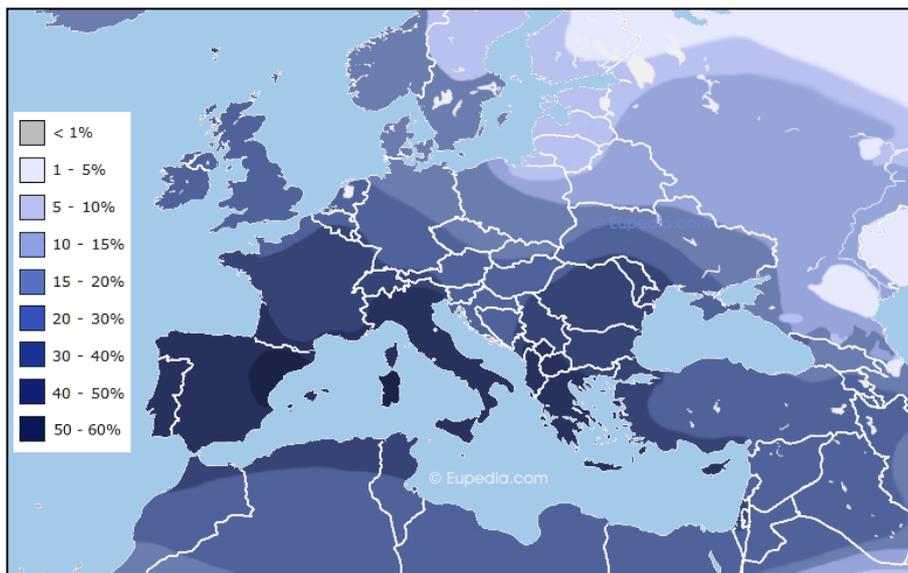


Figura 1.5: Gradiente del componente genético Mediterráneo en Europa, Medio Oriente y Norte de África basado en datos de STR autosómicos del Dodecad Project. Imagen tomada de *Eupedia.com*.

4. BIBLIOGRAFÍA

- [1] A.J. Jeffreys, V. Wilson, S.L. Thein, Individual-specific “fingerprints” of human DNA, *Nature*. 316 (1985) 76–79. doi:10.1038/316076a0.
- [2] L.L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes*, Princeton, NJ, 2007.
- [3] L. Roewer, DNA fingerprinting in forensics: past, present, future., *Investig. Genet.* 4 (2013) 22. doi:10.1186/2041-2223-4-22.
- [4] M.A. Jobling, P. Gill, Encoded evidence: DNA in forensic analysis, *Nat. Rev. Genet.* 5 (2004) 739–751.
- [5] M.L. Pontes, M. Fondevila, M.V. Laréu, R. Medeiros, SNP Markers as Additional Information to Resolve Complex Kinship Cases., *Transfus. Med. Hemotherapy Off. Organ Der Dtsch. Gesellschaft Für Transfusionsmedizin Und Immunhamatologie.* 42 (2015) 385–8. doi:10.1159/000440832.
- [6] S. Willuweit, L. Roewer, Y Chromosome Haplotype Reference Database, *Forensic Sci. Int. Genet.* 15 (2013) 43–48. doi:10.1016/j.fsigen.2014.11.024.
- [7] R. Szibor, X-chromosomal markers: past, present and future., *Forensic Sci. Int. Genet.* 1 (2007) 93–9. doi:10.1016/j.fsigen.2007.03.003.
- [8] L. Li, J. Ge, S. Zhang, J. Guo, S. Zhao, C. Li, et al., Maternity exclusion with a very high autosomal STRs kinship index., *Int. J. Legal Med.* 126 (2012) 645–8. doi:10.1007/s00414-012-0668-8.
- [9] H. Moreira, H. Costa, F. Tavares, L. Souto, Genetic variation of 12 X-chromosomal STR loci in an East Timor sample., *Int. J. Legal Med.* 129 (2015) 257–8. doi:10.1007/s00414-014-1126-6.
- [10] C. Bini, L.N. Riccardi, S. Ceccardi, F. Carano, S. Sarno, D. Luiselli, et al., Expanding X-chromosomal forensic haplotype frequencies database: Italian population data of four linkage groups, *Forensic Sci. Int. Genet.* 15 (2015) 127–130. doi:10.1016/j.fsigen.2014.11.008.
- [11] A. Ibarra, T. Restrepo, W. Rojas, A. Castillo, A. Amorim, B. Martínez, et al., Evaluating the X chromosome-specific diversity of Colombian populations using insertion/deletion polymorphisms., *PLoS One.* 9 (2014) e87202. doi:10.1371/journal.pone.0087202.
- [12] K. Rębała, S.A. Kotova, V.I. Rybakova, T. V Zabauskaya, A.A. Shyla, A.A. Spivak, et al., Variation of X-chromosomal microsatellites in Belarus within the context of their genetic diversity in Europe., *Forensic Sci. Int. Genet.* 16C (2015) 105–111. doi:10.1016/j.fsigen.2014.12.011.
- [13] C. Santos, C. Phillips, M. Fondevila, R. Daniel, R.A.H. van Oorschot, E.G. Burchard, et al., Pacifplex: an ancestry-informative SNP panel centred on Australia and the Pacific region, *Forensic Sci. Int. Genet.* 20 (2016) 71–80. doi:10.1016/j.fsigen.2015.10.003.
- [14] M. de la Puente, C. Santos, M. Fondevila, L. Manzo, Á. Carracedo, M. V Lareu, et al., The Global AIMs Nano set: A 31-plex SNaPshot assay of ancestry-informative SNPs., *Forensic Sci. Int. Genet.* 22 (2016) 81–88. doi:10.1016/j.fsigen.2016.01.015.
- [15] O. Bulbul, G. Filoglu, T. Zorlu, H. Altuncul, A. Freire-Aradas, J. Söchtig, et al., Inference of biogeographical ancestry across central regions of Eurasia., *Int. J. Legal Med.* 130 (2016) 73–9. doi:10.1007/s00414-015-1246-7.

- [16] M. Kayser, P. de Knijff, Improving human forensics through advances in genetics, genomics and molecular biology, *Nat. Rev. Genet.* 12 (2011) 179–192.
- [17] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, et al., A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages., *Forensic Sci. Int. Genet.* 6 (2012) 208–18. doi:10.1016/j.fsigen.2011.04.017.
- [18] K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, Q. Ayub, et al., Toward male individualization with rapidly mutating y-chromosomal short tandem repeats., *Hum. Mutat.* 35 (2014) 1021–32. doi:10.1002/humu.22599.
- [19] S. Turrina, S. Caratti, M. Ferriani, D. De Leo, Are rapidly mutating Y-short tandem repeats useful to resolve a lineage? Expanding mutability data on distant male relationships., *Transfusion.* 56 (2016) 533–8. doi:10.1111/trf.13368.
- [20] P. Sulem, D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, K.P. Magnusson, et al., Genetic determinants of hair, eye and skin pigmentation in Europeans., *Nat. Genet.* 39 (2007) 1443–1452. doi:10.1038/ng.2007.13.
- [21] S.I. Candille, D.M. Absher, S. Beleza, M. Bauchet, B. McEvoy, N.A. Garrison, et al., Genome-Wide Association Studies of Quantitatively Measured Skin, Hair, and Eye Pigmentation in Four European Populations, *PLoS One.* 7 (2012). doi:10.1371/journal.pone.0048294.
- [22] J. Söchtig, C. Phillips, O. Maroñas, A. Gómez-Tato, R. Cruz, J. Alvarez-Dios, et al., Exploration of SNP variants affecting hair colour prediction in Europeans., *Int. J. Legal Med.* 129 (2015) 963–75. doi:10.1007/s00414-015-1226-y.
- [23] H. Lango Allen, K. Estrada, G. Lettre, S.I. Berndt, M.N. Weedon, F. Rivadeneira, et al., Hundreds of variants clustered in genomic loci and biological pathways affect human height., *Nature.* 467 (2010) 832–838. doi:10.1038/nature09410.
- [24] S.E. Medland, D.R. Nyholt, J.N. Painter, B.P. McEvoy, A.F. McRae, G. Zhu, et al., Common Variants in the Trichohyalin Gene Are Associated with Straight Hair in Europeans, *Am. J. Hum. Genet.* 85 (2009) 750–755. doi:10.1016/j.ajhg.2009.10.009.
- [25] A. Fujimoto, R. Kimura, J. Ohashi, K. Omi, R. Yuliwulandari, L. Batubara, et al., A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness, *Hum. Mol. Genet.* 17 (2008) 835–843. doi:10.1093/hmg/ddm355.
- [26] E. Pośpiech, J. Karłowska-Pik, M. Marcińska, S. Abidi, J.D. Andersen, M. van den Berge, et al., Evaluation of the predictive capacity of DNA variants associated with straight hair in Europeans., *Forensic Sci. Int. Genet.* 19 (2015) 280–8. doi:10.1016/j.fsigen.2015.09.004.
- [27] S. Walsh, F. Liu, K.N. Ballantyne, M. Van Oven, O. Lao, M. Kayser, IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information, *Forensic Sci. Int. Genet.* 5 (2011) 170–180. doi:10.1016/j.fsigen.2010.02.004.
- [28] E. Sauer, A.-K. Reinke, C. Courts, Differentiation of five body fluids from forensic samples by expression analysis of four microRNAs using quantitative PCR., *Forensic Sci. Int. Genet.* 22 (2016) 89–99. doi:10.1016/j.fsigen.2016.01.018.
- [29] M.T. Machado, S. Navega, F. Dias, M.J.C. de Sousa, A.L. Teixeira, R. Medeiros, microRNAs for peripheral blood fraction identification: Origin, pathways and forensic relevance., *Life Sci.* 143 (2015) 98–104. doi:10.1016/j.lfs.2015.10.029.
- [30] H.Y. Lee, M.J. Park, A. Choi, J.H. An, W.I. Yang, K.J. Shin, Potential forensic application of DNA methylation profiling to body fluid identification, *Int. J. Legal Med.* 126

- (2012) 55–62. doi:10.1007/s00414-011-0569-2.
- [31] U. Tuckprakhon, K. Lomthaisong, S. Daduang, K. Lomthaisong, Preliminary study of alpha-globin mRNA as a molecular tool for age estimation of human bloodstains, *Walailak J. Sci. Technol.* 13 (2016) 281–290.
- [32] K. Lech, K. Ackermann, A. Wollstein, V.L. Revell, D.J. Skene, M. Kayser, Assessing the suitability of miRNA-142-5p and miRNA-541 for bloodstain deposition timing, *Forensic Sci. Int. Genet.* 12 (2014) 181–4. doi:10.1016/j.fsigen.2014.06.008.
- [33] K. Ackermann, K.N. Ballantyne, M. Kayser, Estimating trace deposition time with circadian biomarkers: a prospective and versatile tool for crime scene reconstruction, *Int. J. Legal Med.* 124 (2010) 387–95. doi:10.1007/s00414-010-0457-1.
- [34] K. Lech, F. Liu, K. Ackermann, V.L. Revell, O. Lao, D.J. Skene, et al., Evaluation of mRNA markers for estimating blood deposition time: Towards alibi testing from human forensic stains with rhythmic biomarkers, *Forensic Sci. Int. Genet.* 21 (2016) 119–25. doi:10.1016/j.fsigen.2015.12.008.
- [35] D. Pushkarev, N.F. Neff, S.R. Quake, Single-molecule sequencing of an individual human genome, *Nat. Biotechnol.* 27 (2009) 847–850.
- [36] Y. Yang, B. Xie, J. Yan, Application of Next-generation Sequencing Technology in Forensic Science, *Genomics. Proteomics Bioinformatics.* 12 (2014) 190–197. doi:10.1016/j.gpb.2014.09.001.
- [37] E.L. van Dijk, H. Auger, Y. Jaszczyszyn, C. Thermes, Ten years of next-generation sequencing technology, *Trends Genet.* 30 (2014). doi:10.1016/j.tig.2014.07.001.
- [38] L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, et al., Comparison of next-generation sequencing systems, *J. Biomed. Biotechnol.* 2012 (2012) 251364.
- [39] Y. He, J. Wu, D.C. Dressman, C. Iacobuzio-Donahue, S.D. Markowitz, V.E. Velculescu, et al., Heteroplasmic mitochondrial DNA mutations in normal and tumour cells, *Nature.* 464 (2010) 610–614. doi:10.1038/nature08802.
- [40] E.C. Berglund, A. Kiialainen, A.C. Syvänen, Next-generation sequencing technologies and applications for human genetic history and forensics, *Investig. Genet.* 2 (2011) 23.
- [41] K.B. Gettings, K.M. Kiesler, S.A. Faith, E. Montano, C.H. Baker, B.A. Young, et al., Sequence variation of 22 autosomal STR loci detected by next generation sequencing, *Forensic Sci. Int. Genet.* 21 (2016) 15–21. doi:10.1016/j.fsigen.2015.11.005.
- [42] C. Van Neste, W. Van Criekinge, D. Deforce, F. Van Nieuwerburgh, Forensic Loci Allele Database (FLAD): Automatically generated, permanent identifiers for sequenced forensic alleles, *Forensic Sci. Int. Genet.* 20 (2016) e1–3. doi:10.1016/j.fsigen.2015.09.006.
- [43] Y. Xue, Q. Wang, Q. Long, B.L. Ng, H. Swerdlow, J. Burton, et al., Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree, *Curr. Biol.* 19 (2009) 1453–1457. doi:10.1016/j.cub.2009.07.032.
- [44] A.D. Greenwood, J. Castresana, G. Feldmaier-Fuchs, S. Pääbo, A molecular phylogeny of two extinct sloths, *Mol. Phylogenet. Evol.* 18 (2001) 94–103.
- [45] J.P. Noonan, M. Hofreiter, D. Smith, J.R. Priest, N. Rohland, G. Rabeder, et al., Genomic sequencing of Pleistocene cave bears, *Science.* 309 (2005) 597–599. doi:10.1126/science.1113485.

- [46] W. Miller, D.I. Drautz, A. Ratan, B. Pusey, J. Qi, A.M. Lesk, et al., Sequencing the nuclear genome of the extinct woolly mammoth., *Nature*. 456 (2008) 387–390. doi:10.1038/nature07446.
- [47] R.E. Green, J. Krause, A.W. Briggs, T. Maricic, U. Stenzel, M. Kircher, et al., A draft sequence of the Neandertal genome, *Science* (80-.). 328 (2010) 710–722.
- [48] J. Krause, Q. Fu, J.M. Good, B. Viola, M. V Shunkov, A.P. Derevianko, et al., The complete mitochondrial DNA genome of an unknown hominin from southern Siberia, *Nature*. 464 (2010) 894–897.
- [49] D. Reich, R.E. Green, M. Kircher, J. Krause, N. Patterson, E.Y. Durand, et al., Genetic history of an archaic hominin group from Denisova Cave in Siberia, *Nature*. 468 (2010) 1053–1060.
- [50] M.F. Hammer, A.E. Woerner, F.L. Mendez, J.C. Watkins, J.D. Wall, Genetic evidence for archaic admixture in Africa., *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 15123–15128. doi:10.1073/pnas.1109300108.
- [51] L. Pagani, J. Lawson, E. Jagoda, A. Mörseburg, F. Clemente, G. Hudjashov, et al., Geographical barriers, environmental challenges, and complex migration events during the peopling of Eurasia, *Nature*. (2016).
- [52] J.Z. Li, D.M. Absher, H. Tang, A.M. Southwick, A.M. Casto, S. Ramachandran, et al., Worldwide human relationships inferred from genome-wide patterns of variation., *Science*. 319 (2008) 1100–4. doi:10.1126/science.1153717.
- [53] S.M. Kim, S.Y. Yoo, S.H. Nam, J.M. Lee, K.W. Chung, Identification of Korean-specific SNP markers from whole-exome sequencing data., *Int. J. Legal Med.* (2016) 1–9. doi:10.1007/s00414-015-1313-0.
- [54] J. Kaiser, DNA sequencing. A plan to capture human diversity in 1000 genomes., *Science*. 319 (2008) 395. doi:10.1126/science.319.5862.395.
- [55] N. Moran, 10,000 rare-disease genomes sequenced, *Nat Biotech*. 32 (2014) 7.
- [56] D.L. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, A. Chakravarti, A.G. Clark, et al., A map of human genome variation from population-scale sequencing, *Nature*. 467 (2010) 1061–1073.
- [57] A. Bhaskar, Y.X.R. Wang, Y.S. Song, Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data, *bioRxiv*. (2014).
- [58] H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences., *Nature*. 475 (2011) 493–496. doi:10.1038/nature10231.
- [59] I. Gronau, M.J. Hubisz, B. Gulko, C.G. Danko, A. Siepel, Bayesian inference of ancient human demography from individual genome sequences, *Nat. Genet.* 43 (2011) 1031–1034. doi:10.1038/ng.937.
- [60] I. Pugach, M. Stoneking, Genome-wide insights into the genetic history of human populations, *Investig. Genet.* 6 (2015) 1–20. doi:10.1186/s13323-015-0024-0.
- [61] S. Anderson, A.T. Bankier, B.G. Barrell, Sequence and organization of the human mitochondrial genome, *Nature*. 290 (1981) 457–465.
- [62] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees., *Mol. Biol. Evol.* 4 (1987) 406–425.
- [63] J. Reynolds, B.S. Weir, C.C. Cockerham, Estimation for the coancestry coefficient: Basis for a short-term genetic distance, *Genetics*. 105 (1983) 769–779.
- [64] M. Slatkin, Linkage disequilibrium in growing and stable populations, *Genetics*. 137 (1994) 331–336.

- [65] L.L. Cavalli-Sforza, M.W. Feldman, The application of molecular genetic approaches to the study of human diversity, *Nat.Genet.* 33 (2003) 267.
- [66] R. Sachidanandam, D. Weissman, S.C. Schmidt, J.M. Kakol, L.D. Stein, G. Marth, et al., A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms, *Nature.* 409 (2001) 928–933.
- [67] M. Jobling, M. Hurles, C. Tyler-Smith, *Human Evolutionary Genetics: Origins, People and Disease*, Garland Science/Taylor & Francis Group, New York, 2003.
- [68] M. Nei, Genetic Distance, in: S. Brenner, J.H. Miller (Eds.), *Encycl. Genet.*, Academic Press, New York, 2001: pp. 828–832. doi:10.1006/rwgn.2001.0532.
- [69] Y. Nagamine, M. Higuchi, Genetic distance and classification of domestic animals using genetic markers, *J. Anim. Breed. Genet.* 118 (2001) 101–109.
- [70] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure, *Evolution (N. Y.)* 38 (1984) 1358–1370. doi:10.2307/2408641.
- [71] H.C. Harpending, Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution., *Hum. Biol. an Int. Rec. Res.* 66 (1994) 591–600.
- [72] A. Márcia Barbosa, R. Real, A.R. Muñoz, J.A. Brown, New measures for assessing model equilibrium and prediction mismatch in species distribution models, *Divers. Distrib.* 19 (2013) 1333–1338. doi:10.1111/ddi.12100.
- [73] J. Bertranpetit, L.L. Cavalli-Sforza, A genetic reconstruction of the history of the population of the Iberian Peninsula, *Ann. Hum. Genet.* 55 (1991) 51–67.
- [74] M. Richards, V. Macaulay, E. Hickey, E. Vega, B. Sykes, Tracing European founder lineages in the Near Eastern mtDNA pool, *Am J Hum Genet.* 62 (2000) 241–260.
- [75] H.B.S.M. Côrte-Real, V.A. Macaulay, M.B. Richards, G. Hariti, M.S. Issad, A. Cambon-Thomsen, et al., Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis, *Ann. Hum. Genet.* 60 (1996) 331–350.
- [76] M. Brion, A. Salas, A. González-Neira, M. V Lareu, A. Carracedo, Insights into Iberian population origins through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite, *Am. J. Phys. Anthropol.* 122 (2003) 147–161. doi:10.1002/ajpa.10231.
- [77] O. Semino, G. Passarino, P.J. Oefner, A.A. Lin, S. Arbuzova, L.E. Beckman, et al., The genetic legacy of paleolithic Homo sapiens sapiens in extant europeans: A Y chromosome perspective, *Science (80-)*. 290 (2000) 1155–1159.
- [78] N. Rodríguez-Ezpeleta, J. Álvarez-Busto, L. Imaz, M. Regueiro, M.N. Azcárate, R. Bilbao, et al., High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations, *Hum. Genet.* 128 (2010) 113–117.
- [79] M. Iriondo, M.C. Barbero, C. Manzano, DNA polymorphisms detect ancient barriers to gene flow in Basques., *Am. J. Phys. Anthropol.* 122 (2003) 73–84. doi:10.1002/ajpa.10212.
- [80] C.L. Hernández, G. Reales, J.-M. Dugoujon, A. Novelletto, J.N. Rodríguez, P. Cuesta, et al., 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 (2014) 11. doi:10.1186/1471-2156-15-11.
- [81] A.F. Pardiñas, A. Roca, E. Garcia-Vazquez, B. Lopez, Mitochondrial diversity patterns and the Magdalenian resettlement of Europe: new insights from the edge

- of the Franco-Cantabrian refuge., *J. Hum. Genet.* 57 (2012) 717–26. doi:10.1038/jhg.2012.100.
- [82] A. Achilli, C. Rengo, C. Magri, V. Battaglia, A. Olivieri, R. Scozzari, et al., 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 (2004) 910–918.
- [83] T. Rhouda, D. Martínez-Redondo, A. Gómez-Durán, N. Elmtili, M. Idaomar, C. Díez-Sánchez, et al., Moroccan mitochondrial genetic background suggests prehistoric human migrations across the Gibraltar Strait, *Mitochondrion.* 9 (2009) 402–7. doi:10.1016/j.mito.2009.07.003.
- [84] C. Gamble, *The Palaeolithic Settlement of Europe*, Cambridge University Press, 1986.
- [85] J.L.E. Carrasco, *Applications of Evolutive Archeology: Migrations from Africa to Iberia in the Recent Prehistory*, in: A. Arnaiz-Villena, J. Martínez-Laso, E. Gómez-Casado (Eds.), *Prehist. Iberia SE - 6*, Springer US, 2000: pp. 125–162. doi:10.1007/978-1-4615-4231-5_6.
- [86] C. Santos, R. Fregel, V.M. Cabrera, L. Alvarez, J.M. Larruga, A. Ramos, et al., Mitochondrial DNA and Y-chromosome structure at the Mediterranean and Atlantic façades of the Iberian Peninsula., *Am. J. Hum. Biol.* 26 (2014) 130–41. doi:10.1002/ajhb.22497.
- [87] M. Brion, B. Quintans, M. Zarrabeitia, A. Gonzalez-Neira, A. Salas, V. Lareu, et al., Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis, *Gene.* 329 (2004) 17–25.
- [88] A. Salas, D. Comas, M. V Lareu, J. Bertranpetit, A. Carracedo, mtDNA analysis of the Galician population: a genetic edge of European variation, *Eur. J. Hum. Genet.* 6 (1998) 365–375.
- [89] M. Pino-Yanes, A. Corrales, S. Basaldúa, A. Hernández, L. Guerra, J. Villar, et al., North African influences and potential bias in case-control association studies in the Spanish population., *PLoS One.* 6 (2011) e18389. doi:10.1371/journal.pone.0018389.
- [90] S.M. Adams, E. Bosch, P.L. Balaesque, S.J. Ballereau, A.C. Lee, E. Arroyo, et al., The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula., *Am. J. Hum. Genet.* 83 (2008) 725–36. doi:10.1016/j.ajhg.2008.11.007.
- [91] N. Maca-Meyer, P. Sánchez-Velasco, C. Flores, J.-M. Larruga, A.-M. González, A. Oterino, et al., Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain), *Ann. Hum. Genet.* 67 (2003) 329–339.
- [92] E. González-Pérez, E. Esteban, M. Via, M. Gayà-Vidal, G. Athanasiadis, J.M. Dugoujon, et al., Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems), *Am. J. Phys. Anthropol.* 141 (2010) 430–9. doi:10.1002/ajpa.21161.
- [93] C. Flores, N. Maca-Meyer, A.M. González, P.J. Oefner, P. Shen, J.A. Pérez, et al., Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography., *Eur. J. Hum. Genet.* 12 (2004) 855–63. doi:10.1038/sj.ejhg.5201225.

ABSTRACT

The former Kingdom of Granada, comprising the provinces of Granada, Málaga and Almería (GMA), was once inhabited for more than 700 years by North African population who influenced its creation and establishment. However, after the Reconquest, Moorish were strongly prosecuted and isolated until they were totally expelled in 1609. Families from the North of the Iberian Peninsula came to repopulate the area.

During the 700 years that North African people inhabited the territory, their influence had a great impact in culture, some traditions as personal hygiene and the forerunner of the Spanish guitar; art and architecture, as it can be seen in many monuments, Alhambra, Mezquita of Córdoba; language, not only in many words but also in idioms; toponymy, cities, villages, rivers...; gastronomy, many food products were introduced as far as how they are cooked.

A total of 245 DNA samples were obtained through informed consent from unrelated individuals from the provinces of Granada, Málaga and Almería for at least three generations. The birthplace and surnames of all the 245 unrelated individuals were queried and annotated in order to determine the origin of each surname and the regions where more bearers are born.

Meanwhile only the 7% of the surnames studied have an Andalusian origin, 37% of the surnames are mostly present in this region. Surname analysis reflected the great isolation of Moorish and the scarce contact between them and the new population. While, most of the surnames have a Castilian or Galician origin; the most of them are more frequent on the South of Spain.

1. INTRODUCCIÓN

El antiguo reino de Granada estaba formado por las actuales provincias de Granada, Málaga y Almería con alguna parte de Cádiz, Jaén, Córdoba y Sevilla. La capital nazarí, Granada, se convirtió en los siglos XIV y XV en una de las ciudades más prósperas de Europa. Era un centro comercial y cultural de primer orden que llegó a contar con unos 165 000 habitantes [1].

1.1 Historia del Reino de Granada

1.1.1. Orígenes

Los orígenes de Granada son muy imprecisos, pero parece que se pueden fijar en las primitivas tribus íberas (túrdulos) que fundaron **Iliberir**. Durante el período cartaginés hubo una convivencia pacífica con los túrdulos, que colaboraron con los cartagineses en la Primera Guerra Púnica contra Roma. Tras la derrota de los cartagineses, el triunfo de Roma tuvo como consecuencia el sometimiento de las tribus ibéricas, aunque parece que durante un tiempo **Iliberis** tuvo estatus de ciudad libre [2,3].

La ciudad mantenía tres enclaves: **Iliberis** ubicado en la colina del Albaicín, **Garnatha**, situado en la colina frente al Albaicín, que después fue poblada principalmente por judíos y sobre la que se edificó la Alhambra. El tercer enclave es **Castilia** situada en los que más tarde sería Sierra Elvira.

Finalizado el periodo romano a principios del siglo V, vándalos, suevos-silingos y alanos invadieron Hispania. Los visigodos se incorporaron a la conquista de Hispania y expulsaron a las tribus de suevos-silingos y alanos al Norte de África. Por último, los bizantinos y romanos también fueron expulsados del territorio hispano. Los visigodos se establecieron en Hispania durante 300 años (415-711) [4,5].

1.1.2. Conquista musulmana

En el año 711, el caudillo bereber *Tariq ibn Ziyad* sometió **Iliberis** con ayuda de los judíos, derrotando a los visigodos. En el 716, *Abd-al-Aziz ibn Musá* conquistó definitivamente todos los territorios del sudeste peninsular. En el año 740 hubo otra rebelión de los bereberes liderados por *Mysara*, en la que se apoderaron de

Tánger. Esto llevó a los bereberes afincados en España a revelarse y reconquistar los territorios que les habían sido arrebatados. Como la España musulmana era una provincia del califato sirio, un ejército se puso en camino para derrotar y expulsar a los bereberes de España, instalando a su jefe en Córdoba en el 741.

Con la desaparición del califato de Córdoba y la creación de los primeros reinos de Taifas, surgieron tres grupos sociales dominantes: *andalusíes*, musulmanes de origen árabe e hispano; *bereberes* y *eslavos*, antiguos esclavos-eunucos. Además estaban los grandes núcleos andalusíes autóctonos, mozárabes (cristianos que vivían en Al-Ándalus) y muladíes (indígenas que habían aceptado el Islam); y las comunidades judías [4].

1.1.3. La dinastía zirí

A la caída y desmembración del califato y el surgimiento de los **primeros reinos de los Taifas**, los bereberes emigraron de al-Ándalus y pasaron a Berbería. Dos años más tarde, en el 1013, llegó desde África la dinastía **Zirí**, fundada por *Zawi ibn Zirí*, a **Ilbira (Elvira)**. No puede considerarse como reino, ya que desde la época romana se comportaba como ciudad libre. No fue hasta 1025, cuando los habitantes de Elvira decidieron solicitar la protección de *Zawi ibn Zirí*, y convertirse en sus vasallos a cambio de que les protegiesen de los constantes ataques de las taifas cercanas. Frente a esta situación de constantes invasiones, se propuso la construcción de una ciudad fortificada en una altura que dominase el territorio, destruyendo la antigua ciudad de Elvira. A partir de entonces, el asentamiento bereber se consolidó y se inició el engrandecimiento y poderío del reino de Granada.

Hacia el 1025 *Zawi ibn Zirí* decidió abandonar la Península dejando el reino a su sobrino. A la muerte del soberano, su hijo *Badis ibn Habus al-Muzaffar* le sucedió en el mandato. En 1038, *Badis ibn Habus al-Muzaffar* derrotó al soberano almeriense anexionando así todo el territorio de Almería y aumentando el poderío del reino de Granada. Tras derrotar a *ibn Abbad* de Sevilla, *Badis ibn Habus al-Muzaffar* tomó conciencia de su poder y se apoderó del principado de Málaga, desterrando a los príncipes a África en 1057 [4,5].

1.1.4. Los almorávides

Tras la toma de Toledo por *Alfonso VI* en 1085 y sus intenciones de continuar conquistando hacia el sur de la Península, *al-Mutamid* de Sevilla y *Abd-Allah* de

Granada decidieron pedir ayuda a los almorávides. Juntos derrotaron al ejército cristiano de *Alfonso VI* en la batalla de Sagrajas (Badajoz). Tras el derrocamiento de *Abd-Allah* en 1090, *Yusif ibn Tasufin* tomó Almería, Sevilla en 1091 y el reino de Badajoz en 1095. Una vez conquistado todo el territorio de al-Ándalus por los almorávides, emprendieron la tarea de atacar a los cristianos.

El régimen rudo político-militar de los almorávides se fue deteriorando paulatinamente, dejando de existir como Imperio hacia el 1145 al ser atacados por los almohades en África. Este hecho da como resultado el resurgimiento por segunda vez de los reinos de Taifas [4,5].

1.1.5. Dominación almohade

Tras la desmembración del imperio almorávide surgió un periodo anárquico capitaneado por familias de nuevos ricos, los **Segundos Reinos de Taifas**. La dominación almohade (1146-1192/1224) terminó con este periodo anárquico. El ocaso del poderío almohade se dió en 1224 con la aparición de los **Terceros Reinos de Taifas**.

Tras la batalla de Navas de Tolosa y el libre acceso al valle del Guadalquivir, las terceras taifas tuvieron una vida efímera debido a las conquistas de *Fernando III el Santo* con Sevilla, y *Alfonso X el Sabio* con el valle del Guadalquivir. De esta manera, el al-Ándalus islámico quedó reducido a los límites del reino nazarí de Granada [4,5].

1.1.6. La dinastía nasri o nazarí

Después de la derrota almohada en las Navas de Tolosa, *Muhammad abu Abd-Allah ibn Yusuf ibn Nasar de Arjona* conquistó las ciudades de Alhama, Loja, Almería, Málaga y Granada constituyendo así el reino de Granada. Al morir *Muhammad I*, su hijo *Muhammad II* le sucedió en el trono y renovó los pactos de su padre con *Alfonso X* de Castilla. Después, en 1275, *Muhammad II*, se alió con pueblos del norte de Marruecos para combatir a los cristianos y recuperar las tierras que les habían arrebatado. Al morir, su hijo *Abd-Allah* comenzó su reinado firmando una paz con Jaime de Aragón y dirigiendo su ejército contra Castilla, pero fue derrotado y tuvo que reconocerse vasallo de *Fernando IV*. Este pacto de vasallaje no fue bien visto por la corte de Granada, provocando una sublevación a favor del príncipe *Nazar*, hermano de *Muhammad III*.

El príncipe *Nazar* accedió al trono en un momento de crisis que no consiguió resolver y que agravó el malestar entre sus partidarios, siendo expulsado del trono por su sobrino *Ismail* en 1312, lo que incrementó la persecución de los cristianos. Tras su muerte en 1324, su hijo *Muhammad IV* tomó el poder del reino, ejerciendo el gobierno a través de varios partidarios suyos, lo que originó una lucha por el poder. *Alfonso XI* aprovechó esta situación para apoderarse de algunas plazas fronterizas. *Muhammad IV* reaccionó apoderándose de Cabra y Priego.

Yusuf I, hermano de *Muhammad IV*, tomó el trono en 1333. Durante su mandato estableció relaciones con los merinidas, interviniendo activamente en sus contiendas. Durante su reinado hubo un extraordinario desarrollo cultural, fundándose la escuela coránica granadina de La Madraza en 1349. A su muerte, su hijo *Muhammad V* procuró la paz con los cristianos peninsulares y con el Emir de Fez, pero su hermano *Ismail* le arrebató el trono en 1358 teniendo que refugiarse en Guadix.

Ismail II se mantuvo en el trono hasta 1360 en que fue asesinado por *Muhammad VI*. *Muhammad V* pidió ayuda a Pedro I de Castilla e inició una guerra contra el usurpador. Con el apoyo de la población, *Muhammad V* pudo ocupar el trono de Granada sin dificultad. Firmó una tregua con Pedro I de Castilla que duró 20 años durante los cuales el reino de Granada alcanzó su máximo esplendor cultural y comercial. Es considerado el constructor de la Alhambra, realmente iniciada en la época zirí, ya que se le atribuyen la mayor parte de las edificaciones.



Figura 2.1. Foto de la Alhambra, edificada durante el reinado de la dinastía Nazarí.

Yusuf II sucedió a su padre *Muhammad V* y contribuyó al enriquecimiento cultural del reino de Granada. Renovó la tregua con Enrique II de Castilla en 1395. Murió envenenado por orden de su hijo, que renovó tratados de amistad con

Enrique III de Castilla. Al morir en 1408, su hermano mayor *Yusuf* tomó el trono. A su muerte se iniciaron unos años de gran confusión, entre 1423 y 1447, que se agravaron en la segunda mitad del siglo XV, que debilitaron el reino, facilitando su conquista por los Reyes Católicos.

El principio del fin del reino de Granada se dió con *Muhammad VII*. En 1427, su primo *al-Zaquir (el Pequeño)* le arrebató el trono y se proclamó rey como *Muhammad VIII*. *Muhammad VII* se exilió a la corte del rey de Túnez. Con la ayuda de éste y de Juan II de Castilla derrotó a *Muhammad VIII*, pero se negó a pagar la fuerte recompensa exigida y se declaró la guerra. El rey de Castilla favoreció a *Yusuf IV* para que subiera al trono pero sólo estuvo en el poder durante seis meses. En 1452, *Muhammad X*, entra en Granada como rey gracias al apoyo de Juan II de Castilla. Firmó vasallaje con Castilla hasta ser destituido por su hijo *Mulay Hacén*, en 1466. Ésta fue una época de grandes luchas internas en el reino entre *Mulay Hacén*, su hijo *Boabdil* y su hermano *Abd-Allah el Zagal*. En 1489, los Reyes Católicos atacaron conquistando varias plazas. Ante estos hechos, *el Zagal* tomó conciencia de su impotencia y pactó con los Reyes Católicos la venta de las ciudades que permanecían fieles, excepto la de Granada, de la que se había apoderado su sobrino *Boabdil* durante su ausencia.

Muhammad XII o *Boabdil el Chico* fue el último rey de Granada. La población de Granada, tras la pérdida de Baza, se amotinó contra *Boabdil*. Entretanto, los Reyes Católicos habían dado un ultimátum al rey nazarí para que entregase la última plaza, a lo que *Boabdil* se negó dando lugar al asedio de Granada y a la entrega de la ciudad el 2 de Enero de 1492. *Boabdil* se retiró a Fez donde murió [4,5].

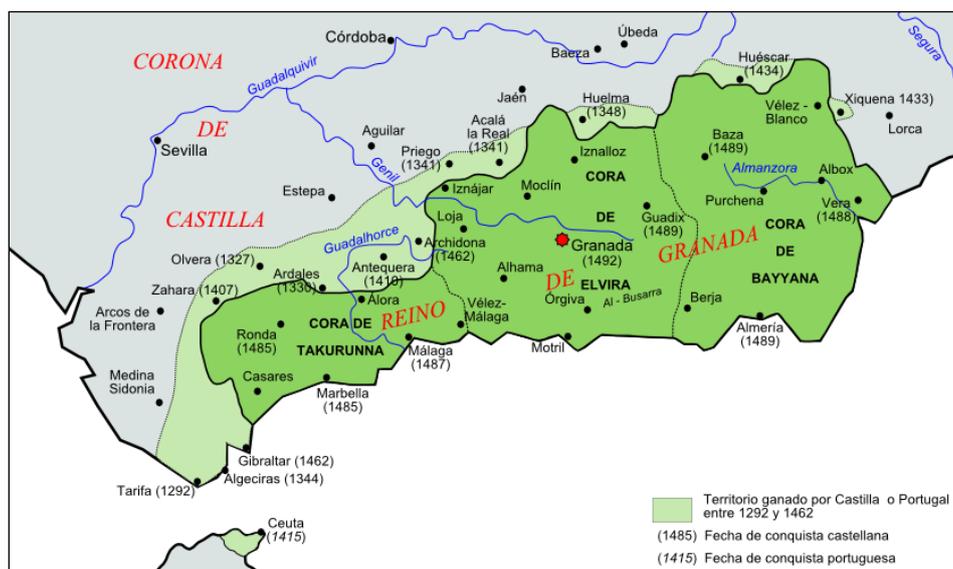


Figura 2.2. Mapa del territorio que ocupaba el reino nazarí de Granada.

1.1.7. El reino cristiano de Granada

La entrada de los Reyes Católicos en la ciudad tuvo lugar el 6 de Enero de 1492. Una vez la ciudad había sido ocupada, se inició el difícil proceso de convertirla en una provincia castellana. Como base de la nueva situación existían las capitulaciones firmadas con *Boabdil*. En ellas se prometía respetar las instituciones de los moros granadinos, aunque estarían sometidos a un gobierno cristiano. El primer arzobispo de Granada optó por la conversión de los moros de forma pacífica, constante y caritativa, con lo que consiguió el respeto del colectivo mudéjar.

Los Reyes Católicos no se atuvieron demasiado a las capitulaciones firmadas con *Boabdil* en lo que a la religión de los vencidos se refería, ya que desde el primer momento consideraron el islamismo como provisional y buscaron los medios para acelerar su conversión. En una visita de los Reyes Católicos a Granada en 1499, un gran número de moras recibieron a los reyes, ante lo que tuvieron la sensación de estar en una ciudad musulmana administrada por cristianos. Ante esta visión se decidió que los moros tenían que convertirse o emigrar, con lo que comenzaron duras medidas para acelerar el proceso de conversión de los moriscos (Figura 2.3). En este mismo año, la mezquita mayor del Albaicín fue transformada en iglesia cristiana. El resultado de los métodos de Cisneros fue el bautismo de un gran número de musulmanes en Granada [4,5].

Los cristianos confiaban que con el bautismo de los moriscos, estos terminarían por integrarse en el colectivo cristiano, pero no fue así, y las diferencias eran cada vez más patentes. En 1526 se pidió la prohibición de todas las peculiaridades moriscas, así como vigilar extremadamente el comportamiento cristiano de los moriscos mediante la instalación en Granada de un Tribunal de la Inquisición. Sin embargo, Carlos V no autorizó tales extremos, optando por una línea cultural y dando lugar a un periodo de buena convivencia durante al menos tres décadas. Durante bastantes años, la marcha económica y social del reino granadino continuó sobre las raíces musulmanas, con las formas y costumbres cristianas. El poder cristiano iría en aumento, no sólo en los altos estamentos de la Administración sino también a nivel popular. La intolerancia religiosa protagonizada por la Inquisición y las drásticas medidas de Cisneros dieron lugar a la Revuelta de las Alpujarras (1560-1570), que terminó con la cruel decisión de la expulsión de los moriscos de Granada y la dispersión de ellos por el interior de Castilla (140 000 personas). Ordenada la expropiación de los bienes de los moriscos, fue necesario enviar nuevos pobladores al reino de Granada. El periodo

de repoblación abarca desde 1571 hasta la real instrucción del año 1595. Al mismo tiempo que el Consejo de la Población trabajaba en Granada para preparar la repoblación, se mandaron comisarios a toda la Península para reclutar pobladores. Se reunieron 12 546 familias con las que se poblaron 270 lugares [4]. La gran mayoría de los repobladores procedían del norte de la Península: Galicia, Asturias, Burgos, León, País Vasco y en menor número de otras zonas [4]. El 9 de Diciembre de 1609, Felipe III firmó la orden de expulsión de los moriscos de toda España. Se cree que la expulsión de los moriscos fue el motivo de la decadencia económica de España durante el siglo de oro [4,5].



Figura 2.3. Mapa de la distribución de los moriscos tras la expulsión de éstos del Reino de Granada.

El siglo XVII se caracteriza en todo el territorio peninsular por una decadencia económica, en negocios, industria y comercio. Granada fue un poco la excepción de la regla, ya que todavía persistía algo de la inercia de la prosperidad del antiguo reino granadino. En la segunda mitad del siglo XVII, a pesar de las epidemias, la población de Granada aumentó debido a la inmigración atraída por su riqueza agrícola. Tras la expulsión de los moriscos, la industria de la seda tuvo un gran parón y decadencia, ya que los castellanos sólo aprendieron a criar el gusano y a exportar la seda en bruto. El estado no empleaba nada de su presupuesto en sanidad, educación, beneficencia, obras públicas... La recaudación del Estado estaba destinada a los gastos de la Casa Real y su burocracia, deuda pública y los gastos militares.

La España del siglo XVII y comienzos del XVIII había llegado a una situación de decadencia extrema. Una vez acabada la Guerra de Sucesión se comenzó con un proteccionismo para detener la importación, al mismo tiempo que se inició la formación de una industria estatal. En apoyo a esta política económica borbónica se crearon las Sociedades Económicas. En Granada, la Sociedad Económica de Amigos del País se creó en noviembre de 1776.

En 1833, tras 341 años de existencia, las provincias de Almería y Málaga se separaron del reino, suprimiendo así el antiguo reino cristiano de Granada [4,5].

Se puede concluir que, durante el establecimiento del Reino de Granada, así como durante todo el tiempo que existió, fueron múltiples los pueblos que en él convivieron. El paso de visigodos, bereberes, sirios, saharauis (almorávides), marroquíes (almohades), judíos y cristianos por el Reino de Granada ha hecho de las actuales ciudades de Granada, Málaga y Almería un crisol de civilizaciones que queda reflejado en sus monumentos arquitectónicos, sus costumbres y su folklore. En la actualidad, existen múltiples y muy diversos movimientos migratorios en Andalucía, no sólo del Norte de África sino procedentes de todas las partes de mundo. En la tabla 2.1 queda resumida la procedencia de la inmigración actual en las provincias de Granada, Málaga y Almería. Estos movimientos no son significativos y por ello ninguno ha sido seleccionado para este estudio.

Tabla 2.1: Procedencia de la inmigración en las provincias de Granada, Málaga y Almería (Instituto Nacional de Estadística, datos de 2013).

	% Europa Occ.	% Europa Or.	% Marruecos	% África	% América Latina	% Asia
Almería	31,10	2,17	15,48	22,03	6,91	2,11
Granada	31,11	2,59	9,88	14,08	15,68	4,72
Málaga	45,86	8,59	6,80	9,74	11,23	4,40

1.2 Influencias en la Cultura

Actualmente, se observan grandes influencias de la cultura andalusí en la cultura del sur de la Península Ibérica.

Los sistemas de canalización de agua, *qanats*; así como los sistemas empleados como medio de riego datan de esta época, la noria (*na'ura*), el molino de agua o las acequias (*saqiya*). Lo mismo ocurre con los aljibes de la ciudad de Granada que hoy se conservan prácticamente intactos. Con esa tradición han llegado hasta nuestros

días el Tribunal de las Aguas de Valencia o la gran variedad de Comunidad de Regantes.

El paisaje y el entramado urbano característico de las ciudades como Córdoba y Granada, el casco antiguo de Málaga e infinidad de pueblos de Andalucía, muestran calles angostas con casas encaladas o con ajimeces desde donde las mujeres podían asomarse sin ser vistas; herencia directa de la cultura andalusí.

La higiene y la belleza, costumbres cotidianas, derivan directamente de las enseñanzas de *Ziryab*, árbitro de la elegancia cordobesa y famoso músico venido de Bagdad, que inventó una crema de dientes y un desodorante eficaz. Así mismo, innovó en el corte de pelo de los hombres y estableció un código de colores para la vestimenta según la época del año. Los andalusíes introdujeron en España la explotación del algodón y el cultivo de la seda.

Costumbres como derramar arroz sobre los recién casados; símbolo de abundancia, siguen manteniéndose intactas. Se dice que *Zyriab* fue el inspirador de ciertas supersticiones como la fobia al número 13 y el convencimiento de que existen antojos en las embarazadas. También distribuyó ciertas creencias como que los rabos de pasa aseguran la memoria o que los niños que juegan con fuego se orinan en la cama. Además, creó una escuela de música que puede considerarse como el primer conservatorio musical e inventó un laúd, al que incorporó una quinta cuerda para darle mejor sonido. Este laúd, que aún tocan en zonas de África, es el padre directo de la guitarra española.

También de la época de al-Ándalus viene uno de los símbolos institucionales que identifican a la actual Comunidad Autónoma de Andalucía, la bandera blanca y verde. El primer texto referente a una bandera blanca y verde es de la época de al-Ándalus. El verde y blanco era un color muy habitual en los estandartes andalusíes.

1.3 Influencia en el Arte

Una parte importante del rico patrimonio andaluz se debe a la civilización andalusí. No solo los conocidos monumentos de la Alhambra, la Mezquita de Córdoba, la Giralda, la Alcazaba de Málaga, sino también esa multitud de pequeñas construcciones que destacan en el territorio andaluz como son las mezquitas convertidas en iglesias, las almenaras, los baños, aljibes, albercas y antiguas acequias, algunas de ellas en uso, así como casas y palacios de la época que han sufrido restauraciones posteriores.

El arte andalusí concede una gran importancia a la casa cuidando todos los detalles, no sólo del interior sino también del exterior. Solían ser casas encaladas

con tejas rojizas, patio interior y un zaguán a la entrada; todo ello adornado de plantas y flores.

Algunas de las soluciones arquitectónicas introducidas por los andalusíes tuvieron una gran aceptación en los estilos importados de Europa, como la bóveda nervada o el recubrimiento de las naves con arcos de madera, que se caracterizan por su decoración geométrica formada por líneas o cintas que se entrecruzan formando polígonos o estrellas (Figura 2.4).



Figura 2.4: Detalle de un arcosado de madera.

1.4 Influencias en la Lengua

A pesar de toda la fuerza del árabe, es cierto que junto a él pervivió la lengua romance proveniente del latín, y no sólo entre los cristianos mozárabes sino también entre los propios musulmanes y judíos. El bilingüismo en la sociedad andalusí ha hecho que apenas un 7% de las palabras actuales provengan del árabe y se hayan mantenido un gran número de palabras de origen latino provenientes de romance aljamiado [6].

Aún así, hay más de cuatro mil palabras procedentes del árabe. Son claramente visibles las influencias lingüísticas árabes según oficios, como albañil, alarife, alfarero; términos jurídicos como alcalde, alguacil; términos rurales como arroba, fanega; relacionados con el regadío, noria, aljibe, alberca; con la agricultura, naranja, albaricoque, limón, alcachofa [6].

Otras palabras son comúnmente usadas y se desconoce su etimología, como la interjección “ojalá” de la que se deduce su procedencia de la expresión “*in shaa Allah*”, o si Dios quisiera. Lo mismo ocurre con el término “algarabía”, que siendo en su origen el nombre que daban los cristianos a la lengua árabe (claramente ruidosa), terminó empleándose para designar alboroto. También las frases hechas han recogido la huella andalusí, como “De higos a brevas”, “No hay moros en la costa”, “Esto o es moro o es cristiano” o “Ya sabes dónde está tu casa” que siempre se recalca al presentar nuestro hogar a un desconocido, expresión propia de la hospitalidad islámica y que parece ha quedado en el recuerdo colectivo [6].

1.5 Influencias en la Toponimia

En el caso de la Toponimia está clara la procedencia andalusí de multitud de nombres de pueblos, ciudades, ríos, lugares, pagos agrícolas, etc, ya sea directamente del árabe, o bien a través de la arabización de un topónimo anterior.

Está claro que el nombre de Andalucía viene del término *Ándalus* al que se le añadió el artículo árabe *al*. El primer documento que se conoce con el nombre de *Ándalus* son las monedas bilingües del año 716 en las que aparece en una de las caras de la moneda escrito en árabe, y en la otra cara, escrito en latín, como sinónimo el término de *Hispania*.

Si bien muchos topónimos vienen directamente del árabe, aunque sufriendo alguna que otra corrupción, como Guadalquivir, Almería, Jaén; hay otros que vienen de un topónimo anterior pero que han sufrido una arabización fonética y una corrupción. Es el caso de Sevilla, que del *Ispalis* latino se ha ido transformando hasta dar el Sevilla actual.

1.6 Influencias en la Gastronomía

La entrada del Islam en la Península Ibérica, supuso la ruptura de una trilogía básica propia de la manutención cristiana de la época, centrada en el trigo, la carne y el vino. Las influencias de Oriente posibilitaron el cambio hacia nuevos hábitos alimenticios ante la incorporación de las verduras como elemento imprescindible para su consumo en solitario, o como acompañamiento a sopas, carnes o pescados. La instauración de sistemas de regadío por todo el territorio permitió el consumo durante todo el año de verduras y hortalizas frescas.

Especial importancia adquirieron las legumbres como la soja, los garbanzos, las lentejas y las judías, considerados como alimentos altamente nutritivos. El flujo comercial con otras zonas del mediterráneo, permitía el consumo constante de frutas que posteriormente formaron parte de las cosechas andalusíes: sandía, melón, albaricoque, membrillo y granada, entre otras frutas, convertida esta última en el símbolo de referencia de *al-Ándalus* [7,8].

El uso y consumo de especias en la cocina andalusí permitió la introducción de la canela así como el azafrán, el comino, el jengibre, la nuez moscada, el anís, el sésamo o el cilantro. Otro de los reconocimientos a la aportación andalusí se debe a la introducción de la caña de azúcar en todo el territorio europeo, como excelente sustituto de la miel en algunos platos y usos cotidianos.

El *Sherbet*, sorbete, es otra de las herencias andalusíes que siguen siendo utilizados hasta nuestros días para refrescar el calor propio del verano, a base de esencia de flores y frutos con agua fría o hielo.

El impulso del desarrollo agrícola promovió una variación confortable en la alimentación. Es así como los cereales adquieren una base fundamental en la alimentación andalusí, cocinados como pan, sémolas, sopas o gachas [7,8].

De la influencia de otras culturas, los platos eran preparados de forma diversa, tanto en asado, guiso o frito. El pescado tenía una base crucial en la gastronomía andalusí preparándose frito en aceite de oliva, relleno, en escabeche (*iskabay*) o guisado.

Eran de gran valor los frutos secos y a los productos lácteos. Se conservan gran número de postres, elaborados con almendras, pistachos, piñones, agua de azahar y de rosas, miel, hojaldres (de origen judío), sésamo o aceite de almendras dulces, presentados tanto fritos como al horno.

El refinamiento de la cocina andalusí, desconocido en Europa, transformó muchas costumbres. Fueron los andalusíes los que crearon la sala de comer o el comedor y el actual orden de servicio de los platos. *Zyryab* propone el orden en el que han de servirse los alimentos en las comidas: "sopas y caldos primero, entradas de carne y aves sazonadas después, y finalmente, postres azucarados, dulces y pasteles de nueces." De igual forma, fue el responsable del uso de la cuchara y de las copas en la mesa [7,8].

2. MATERIALES Y MÉTODOS

2.1 Toma de Muestras

Se han estudiado un total de 245 individuos adultos no relacionados entre sí de la región Sur Este de la Península Ibérica [Almería (79), Granada (94), Málaga (79)] (Figura 2.5).

En la toma de muestras se tomaron indistintamente hombre y mujeres, en todos los casos no emparentados entre sí. Solamente se tomaron muestras a aquellos individuos que determinaron que al menos tres generaciones anteriores habían nacido en la zona de estudio.

Todos los donantes fueron informados del propósito de este estudio, aceptaron su participación y firmaron un consentimiento informado según la guía ética de la declaración de Helsinki.

A todos los donantes se les tomó una muestra de epitelio de la mucosa bucal mediante hisopado. Todas las muestras se dejaron secar y fueron remitidas al Laboratorio de Identificación Genética del Departamento de Medicina Legal, Toxicología y Antropología Física de la Universidad de Granada. Allí se codificaron, anonimizaron, almacenaron y procesaron todas las muestras.

2.2 Estudio de Apellidos

Se anotaron ambos apellidos de los 245 individuos no relacionados estudiados. Se consultó la página web del Instituto Nacional de Estadística (www.ine.es) para determinar la provincia española con mayor número de individuos nacidos con cada apellido, así como la frecuencia de cada apellido tanto en la población de las provincias de Granada, Málaga y Almería como en la población española. Además, se consultaron diversas páginas de heráldica y linaje para determinar el origen histórico de los apellidos. Se dividió la población en ocho grupos y se clasificó en función del origen de los apellidos y del lugar de nacimiento de los portadores de los mismos. Se han comparado todos los apellidos con una lista de 1620 apellidos de origen árabe presentes en el libro "*Apellidos españoles de origen árabe*" [9].

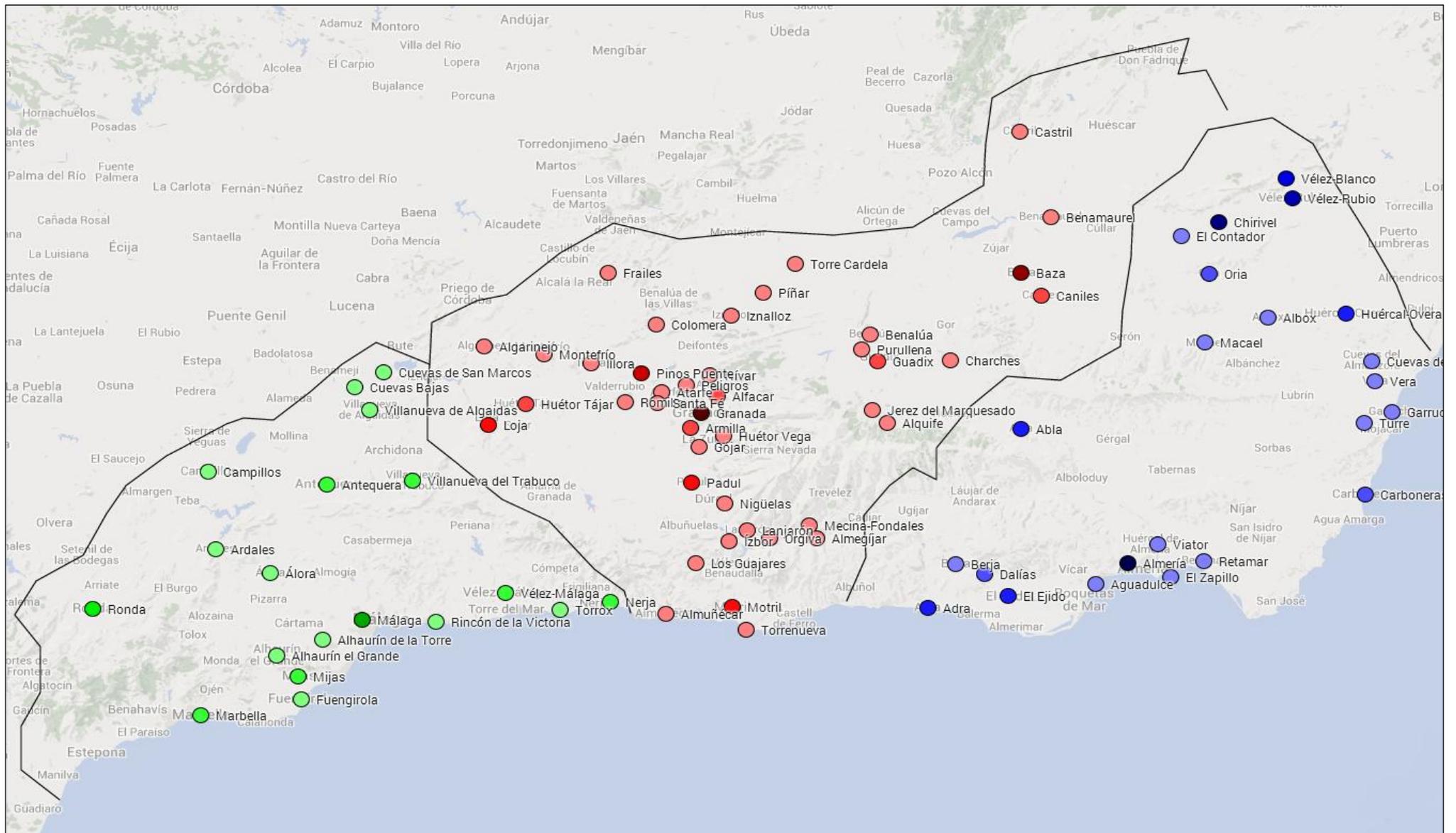


Figura 2.5: Mapa de la región de estudio y municipios de procedencia de las muestras. El gradiente de colores representa el número de muestras disponibles en cada municipio; a mayor número de muestras color más oscuro.

3. RESULTADOS Y DISCUSIÓN

3.1 Importancia del Estudio de Apellidos

El uso de apellidos en estudios poblacionales tuvo su origen en el siglo XIX [10]. Desde entonces, se han desarrollado múltiples técnicas para deducir la estructura genética de una población a través del estudio de la distribución de los apellidos en la misma [11–14].

Los primeros estudios de genética de poblaciones clasificaban a los individuos estudiados según su residencia o lugar de nacimiento, su etnicidad y su lengua. Sin embargo, se necesitan más datos relacionados con la región de estudio para seleccionar a los individuos. Actualmente, la gran mayoría de los estudios poblacionales aplican el criterio de que como mínimo los abuelos de los individuos seleccionados ya residían en la región de estudio [15,16]. Otro criterio de selección, es el de elegir a individuos de zonas rurales, ya que se prevén menores tasas de migración reciente.

Además de estos criterios de selección, diversos estudios poblacionales emplean el apellido de los donantes en regiones donde se da una herencia patrilineal del mismo [14,17]. El apellido de un individuo da información extra sobre la genealogía de una persona, pero está limitada al linaje paterno. Los datos genealógicos ayudan a:

- Verificar la estructura genética poblacional observada, así como a interpretarla correctamente a través de la mejora de la representatividad de la muestra.
- Estimar más certeramente la data de los eventos genéticos observados.
- Determinar la diferenciación genética temporal de una población; únicamente aplicable a marcadores haploides [14].

Ciertos estudios han establecido unas recomendaciones para la correcta selección de muestras auténticas ideales en estudios poblacionales de regiones concretas [18].

- La genealogía completa del donante ha de confirmar una ancestralidad ligada a la región durante los últimos dos siglos.
- No se han de incluir parientes recientes a los donantes basándose en su genealogía completa.
- Los apellidos de los participantes, así como de sus ancestros maternos, han de estar presentes en documentos donde se confirme su presencia en la región con anterioridad a grandes eventos migratorios.

- El origen, la toponimia y el dialecto del apellido de los donantes y de sus parientes maternos conocidos han de pertenecer a la región de estudio.
- La distribución de los apellidos de los participantes, así como de sus parientes maternos conocidos, han de estar restringidos al área de estudio.

Larmuseau y cols. han encontrado una diferenciación temporal significativa en el gradiente norte-sur de Brabante basándose en exhaustivos datos genético-genealógicos [14]. Además, otro estudio del mismo grupo revela que los eventos de mestizaje ocurridos en Flandes durante el siglo XVI pueden ser trazados mediante la antroponimia de los apellidos. Sin embargo, este fenómeno no se detecta con datos genealógicos, hecho que corrobora la importancia del origen del apellido en la selección de los individuos en estudios poblacionales [13].

3.2 Estudio de Apellidos

En la muestra de 245 individuos pertenecientes a las provincias de Granada, Málaga y Almería se han observado 266 apellidos diferentes en el total de la población estudiada (165 si sólo se tiene en cuenta el primer apellido). De ellos, 197 apellidos se observaron una única vez (*singletons*) y el resto (69 apellidos) se presentaron con frecuencias entre 2 y 23 (Tabla 2.2).

Tabla 2.2: Apellidos observados en la población GMA

Apellidos	n	k	Singletons*
Primer apellido	245	165	136
Segundo apellido	245	149	107
Primer + Segundo apellido	490	266	197

n, tamaño muestral; k, número de apellidos diferentes, * apellidos que se observan una única vez

En España, hay 74 015 apellidos con una frecuencia mayor o igual a 20 según el censo de 2014 (Instituto Nacional de Estadística, www.ine.es). El apellido más frecuente en la población es García (4,63% de los apellidos). La figura 2.6 muestra la distribución de los 69 apellidos que se encuentran en más de una ocasión en la población GMA, en la población total de Granada, Málaga y Almería y en la población total española. Se observan patrones similares de distribución de las frecuencias en las tres poblaciones. A nivel nacional, se han observado 29 apellidos con una frecuencia en la población inferior a 0,001 y 5 con frecuencias inferiores a

0,0001. A nivel de la región de estudio, se han determinado 19 y 2 apellidos respectivamente.

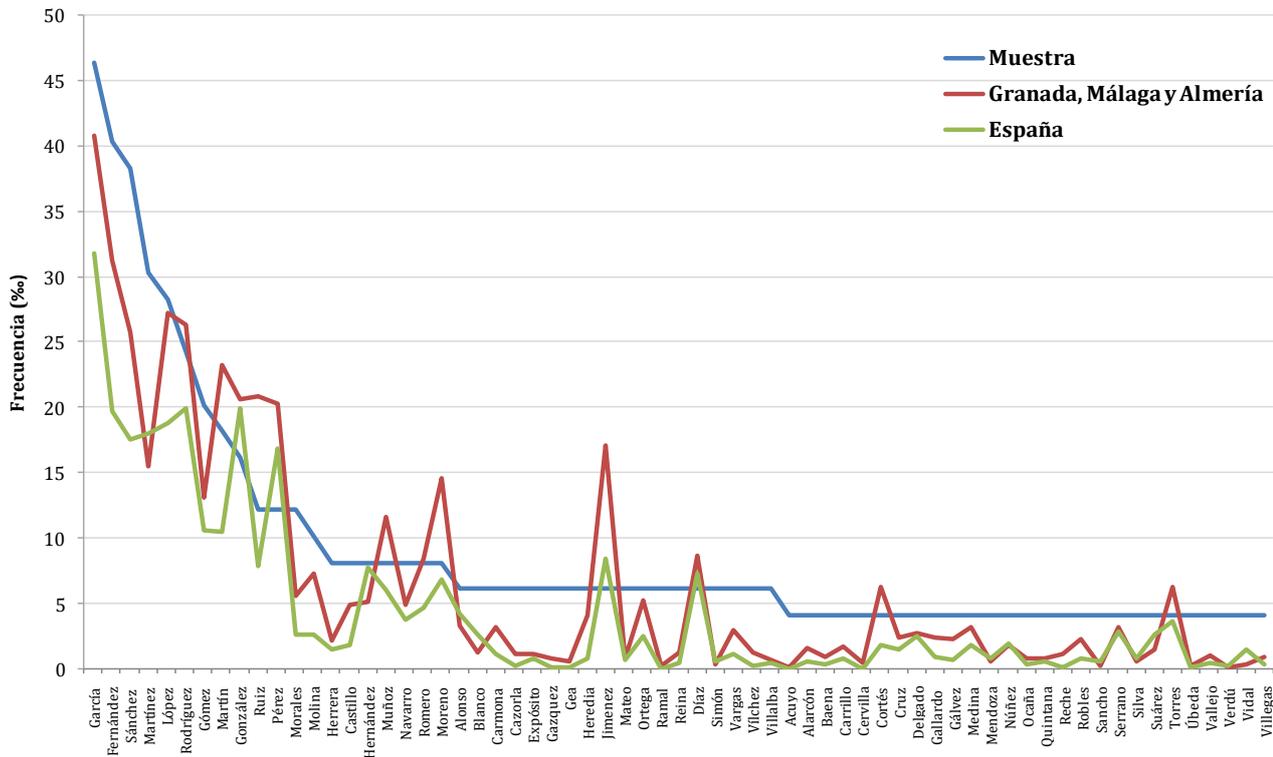


Figura 2.6: Distribución de los apellidos más comunes en la población GMA. Los datos de las poblaciones de Granada, Málaga, Almería y la población total española se han tomado de la Estadística del Padrón Continuo a 1 de enero de 2014, INE.

La distribución de la ocurrencia de los apellidos frente a la abundancia de los apellidos sigue una distribución en forma de ley potencial $y = 74,74x^{-1,66}$, $R^2 = 0,820$ (Figura 2.7).

En la mayoría de las poblaciones, el valor del exponente de la ley potencial varía entre -2 y -1,5. El exponente de la ley potencial mide el ratio de creación de apellidos frente al crecimiento de las familias de apellidos (número de individuos que comparten un mismo apellido), teniendo en cuenta el tiempo que hace que este apellido existe, y viene determinado por parámetros asociados con procesos de nacimientos, inmigración y la distribución de la descendencia [19]. En los casos de mestizaje o una importante presencia de apellidos de diferentes idiomas, el exponente es muy cercano a -1,5 [19–21]. El valor obtenido para la población GMA (-1,66) sugiere que el ratio de crecimiento del tamaño de las familias de apellidos excede al ratio en el que nuevos apellidos evolucionan de los ya existentes, hecho común en las sociedades en las que el apellido se hereda de padres a hijos, como es la población GMA.

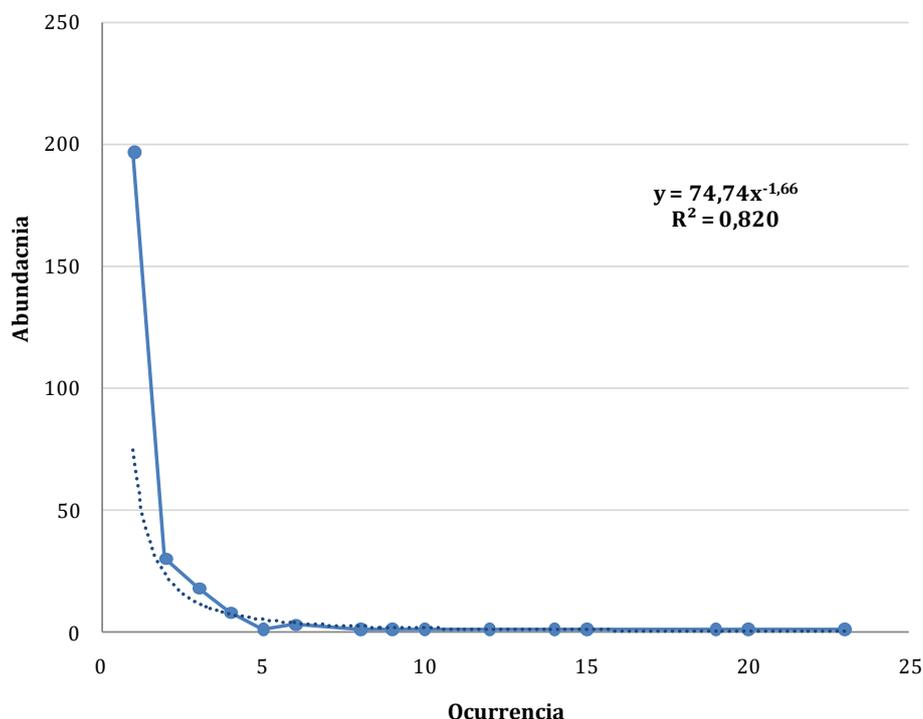


Figura 2.7: Ocurrencia de los apellidos *versus* su abundancia en la población GMA. *Ocurrencia* se refiere al número de veces que se observa a un apellido en la población. *Abundancia* es el número de apellidos en cada ocurrencia. La línea punteada representa la distribución de ley potencial esperada.

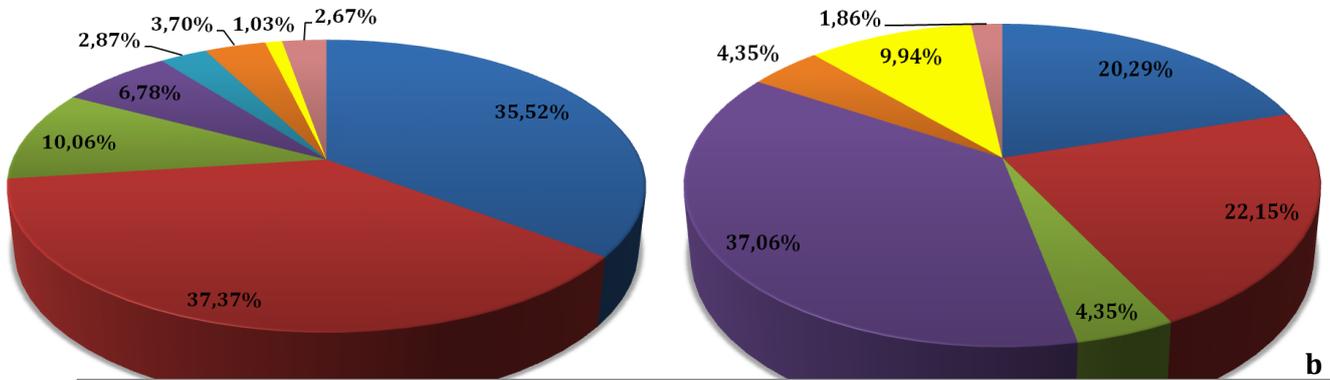
A través del estudio de los apellidos se puede dividir la población en ocho subgrupos; basados en la localización geográfica y su origen histórico.

Al analizar el conjunto de ambos apellidos de los individuos estudiados se ha determinado que la gran mayoría de los apellidos tienen origen en las regiones norte y centro de la Península. Un 37% de los apellidos encontrados se originan en el centro de España, sede de la corona de Castilla. 35% de los apellidos provienen del norte de la Península, cuyos habitantes descienden de los pueblos celtas, 10% provienen de Aragón y 4% de Cataluña. Solamente un 7% de los apellidos de los individuos estudiados provienen de Andalucía, del Antiguo Reino de Granada (Figura 2.8a).

Sin embargo, al estudiar la provincia con mayor número de individuos nacidos con cada apellido, se observa que un 37% de los apellidos están distribuidos por Andalucía, 22% en Castilla, 20% en el norte de la Península y un 10% se incluyen en otras regiones como las Islas Canarias (Figura 2.8b).

Por último, 10 de los 266 apellidos presentes en la muestra tienen origen etimológico árabe (8 primeros apellidos y 4 segundos apellidos; dos de ellos

presentes tanto en el grupo de primer apellido como de segundo apellido) 8 de ellos únicos para cada grupo. Los apellidos presentes tanto en primer como en segundo apellido se encuentran, en el caso de Simón, en 4 individuos (3 primer apellido y 1 segundo apellido) y Medina en 2 individuos (uno en cada).

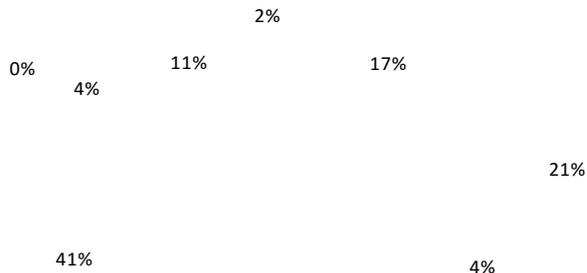


a

b

Título del gráfico

- Galicia-Asturias-Cantabria
- Andalucía-Extremadura-Murcia
- Navarra-Aragón-La Rioja
- País Vasco
- Cataluña-C.Valenciana-Baleares
- Indeterminado
- Castilla
- Otras regiones



en la historia de esta región. Los datos históricos señalan que los moriscos que habitaban el Reino de Granada fueron expulsados y aislados. Muy pocos de ellos se convirtieron al cristianismo y permanecieron en la región, conocidos como cristianos nuevos. Como ya se ha comentado anteriormente, durante este periodo los musulmanes y judíos fueron adoptando apellidos cristianos, así como el sistema de herencia masculina de estos. Posteriormente, tras la ocupación de la ciudad de Granada por los españoles en 1492, los moriscos fueron relegados a la zona de la Alpujarra hasta 1570 de donde fueron totalmente expulsados. Todas estas regiones fueron repobladas con gente traída del norte y centro de la Península. No existía ningún tipo de contacto entre los cristianos nuevos y los

cristianos viejos. En 1609 se expulsaron a todos los moriscos o cristianos nuevos de la Península Ibérica hacia el norte de África.

El gran aislamiento de los moriscos así como el escaso contacto entre los nuevos pobladores y estos, queda reflejado en los resultados obtenidos en el estudio de apellidos. Se puede ver cómo la gran mayoría de los apellidos tienen un origen castellano o gallego, mientras que la gran mayoría de apellidos tienen mayores frecuencias en el sur de España. Un 12,57% de los apellidos con mayor número de individuos nacidos en el sur de la Península son de origen gallego, asturiano o cántabro. Entre estos podemos encontrar apellidos como Carmona, Cubero, Ferrón, Montes, Padiel, Rojas o Santiago. Por otro lado, el 18,03% de los apellidos con mayor número de individuos nacidos en el sur de la Península provienen del centro de la Península, Castilla León y Castilla la Mancha. Son ejemplos de estos Burgos, Castillo, Domínguez, Guerrero, León o Romero. Por último, 7,65% de los apellidos con mayor número de individuos nacidos en el sur de la Península son de origen navarro-aragonés; Aragón, Cortés, Navas o Soto. Aunque históricamente se han hablado muchas lenguas en la Península Ibérica, castellano, portugués, gallego, vasco, catalán, árabe y hebreo; dando lugar a ciertos apellidos característicos; la gran mayoría de la población española tiene apellidos de origen castellano-leonés, predominantes en todas las provincias españolas [21].

4. BIBLIOGRAFÍA

- [1] T. Chandler, *Four Thousand Years of Urban Growth: An Historical Census*, St. David's University Press, Lewiston, N.Y., U.S.A, 1987.
- [2] E. Gozalbes-Cravioto, M. Pastor-Muñoz, La ciudad de Granada en época romana, *Rev. Del Cent. Estud. Históricos Granada Y Su Reino*. 25 (2013) 7–13. <http://cehgr.es/revista/index.php/cehgr/article/download/39/34>.
- [3] M. Pastor-Muñoz, La estructura social de Iliberis según la epigrafía, *Rev. Del Cent. Estud. Históricos Granada Y Su Reino*. 25 (2013) 29–48.
- [4] P. Bueno, *El reino de Granada (De orígenes a 1936)*, Don Quijote Editorial, Granada, 2004.
- [5] M. Garzón, *Historia de Granada, Gráficas del Sur*, Granada, 1980.
- [6] J.A. Chavarría Vargas, Herencia lingüística de mudéjares y moriscos en la lengua castellana, *Estud. Sobre Patrimonio, Cult. Y Cienc. Mediev.* 7-8 (2006) 115–134.
- [7] E. García Sánchez, La gastronomía andalusí, in: Lunweg Editores (Ed.), *El Zoco Vida Económica Y Artes Tradic. En Al-Andalus Y Marruecos*, Barcelona, 1995: pp. 49–57.
- [8] E. García Sánchez, La alimentación popular urbana en al-Andalus, *Arqueol. Mediev.* 4 (1996) 219–235.
- [9] J.M. Calvo Baeza, *Apellidos españoles de origen árabe*, Darek-Nyumba, Madrid, 1990.
- [10] G.H. Darwin, Marriages between first cousins in England and Wales and their effects, *J. Stat. Soc.* 38 (1875) 153–184.
- [11] M.A. Jobling, In the name of the father: Surnames and genetics, *Trends Genet.* 17 (2001) 353–357. <http://www.scopus.com/inward/record.url?eid=2-s2.0-0035368897&partnerID=40>.
- [12] T.E. King, M.A. Jobling, What's in a name? Y chromosomes, surnames and the genetic genealogy revolution, *Trends Genet.* 25 (2009) 351–360. <http://www.scopus.com/inward/record.url?eid=2-s2.0-68349128219&partnerID=40>.
- [13] M.H.D. Larmuseau, J. Vanoverbeke, G. Gielis, N. Vanderheyden, H.F.M. Larmuseau, R. Decorte, In the name of the migrant father—Analysis of surname origins identifies genetic admixture events undetectable from genealogical records, *Heredity (Edinb)*. 109 (2012) 90–95. doi:10.1038/hdy.2012.17.
- [14] M.H. Larmuseau, C. Ottoni, J.A. Raeymaekers, N. Vanderheyden, H.F. Larmuseau, R. Decorte, Temporal differentiation across a West-European Y-chromosomal cline: genealogy as a tool in human population genetics, *Eur. J. Hum. Genet.* 20 (2012) 434–440. doi:10.1038/ejhg.2011.218.
- [15] S.M. Adams, E. Bosch, P.L. Balaesque, S.J. Ballereau, A.C. Lee, E. Arroyo, et al., The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula., *Am. J. Hum. Genet.* 83 (2008) 725–36. doi:10.1016/j.ajhg.2008.11.007.
- [16] E. González-Pérez, E. Esteban, M. Via, M. Gayà-Vidal, G. Athanasiadis, J.M. Dugoujon, et al., Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems), *Am. J. Phys. Anthropol.* 141 (2010) 430–9. doi:10.1002/ajpa.21161.
- [17] M.H.D. Larmuseau, N. Vanderheyden, M. Jacobs, M. Coomans, L. Larno, R. Decorte,

- Micro-geographic distribution of Y-chromosomal variation in the central-western European region Brabant, *Forensic Sci. Int. Genet.* 5 (2011) 95–99. doi:10.1016/j.fsigen.2010.08.020.
- [18] M.H.D. Larmuseau, A. Van Geystelen, M. Van Oven, R. Decorte, Genetic genealogy comes of age: Perspectives on the use of deep-rooted pedigrees in human population genetics, *Am. J. Phys. Anthropol.* 150 (2013) 505–511. doi:10.1002/ajpa.22233.
- [19] W.J. Reed, B.D. Hughes, On the distribution of family names, *Phys. A Stat. Mech. Its Appl.* 319 (2003) 579–590. doi:10.1016/S0378-4371(02)01455-3.
- [20] S. Miyazima, Y. Lee, T. Nagamine, H. Miyajima, Power-law distribution of family names in Japanese societies, *Phys. A Stat. Mech. Its Appl.* 278 (2000) 282–288. doi:10.1016/S0378-4371(99)00546-4.
- [21] R. Calderón, C.L. Hernández, P. Cuesta, J.M. Dugoujon, Surnames and Y-chromosomal markers reveal low relationships in southern Spain, *PLoS One.* 10 (2015). doi:10.1371/journal.pone.0123098.

ABSTRACT

The former Kingdom of Granada, comprising the provinces of Granada, Málaga and Almería (GMA), was once inhabited for more than 700 years by North African population who influenced its creation and establishment. This essay analyzes that influence on those who inhabited the three provinces. In this work, the genetic data of fifteen autosomal short tandem repeats (STRs) in 245 unrelated donor residents of the provinces were analyzed. Observed-allelic frequencies were also analyzed. AMOVA and STRUCTURE analysis show that the populations of the three provinces can be treated genetically as a single population.

For better characterization, the population of GMA was compared with that of North Africa, Iberian Peninsula and South Europe. Analysis of allelic frequencies and genetic distance shows that GMA population was located in the Spanish populations group, but more similar to the North African populations than the rest of the Spanish ones. The analytical results affirm that, owing to the studied genetic markers, the GMA population shows no characteristics that reflect genetic influence of North African people. This feature is consistent with the historical data that African inhabitants were expelled or isolated during the repopulation of the region with Spaniards from northern Spain.

1. INTRODUCCIÓN

El genoma eucariótico está lleno de secuencias de ADN repetidas. Éstas presentan tamaños variables y se denominan en función de la longitud del núcleo de la unidad de repetición y del número de unidades de repetición continuas o de la longitud total de la región de repetición [1].

Los polimorfismos con regiones de repetición de 2 a 7 pb de longitud se llaman repeticiones cortas en tándem (STRs o *Short Tandem Repeats*) o microsatélites. Los STRs son numerosos por todo el genoma, siguen un patrón de herencia mendeliana [2] y han sido descritos en regiones génicas y extragénicas. Los situados en regiones génicas están presentes tanto en intrones y secuencias flanqueantes como en regiones codificantes. El alto grado de polimorfismo de los loci STRs se debe a la variación en el número de unidades de repetición [3,4]. Además, las secuencias STRs varían también en la longitud de la unidad de repetición y el rigor con el que se conforma el modelo de repetición. Las unidades “*simples*” contienen repeticiones de idéntica longitud y secuencia, las “*complejas*” contienen varios bloques de repetición de unidades de longitud variable y las “*compuestas*” contienen varios bloques de repetición con unidades de longitud variables así como las secuencias intermedias variables [1]. Existen también las “*repeticiones hipervariables complejas*” que presentan un número de alelos no consensuado que difieren tanto en tamaño como en secuencia y por ello son difíciles de reproducir los genotipos. Esta última categoría no se emplea de rutina en los análisis forenses por la complejidad de la nomenclatura de los alelos y la variabilidad entre laboratorios aunque existen kits comerciales que ya los incluyen, locus SE33 [5].

Para poder analizar los STRs, se ha de determinar la secuencia no variable que rodea al núcleo de repetición. Una vez conocida ésta, se pueden diseñar los primers que amplifiquen la región repetida. Nuevos STRs se identifican de dos formas: buscando regiones de ADN con más de seis unidades de repetición contiguas en bases de datos de secuencias de ADN, o mediante métodos de aislamiento de biología molecular.

Los STRs se denominan en función de la longitud de la unidad de repetición. Los STR tetraméricos y pentaméricos son los más usados ya que:

- Poseen un rango alélico estrecho que permite reacciones multiplex; se han descrito multiplex que coamplifican hasta 22 loci en una sola reacción de PCR.
- Poseen un rango alélico que reduce la generación de alelos *dropout*.

- Se generan productos de PCR pequeños favoreciendo el análisis de muestras degradadas.
- Se reducen los productos *stutter* facilitando la interpretación de mezclas.

Gracias al Human Genome Project se sabe que existen unos 20 000 loci STR tetraméricos en el genoma humano y más de un millón de microsatélites [6]. Sin embargo, los microsatélites forman un 3% de la variación total del genoma humano [5].

En 1997, en USA, se implantó el empleo de 13 marcadores y el loci de la amelogenina en los laboratorios estatales con el fin de crear una base de datos nacional que posibilitara el intercambio de datos entre diferentes laboratorios (CODIS). Bases de datos similares se han establecido en la Unión Europea (EES).

Los requisitos imprescindibles para el empleo de un marcador en rutina forense son:

1. Alto poder de discriminación ($>0,9$) y de heterozigosidad observada ($>70\%$). Alto grado de polimorfismo. Cuanto mayor sea el número de variables que presenta el marcador, menos marcadores será necesario emplear.
2. Tamaño entre 90 y 500 pares de bases. Éstos son buenos candidatos para trabajar con muestras degradadas. Si se empleasen marcadores genéticos más largos es muy probable que se pierdan por degradación.
3. Los alelos de un marcador han de estar dentro de un margen estrecho y no solaparse entre ellos para poder discriminarlos y así poder trabajar en sistemas multiplex.
4. Los marcadores han de estar en cromosomas independientes asegurando la independencia de loci, con lo que se evita una transmisión en bloque de los alelos.
5. Han de ser reproducibles en reacciones de PCR-Multiplex. Los resultados obtenidos por PCR convencional han de ser los mismos que los obtenidos cuando realizamos una PCR-Multiplex.
6. Han de tener bajas tasas de mutación, especialmente para casos de paternidad.
7. Es preferible que se trate de repeticiones de tetra y pentanucleótidos ya que tienen niveles bajos de *stutter band*.

En los análisis forenses se realiza una PCR multiplex para más de 15 marcadores genéticos a la vez, incluyendo el locus de la amelogenina. Para dicha amplificación se emplean primers que flanquean el fragmento a estudiar y se encuentran marcados con una molécula que emite fluorescencia. La fluorescencia emitida por

esta molécula y el tamaño del fragmento que la emite nos permitirá determinar la longitud del marcador que se está estudiando. Para el trabajo de rutina se suelen emplear kits comerciales que contienen la mezcla de productos (incluidos los primers marcados y sin marcar) precisos para realizar una PCR múltiple (Tabla 3.1) [7].

Tabla 3.1: Características de algunos STRs autosómicos más empleados [1].

<i>Locus</i>	<i>Localización cromosómica</i>	<i>Posición Física</i>	<i>Categoría; motivo de repetición</i>	<i>Rango alélico</i>
D1S1656	1q42	Chr 1 (230.905 Mb)	Compuesta; TAGA	8 a 20.3
TPOX	2p25.3 thyroid peroxidase, Intrón 10	Chr 2 (1.493 Mb)	Simple; AATG	4 a 16
D2S441	2p14	Chr 2 (68.239 Mb)	Compuesta; TCTA/TCAA	8 a 17
D2S1338	2q35	Chr 2 (218.879 Mb)	Compuesta; TGCC/TTCC	10 a 31
D3S1358	3p21.31	Chr 3 (45.582 Mb)	Compuesta; TCTA/TCTG	6 a 26
FGA	4q31.3 alpha fibrinogen, Intrón 3	Chr 4 (155.509 Mb)	Compuesta; CTTT/TTCC	12.2 a 51.2
D5S818	5q23.2	Chr 5 (123.111 Mb)	Simple; AGAT	4 a 29
CSF1PO	5q33.1 c-fms protooncogene, Intrón 6	Chr 5 (149.455 Mb)	Simple; AGAT	5 a 17
SE33 (ACTBP2)	6q14 beta-actin related pseudogene	Chr 6 (88.987 Mb)	Compleja; AAAG	3 a 49
D6S1043	6q15	Chr 6 (92.450 Mb)	Compuesta; AGAT/AGAC	8 a 25
D7S820	7q21.11	Chr 7 (83.789 Mb)	Simple; GATA	5 a 16
D8S1179	8q24.13	Chr 8 (125.907 Mb)	Compuesta; TCTA/TCTG	6 a 20
D10S1248	10q26.3	Chr 10 (131.093 Mb)	Simple; GGAA	7 a 19
TH01	11p15.5 tyrosine hydroxylase, Intrón 1	Chr 11 (2.192 Mb)	Simple; TCAT	3 a 14
vWA	12p13.31 Factor de von Willebrand, Intrón 40	Chr 12 (6.093 Mb)	Compuesta; TCTA/TCTG	10 a 25
D12S391	12p13.2	Chr 12 (12.450 Mb)	Compuesta; AGAT/AGAC	13 a 27.2
D13S317	13q31.1	Chr 13 (82.692 Mb)	Simple; TATC	5 a 17
Penta E	15q26.2	Chr 15 (97.374 Mb)	Simple; AAAGA	5 a 32
D16S539	16q24.1	Chr. 16 (86.386 Mb)	Simple; GATA	4 a 17
D18S51	18q21.33	Chr 18 (60.949 Mb)	Simple; AGAA	5.3 a 40
D19S433	19q12	Chr 19 (30.416 Mb)	Compuesta; AAGG/TAGG	5.2 a 20
D21S11	21q21.1	Chr 21 (20.554 Mb)	Compleja; TCTA/TCTG	12 a 43.2
Penta D	21q22.3	Chr 21 (45.056 Mb)	Simple; AAAGA	1.1 a 19
D22S1045	22q12.3	Chr 22 (37.536Mb)	Simple; ATT	7 a 20

1.1. Generación de Diversidad en Microsatélites

La tasa de mutación en microsatélites es relativamente alta, entre 10^{-2} y 10^{-9} mutaciones por locus por generación en eucariotas (Tabla 3.2) [8,9]. La tasa de mutación depende y varía en función del motivo de repetición, el contenido GC en la región flanqueante, el tamaño alélico, la posición cromosómica, el sexo, la edad, el tipo de cromosoma y el genotipo. Las alteraciones se observan principalmente en cambios en el número de repeticiones [8–10]. Se ha visto que la tasa de mutación incrementa en un determinado locus a medida que aumenta el número de unidades de repetición perfectas [11]. Se ha observado que existen diferencias entre las tasas de mutación en el aumento o disminución del número de repeticiones; la tasa de expansión es aparentemente lineal y constante, independientemente del tamaño del microsatélite. Sin embargo, la tasa de contracción es baja pero aumenta exponencialmente a medida que incrementa el tamaño [12]. Además, se ha comprobado que la estructura interna del microsatélite también influye en la tasa de mutación; ésta es mucho mayor en microsatélites complejos que en simples, y el número de repeticiones adicionales fuera del mayor casete incrementa la variabilidad de éste [13].

Tabla 3.2: Tasa de mutación de los STR autosómicos más empleados [14].

Locus	Tasa de mutación	Locus	Tasa de mutación
CSF1PO	0,16%	D16S539	0,11%
FGA	0,28%	D18S51	0,22%
TH01	0,01%	D21S11	0,19%
TPOX	0,01%	Penta D	0,14%
VWA	0,17%	Penta E	0,16%
D3S1358	0,12%	D2S1338	0,12%
D5S818	0,11%	D19S433	0,11%
D7S820	0,10%	SE33	0,64%
D8S1179	0,14%	D13S317	0,14%

Existen dos mecanismos principales que explican estas tasas de mutación tan altas; deslizamiento durante la replicación y recombinación.

- *Deslizamiento durante la replicación:* Este tipo de mutaciones se dan debido a la disociación de la hebra de ADN seguido de un realineamiento erróneo dando lugar a la ganancia o pérdida de una unidad de repetición. Aunque generalmente este proceso da lugar a mutaciones de un sólo paso, estudios en

dinucleótidos revelan que un 63% de los eventos observados implican múltiples unidades de repetición [9,15].

- *Recombinación*: Otro mecanismo de mutación es el proceso de recombinación, que interfiere en la longitud del núcleo de repetición por un entrecruzamiento asimétrico o por conversión génica [9,10]. Fenómenos de recombinación no recíproca también dan lugar a cambios en el número de repeticiones [9].

1.2. Aplicaciones

1.2.1. *aSTRs como herramienta en estudios de criminalística*

Cuando se necesita analizar vestigios mínimos, degradados u otras muestras (como esperma, saliva, pelos, cabellos sin bulbo, fragmentos óseos, etc.), los STRs son una herramienta muy adecuada. El uso de los STRs se ha extendido tanto, que cada vez se conocen y reconocen mejor los distintos resultados posibles que se esperan del análisis de muestras de origen criminalístico; debido, entre otros factores, a la posibilidad de diferenciar fácilmente muestras compuestas por mezclas de fluidos procedentes de distintos individuos [16,17]. Además, la mejora de las técnicas de recuperación de ADN han permitido la obtención de éste, no sólo de restos biológicos encontrados en la escena del crimen, sino también de objetos que han estado en contacto con la víctima o el sospechoso [18–20].

1.2.2. *aSTRs como herramienta en estudios de identificación genética*

La identificación de cadáveres y restos óseos sólo era posible por medio de la antropología física y la odontología forense hasta la aparición de los STRs. Diversos estudios han demostrado la utilidad de la amplificación del ADN mediante PCR en la identificación de cadáveres humanos [21]. Aún en restos muy antiguos es posible analizar los STRs partiendo, por ejemplo, de piezas dentales o fragmentos óseos [22–25], sobre todo cuando se emplean marcadores mini-STRs, marcadores con tamaños de amplicón pequeños, ya que los primers hibridan muy cercanos al motivo de repetición [24].

En situaciones de grandes catástrofes, generalmente se encuentran restos cadavéricos fragmentados o quemados y son difícilmente identificables. Los STRs se revelan como un procedimiento de gran utilidad respecto a la asignación de restos cadavéricos a individuos y la identificación de los mismos, como se ha

demostrado en los atentados del 11-S en EEUU [26], 11-M en Madrid [27] o accidentes como el de Barajas el 20 de Agosto del 2008.

1.2.3. aSTRs como herramienta en estudios de filiación

El análisis de los polimorfismos de ADN tiene como objetivo la confirmación de que un individuo es o no el padre biológico de otro. Los test de paternidad se basan en que, en ausencia de mutaciones, los menores reciben un alelo de cada progenitor en cada locus genético estudiado. En muy poco tiempo esta prueba se ha estandarizado, obteniendo gran información con mucha rapidez y menor coste relativo. Sin embargo, se han de tener en cuenta las mutaciones que se pueden dar en la línea germinal de alguno de los progenitores. Por ello, existen múltiples estudios que analizan la tasa de mutación de los marcadores genéticos de rutina a través del análisis de comparaciones de los alelos entre progenitores y descendientes [14,28].

No obstante, en múltiples ocasiones no se encuentra el escenario ideal en estudios de filiación y se han de recurrir a parientes lejanos; tíos, sobrinos, semihermanos... En estos casos, los STR disponibles comercialmente no son suficientes para resolverlos. Debido al patrón de herencia única, se han determinado haplotipos genéticos de marcadores STR autosómicos como solución en este tipo de relaciones familiares complejas [29].

1.2.4. Estudio de la estructura de las poblaciones

Es de gran importancia determinar las frecuencias alélicas poblacionales de los marcadores STR autosómicos. Se ha comprobado que las frecuencias alélicas no son homogéneas en todas las poblaciones sino que varían entre ellas, habiendo incluso alelos casi exclusivos de ciertas poblaciones o con frecuencias alélicas mucho mayores a las encontradas en el resto. Actualmente, se han descrito un gran número de poblaciones humanas; no sólo se han caracterizado grandes poblaciones [30–33] sino también poblaciones minoritarias [34–36]. Además, el estudio de los marcadores STR autosómicos permite establecer relaciones genéticas entre poblaciones cercanas [37,38].

1.2.5. Estudio de la evolución de las poblaciones

Los marcadores aSTR se han empleado en antropología genética como marcadores informativos de ancestralidad para reconstruir la expansión humana, así como para interpretar la historia evolutiva de las poblaciones humanas para conocer mejor el origen de las poblaciones, sus migraciones y los fenómenos de mestizaje.

Un ejemplo de esto se puede observar en poblaciones nativas americanas; el alelo con repeticiones 9 en el marcador D9S1120 únicamente está presente con frecuencias variables a lo largo de todo el continente americano. Sin embargo, este mismo alelo no se ha observado en poblaciones asiáticas. Las únicas poblaciones asiáticas que lo presentan son las poblaciones del oeste de Bering, cosa que apoya la teoría de un ancestro común entre las poblaciones modernas de ambos continentes (Figura 3.1) [39,40].



Figura 3.1: Distribución del alelo 9 *D9S1120* (naranja). Imagen tomada de Crawford et al. 2013 [40].

1.2.6. Otras aplicaciones

- Caracterización de líneas celulares

Las líneas celulares se pueden contaminar muy fácilmente durante su manipulación en el laboratorio dando lugar a mezclas o a líneas celulares nuevas.

La *American Type Culture Collection* junto con otros distribuidores internacionales de líneas celulares están llevando a cabo la identificación genética de las líneas

celulares humanas [41]. El tipado de los STRs permite descubrir rápidamente las posibles contaminaciones cruzadas entre líneas celulares así como desarrollar un estándar de referencia universal para la caracterización de las líneas celulares humanas.

- *Estudio de enfermedades*

Se desconoce una gran parte de las funciones de los microsateélites, pero se han relacionado tanto con la regulación de la expresión génica [42] como con enfermedades humanas [43]. Son muchas las enfermedades humanas que están asociadas con STRs. Las más conocidas son las relacionadas con trinucleótidos, tanto las causadas por expansiones dentro del gen, como la enfermedad de Huntington [43,44] o las ataxias espinocerebelosas [45], como las causadas por expansiones en regiones no codificantes, como el retraso mental por X-Frágil o la distrofia miotónica [43].

- *Detección de tumores*

Se ha demostrado que los microsateélites son inestables en muchas enfermedades hereditarias y en gran variedad de cánceres humanos [2,46]. La pérdida de la heterozigosidad (deleciones alélicas) y la inestabilidad de los microsateélites (inserciones alélicas) son el resultado de problemas en el mecanismo de reparación del ADN en casos de pérdida alélica.

La inestabilidad genética detectada en tejidos tumorales puede dar lugar a problemas en la identificación forense del individuo, sobre todo en los casos en los que la única muestra de referencia del individuo es una biopsia de tejido tumoral [47].

2. MATERIAL Y MÉTODOS

2.1 Población de Estudio

Se han estudiado un total de 245 individuos adultos no relacionados entre sí de la región sureste de la Península Ibérica [Almería (79), Granada (94), Málaga (79)], ver capítulo 2 páginas 39-40. Todos los donantes fueron informados del propósito de este estudio, aceptaron su participación y firmaron un consentimiento informado según la guía ética de la declaración de Helsinki.

2.2 Análisis Genético Molecular de la Muestra

La toma de muestras, así como su posterior análisis, están estrictamente registrados por rigurosos controles, tanto intralaboratorio como interlaboratorio.

2.2.1. Extracción

Se tomaron muestras del epitelio bucal de todos los individuos mediante hisopos estériles de algodón. El ADN genómico se obtuvo mediante un protocolo de extracción orgánica con fenol:cloroformo:alcohol isoamílico (25:24:1) (Apéndice I).

2.2.2. Cuantificación

La estimación de la cantidad y calidad de ADN nuclear de las muestras se ha realizado a través de electroforesis en geles de agarosa al 0,8% (Apéndice I).

2.2.3. Amplificación

Las muestras se amplificaron para 16 marcadores de tipo STR: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, AMELOGENINA, D5S818, FGA incluidos en un kit tipo multiplex preparado para ser detectado por electroforesis capilar con emisión de fluorescencia [48]. La mitad de las muestras se amplificaron con el kit AmpFlSTR® Identifiler® (Tabla 3.3) y la otra mitad con el kit AmpFlSTR® Identifiler®Plus (Tabla 3.4) (Applied Biosystems, Foster City, CA).

Mejoras en los reactivos y los parámetros de amplificación del kit AmpFlSTR® Identifiler®Plus:

- Aumenta la calidad de los resultados en casos que presentan inhibición.
- Proporciona mayor altura de los picos en muestras mínimas y en mezclas con contribuciones desbalanceadas.
- Proporciona mayor flexibilidad analítica a través de protocolos de amplificación de 28 y 29 ciclos.
- Agiliza el flujo de trabajo reduciendo el tiempo de amplificación.
- Mantiene los amplicones pequeños presentes en el kit AmpFlSTR® Identifiler® para el trabajo con muestras degradadas, y además, incluye características del kit MiniFiler™ para maximizar los resultados con muestras extremadamente degradadas.
- Mantiene las secuencias de los primers del kit AmpFlSTR® Identifiler® para facilitar la comparación de las muestras.

Estos permiten el análisis simultáneo de 15 loci tipo STR y un locus indicador de sexo (amelogenina). Uno de los primers de cada pareja está marcado con un fluorocromo que confiere a cada marcador un color (verde, amarillo, azul y rojo) cuando son analizados en el aparato de electroforesis capilar [49,50].

- Azul -FAM: D8S1179, D21S11, D7S820, CSF1PO
- Verde-VIC: D3S1358, TH01, D13S317, D16S539, D2S1338
- Amarillo-NED: D19S433, vWA, TPOX, D18S51
- Rojo-PET: AMELOGENINA, D5S818, FGA

Tabla 3.3: Componentes, composición y volumen del Kit AmpFlSTR® Identifiler® [49].

Reactivo	Composición	Volumen
PCR Reaction Mix	MgCl ₂ , dNTPs, Suero de albúmina bovina y 0,5% NaO ₃	4µl
Identifiler® Primer Pair Mix	Primers marcados para los loci	2µl
Ampli Taq Gold® DNA Polymerase	Enzima polimerasa 5u/µl	0,2µl(2,5u)
Control DNA 9947A	ADN humano 0,1ng/µl	4µl

Tabla 3.4: Componentes, composición y volumen del Kit AmpFISTR® Identifiler® Plus.

Reactivo	Composición	Volumen
PCR Reaction Mix	MgCl ₂ , dNTPs, Suero de albúmina bovina y 0,5% NaO ₃ , AmpliTaq Gold® DNA Polimerasa	4µl
Identifiler® Plus Primer Set	Primers marcados para los loci	2µl
Control DNA 9947A	ADN humano 0,1ng/µl	4µl

La cantidad de ADN por muestra osciló entre 0,5-1ng/µl y el volumen final de reacción fue de 10µl.

La amplificación de todos los loci se realizó en dos termocicladores 2720 Applied Biosystems y un termociclador T3 Biometra® (Figuras 3.2 y 3.3).

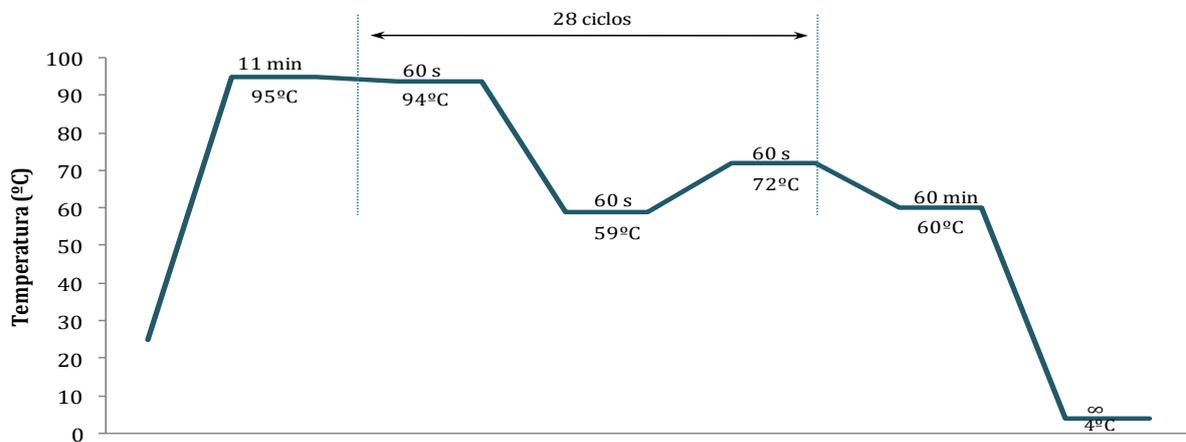


Figura 3.2: Gráfico de las condiciones de amplificación del kit AmpFISTR® Identifiler®,

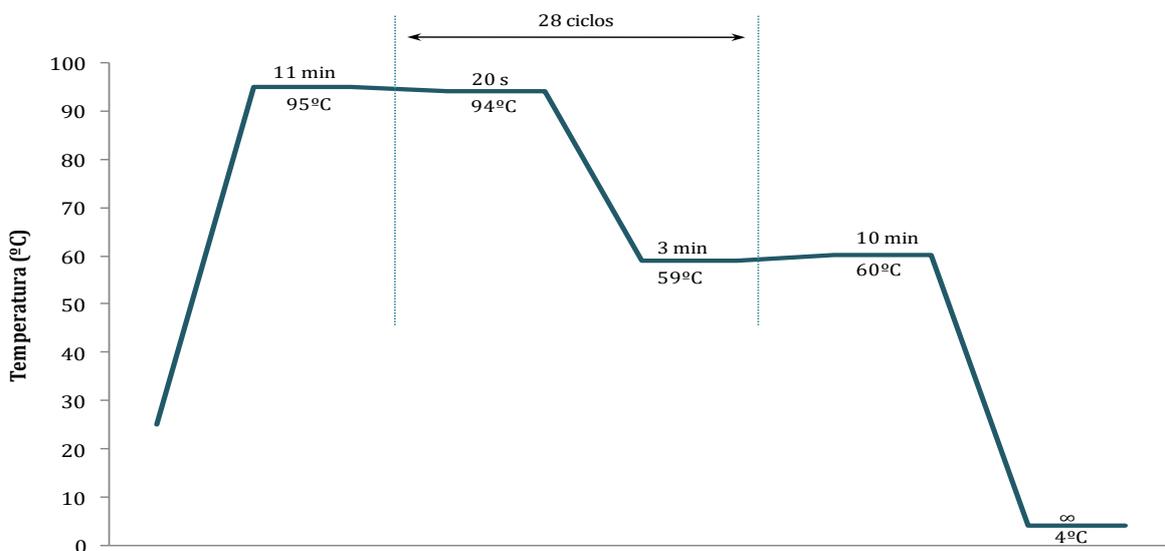


Figura 3.3: Gráfico de las condiciones de amplificación del kit AmpFISTR® Identifiler® Plus.

En todas las reacciones se han amplificado sistemáticamente dos controles:

- Control positivo: verificar que la reacción de amplificación ha sido correcta.
- Control negativo: descartar una posible contaminación por ADN extraño.

2.2.4. Visualización de los resultados

Las muestras amplificadas fueron visualizadas por electroforesis capilar en un analizador genético ABI Prism 310 de Applied Biosystems®.

Para la correcta detección de los fragmentos de ADN nuclear amplificados, se han de preparar las muestras antes de separarlas por electroforesis capilar (Apéndice II).

Para preparar las muestras, se añadió en cada tubo los siguientes volúmenes por muestra:

- 25 µl de formamida desionizada
- 0,5 µl del estándar de tamaño (500pb marcado en naranja- LIZ)
- 1 µl de producto amplificado

La electroforesis se llevó a cabo con Performance Optimized Polymer 4 (POP-4™; Life Technologies) con un filtro G5 para las muestras amplificadas con el kit AmpFlSTR® Identifiler®, y un filtro G5v2 para las muestras amplificadas con el kit AmpFlSTR® Identifiler® Plus.

Las condiciones de electroforesis fueron las siguientes:

- Tiempo de inyección..... 5s
- Voltaje de inyección..... 15kV
- Voltaje de electroforesis..... 15kV
- Temperatura de electroforesis..... 60°C
- Tiempo de electroforesis28min

Los electroferogramas fueron visualizados mediante GeneMapper ID-X 1.1; que permite determinar el tamaño de los fragmentos analizados. Se emplearon los *bin set* para STRs autosómicos facilitados por la casa comercial. Los alelos se designaron de acuerdo con el número de unidades de repetición basándose en un *ladder* alélico (recomendaciones de ISFG) [51].

La calidad del tipado de las muestras, así como de la designación de alelos fueron garantizadas por el análisis simultáneo de una muestra control de genotipo conocido.

2.3 Análisis Estadístico de la Muestra

Los datos de los perfiles genéticos de los individuos que se prestaron para realizar el estudio poblacional fueron introducidos en tablas mediante un código y su perfil genético.

El cálculo de las frecuencias de las repeticiones encontradas en cada una de las poblaciones se realizó por distintos métodos. En primer lugar, la obtención de las frecuencias alélicas muestrales se determinó por el método clásico del recuento en base a los alelos observados (Li, 1976). En segundo lugar, se recalculó este mismo parámetro con dos programas estadísticos, *Arlequin 3.5.1.3* [52] y *PowerStatsv12* (Promega® Corporation) [53]. En todos los casos se han de obtener los mismos resultados.

La heterocigosidad observada y esperada se calcularon mediante estos dos últimos programas, *PowerStatsv12* y *Arlequin 3.5.1.3*. Éste último también sirvió como herramienta para calcular el equilibrio de Hardy-Wienberg con 1 000 000 de pasos en la cadena de Markov y 100 000 pasos de dememorización.

Todos los parámetros forenses más relevantes se calcularon mediante la hoja de cálculo *PowerStatsv12*, obteniendo el poder de discriminación (PD), el poder de exclusión (PE), el índice de contenido polimórfico (PIC), la probabilidad de coincidencia (MP) y el índice típico de paternidad (TPI).

Para determinar la subestructura genética de la población del Sudeste de la Península Ibérica, se testó la hipótesis de una distribución aleatoria de los individuos entre parejas de poblaciones, mediante un test exacto de diferenciación de la población y el análisis de la varianza molecular (AMOVA) con 10 000 repeticiones experimentales mediante *Arlequin 3.5.1.3*. Además, se utilizó *STRUCTURE 2.3.4* [54,55] para inferir grupos de individuos a partir de los datos genéticos de loci no ligados.

Para calcular la distancia genética con poblaciones tomadas desde la bibliografía se utilizó la aplicación *Gendist*, del paquete informático, Phylip 3.69 [56], que calcula las distancias entre poblaciones a partir de las frecuencias observadas en los alelos de cada una de las poblaciones estudiadas. Todas las poblaciones estudiadas quedan resumidas en la tabla 3.5. Para una mejor visualización de los resultados, las distancias genéticas calculadas (Nei, Reynold y Cavalli-Sforza) se representaron mediante escalamiento multidimensional no métrico [57] con *IBM SPSS Statistics 20* y mediante un árbol filogenético basado en el método Neighbor-Joining con el software *Treeview 1.6.6* (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>) [58]. Para poder realizar la representación en dendograma, se ha

de realizar un tratamiento previo de los datos resultantes de las distancias genéticas mediante la aplicación *Neighbor*, del paquete informático, Phylip 3.695 [56].

Se llevó a cabo un análisis de correspondencias con el programa *Statistica v9.1* (Statsoft Inc. Tulsa USA), a fin de entender la asociación entre las frecuencias alélicas y las poblaciones. Se eliminaron dos marcadores del análisis ya que no se tenía información de ellos en todas las poblaciones (D2S1338 y D19S433).

Para estimar la proporción individual de ancestralidad se utilizó *STRUCTURE 2.3.4* [54,55]. Para todas las simulaciones y cálculos, se asumieron los modelos de mestizaje y no mestizaje, incluyendo información poblacional y asumiendo un modelo de frecuencias alélicas independientes. Se estableció un periodo de *burn-in* de 50 000 iteraciones seguido de 100 000 iteraciones adicionales, estimando de 1 a 10 contribuciones (K). Cada carrera se repitió 10 veces y se calcularon las probabilidades para cada K en cada grupo de carreras. Las 10 réplicas para cada K se evaluaron con *CLUMPP 1.1.2* (<http://www.stanford.edu/group/rosenberglab/clumpp.html>) [59]. Los grupos combinados resultantes se visualizaron mediante *DISTRUCT 1.1* (<http://www.stanford.edu/group/rosenberglab/distruct.html>) [60].

Tabla 3.5: Poblaciones, abreviatura, situación geográfica, marcadores estudiados y autor del estudio publicado.

POBLACIÓN	Código	País	a-STRs	N	Bibliografía
ANÁLISIS DE FRECUENCIAS ALÉLICAS					
Granada_Málaga_Almería	GMA	España	15	245	Presente estudio
Bereberes de Asni	BAsMOR	Marruecos	15	105	Coudray <i>et al.</i> , 2007 [61]
Bereberes de Bouhria	BBoMOR	Marruecos	15	104	Coudray <i>et al.</i> , 2007 [61]
Árabes de Marruecos	ArbMOR	Marruecos	15	138	Abdin <i>et al.</i> , 2003 [62]
Alto Egipto	UppEGY	Egipto	15	265	Omran <i>et al.</i> , 2009 [63]
Túnez	TunTUN	Túnez	15	200	Brandt-Casadevall <i>et al.</i> , 2002 [64]
Bereberes de Kesra	BKeTUN	Túnez	15	44	Cherni <i>et al.</i> , 2005 [65]
Turquía	TurTUR	Túnez	13	198	Akbasak <i>et al.</i> , 2001 [66]
Siria	SyrSYR	Siria	15	121	Abdin <i>et al.</i> , 2003 [62]
País Vasco	BasSPA	España	13	100	Pérez-Lezaun <i>et al.</i> , 2000 [67]
Cataluña	CatSPA	España	13	96	Pérez-Lezaun <i>et al.</i> , 2000 [67]
Andalucía	AndSPA	España	13	72	Pérez-Lezaun <i>et al.</i> , 2000 [67]
Norte de Portugal	NorPOR	Portugal	16	200	Pinheiro <i>et al.</i> 2005 [68]
Centro de Portugal	CenPOR	Portugal	15	2125	Lopes <i>et al.</i> 2009 [69]
Norte de España	NEsSPA	España	16	204	Paredes <i>et al.</i> , 2003 [70]
Italia	ItaITA	Italia	15	400	Brisighelli <i>et al.</i> 2009 [71]
Grecia	GreGRE	Grecia	16	300	Sánchez-Diz <i>et al.</i> 2008 [72]
Cabinda	CabANG	Angola	16	110	Beleza <i>et al.</i> 2004 [73]
Maputo	MapMOZ	Mozambique	15	135	Alves <i>et al.</i> 2004 [74]
ANÁLISIS DE ESTRUCTURAS					
Marruecos	Mor	Marruecos	15	210	Bouabdellah <i>et al.</i> , 2008 [75]
Libia	Lib	Libia	15	230	Elmrghni <i>et al.</i> , 2012 [76]
Somalia	Som	Somalia	15	404	Tillmar <i>et al.</i> , 2009 [77]
Sud África	SrAf	Sud África	15	324	Schlebusch <i>et al.</i> , 2012 [78]
España	Spa	España	15	202	Martínez-González <i>et al.</i> Datos no publicados
Huelva	Hue	España	15	114	Coudray <i>et al.</i> , 2007 [79]
Italia	Ita	Italia	13	400	Brisighelli <i>et al.</i> , 2009 [71]
Suecia	Swe	Suecia	15	200	Tillmar <i>et al.</i> Datos no publicados

3. RESULTADOS Y DISCUSIÓN

3.1 Estudio Poblacional

La distribución de las frecuencias alélicas de los 15 loci STR estudiados queda resumida en la tabla 3.7. La tabla 3.6 muestra el poder de discriminación, poder de exclusión, el índice de contenido polimórfico, la probabilidad de coincidencia, el índice típico de paternidad y la heterozigosis observada y esperada.

Los marcadores más informativos son D18S51 y FGA; mientras que el marcador menos informativo es TPOX. El poder de discriminación combinado de toda la población es $1 - 5,55362 \cdot 10^{-18}$.

Tabla 3.6: Resumen de los parámetros forenses más importantes, heterozigosis observada y esperada y desviación del equilibrio Hardy-Weinberg (HWE).

	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S739
PD	0,940	0,947	0,939	0,866	0,916	0,926	0,924	0,915
PE	0,577	0,677	0,526	0,449	0,562	0,562	0,540	0,484
PIC	0,790	0,670	0,790	0,670	0,740	0,720	0,750	0,730
MP	0,060	0,125	0,061	0,134	0,084	0,074	0,076	0,085
TIP	2,360	1,830	2,080	1,730	2,270	2,270	2,150	1,880
Ho	0,788	0,841	0,759	0,714	0,780	0,780	0,767	0,735
He	0,814	0,827	0,816	0,725	0,781	0,798	0,784	0,770
p	0,633	0,193	0,134	0,079	0,904	0,585	0,634	0,613

	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
PD	0,965	0,912	0,934	0,832	0,966	0,875	0,964
PE	0,725	0,474	0,491	0,332	0,766	0,451	0,708
PIC	0,850	0,680	0,780	0,610	0,860	0,670	0,804
MP	0,035	0,088	0,066	0,168	0,034	0,125	0,036
TIP	3,710	1,840	1,910	1,360	4,380	1,750	3,490
Ho	0,865	0,730	0,739	0,633	0,886	0,714	0,857
He	0,868	0,767	0,807	0,665	0,873	0,725	0,865
p	0,594	0,223	0,093	0,124	0,632	0,425	0,502

Todos los loci estudiados se encuentran en equilibrio Hardy-Weinberg para esta población ($p > 0,05$). Todos presentan valores normales de heterozigosis y no existen signos de ligamiento entre ellos.

Tabla 3.7: Frecuencias alélicas para los 15 loci STR estudiados en la población GMA.

<i>Alelo</i>	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S739	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
6			0,004			0,218									
6.3						0,002									
7			0,033	0,002		0,159									
7.3						0,002									
8	0,006		0,131	0,008		0,147	0,161	0,022				0,484		0,004	
9	0,018		0,131	0,014		0,188	0,061	0,118				0,135		0,022	
9.3						0,271									
10	0,086		0,280	0,290		0,010	0,041	0,047				0,080	0,006	0,080	
11	0,096		0,202	0,318		0,002	0,302	0,251		0,006		0,278	0,012	0,353	
12	0,118		0,171	0,296			0,290	0,339		0,109	0,002	0,024	0,182	0,327	
12.2										0,002					
13	0,294		0,047	0,055	0,004		0,090	0,190		0,246	0,002		0,153	0,196	
13.2										0,008					
14	0,233		0,002	0,012	0,076		0,051	0,031		0,371	0,116		0,159	0,014	
14.2										0,016					
15	0,127				0,300		0,004			0,145	0,129		0,129	0,002	
15.2										0,035					
16	0,020				0,235			0,002	0,045	0,041	0,282		0,116		
16.2										0,012					
17	0,002				0,171				0,247	0,006	0,229		0,104		0,002
17.2										0,002					
18					0,200				0,086		0,169		0,055		0,016
19					0,014				0,114		0,061		0,045		0,063
20									0,147		0,010		0,016		0,129
21									0,039				0,016		0,210

3.2 Estudio de la Estructura de la Población

Muchos autores han estudiado las relaciones genéticas que existen entre el Norte de África y las poblaciones del Sur de Europa [80,81] y de la Península Ibérica [67,82–84] para determinar el legado genético que queda presente en las poblaciones actuales. Para enriquecer la información que se posee acerca de las relaciones genéticas entre el Sur de la Península Ibérica y los invasores del Norte de África, se analizarán las poblaciones de las provincias que formaban el antiguo Reino de Granada y se compararán con otras poblaciones españolas y norteafricanas.

Pero antes de comparar la población GMA con otras poblaciones se ha de determinar si puede ser tratada como una única población o se han de considerar como tres poblaciones independientes.

Para determinar si el estudio de los aSTRs permite establecer una subestructura de la población de estudio en función de las tres provincias de las que proceden las muestras, se realizaron estudios de AMOVA y STRUCTURE.

El análisis AMOVA no muestra una subestructura genética significativa entre los individuos de las tres provincias de estudio, por lo que podemos agruparlos dentro de la misma población ($p > 0,05$). Cabe destacar que el 97,77% de la variación determinada a través del estudio de los aSTRs es la que se observa entre los individuos de una misma población (Tabla 3.8). Estos resultados confirman a los STRs como unos buenos marcadores forenses capaces de diferenciar individuos de una misma población entre sí de una forma muy eficaz.

Tabla 3.8: Diseño AMOVA y resultados de las tres provincias del Sudeste de la Península Ibérica.

Diseño	Fuente de Variación	d.f	Suma de cuadrados	Componentes de Varianza	Porcentaje de variación
3 Grupos: Grupo 1= Granada Grupo 2= Málaga Grupo 3= Almería	Entre poblaciones	2	10,234	-0,00542 Va	-0,09
	Entre individuos dentro de las poblaciones	242	1451,158	0,13602 Vb	2,32
	Dentro de las poblaciones	245	1402,500	5,72449 Vc	97,77
	Total	489	2863,892	5,85509	
					Índices de fijación
					FIS: 0,02321
					FST: -0,00099
					FIT: 0,02230

El estudio de la subestructura genética entre estos tres grupos mediante el programa STRUCTURE corrobora estos resultados. El modelo con mayor valor de probabilidad es $K=1$ ($\ln P[D] = -12981,88$) comparado con $K=3$ ($\ln P[D] = -13300,52$). Esto implica que los datos genéticos favorecen un único clúster para las tres subpoblaciones estudiadas.

Estos resultados se ajustan a las expectativas históricas y socioculturales ya que comparten el mismo origen además de una gran proximidad geográfica.

3.3 Estudio de Frecuencias Alélicas

Se realizaron estudios de análisis de correspondencias a través de las frecuencias alélicas de las poblaciones españolas y del norte de África con *Statistica v9.1*. Para facilitar la interpretación de los resultados se omitieron los marcadores en las figuras.

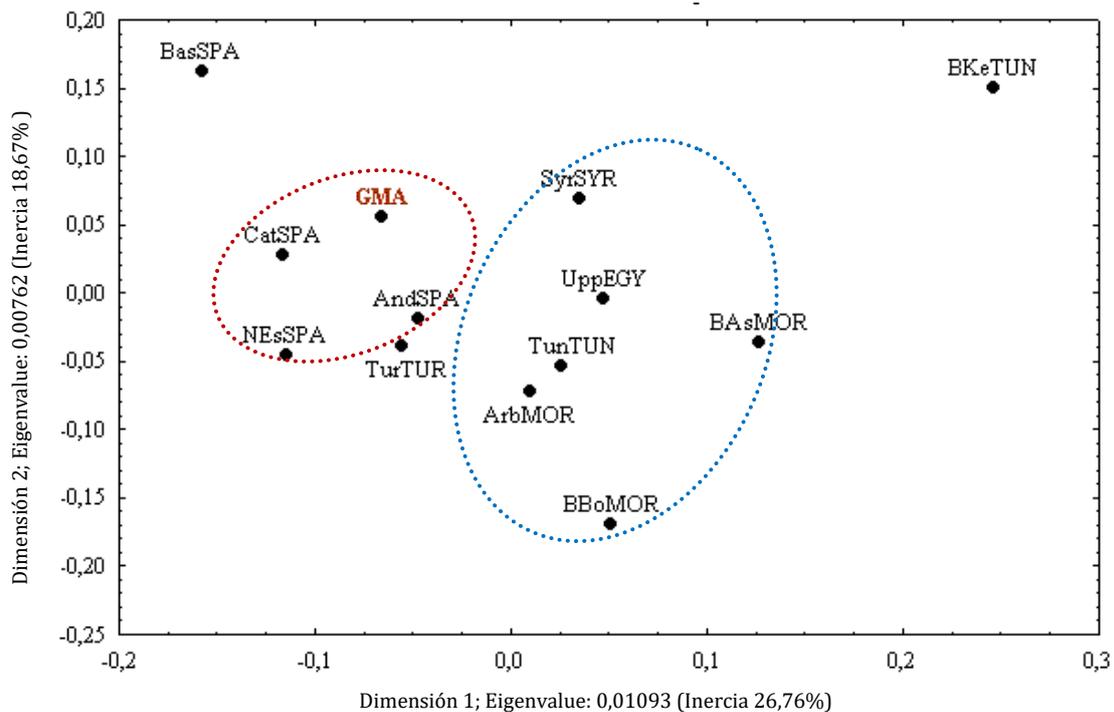


Figura 3.4: Análisis de correspondencia en 2 dimensiones que establece la relación entre las frecuencias alélicas y las poblaciones estudiadas.

En la figura 3.4, se observan dos grupos centrales, uno formado por las poblaciones españolas y la población turca, enmarcadas en color rojo; y las poblaciones marroquí, siria, egipcia y tunecina, enmarcadas en color azul. La población GMA se localiza dentro del grupo de poblaciones españolas siendo las más cercanas a ella la andaluza y catalana.

Hay dos poblaciones que, debido a sus frecuencias genéticas; se distancian del resto, los bereberes de Kesra y la población vasca española. La población del País Vasco se localiza alejada del resto de poblaciones españolas aunque en el mismo cuadrante. Este fenómeno se debe a las diferencias observadas en las frecuencias alélicas de los marcadores D13S317 (alelo 8), TPOX (alelo 12) y TH01 (alelo 9.3). Se sabe que la población del País Vasco tiene una alta proporción de alelos paleolíticos y una baja diversidad genética [85], esto confirma la posición que tiene esta población en el análisis de correspondencias.

3.4 Estudio de Distancias Genéticas

Las distancias genéticas de Nei, Reynold y Cavalli-Sforza se calcularon con la aplicación *Gendist*, del paquete informático, Phylip 3.69 [56]; para descifrar las relaciones genéticas entre la población GMA y 12 poblaciones tomadas de la literatura (Tabla 3.5). Se han tenido en cuenta las frecuencias alélicas de 13 loci STRs contemplados dentro del sistema CODIS.

Para una mejor visualización de los resultados, las distancias genéticas calculadas (Nei, Reynold y Cavalli-Sforza) se representaron mediante un árbol filogenético basado en el método Neighbor-Joining (Figura 3.5). Se pueden observar tres agrupaciones claramente; en verde las poblaciones europeas, con un color más oscuro las poblaciones españolas; en rojo las poblaciones del norte de África; y en azul las poblaciones subsaharianas.

Tras el estudio de los resultados del árbol filogenético se ha de destacar cómo las poblaciones italiana y griega son las más cercanas a los pueblos del norte de África, sobre todo aquellos más próximos a ellos en el Mediterráneo como son turcos, sirios y tunecinos. El hecho de que las poblaciones del norte de África se encuentren agrupadas junto a las poblaciones mediterráneas concuerda con estudios previos donde se detectan trazas genéticas y haplotipos similares a ambos lados del mediterráneo [86].

La población GMA se localiza dentro del grupo de poblaciones españolas. Cabe destacar, que la población del País Vasco, aún estando dentro del grupo de poblaciones españolas; se aleja mucho de las mismas debido a la gran conservación de alelos ancestrales paleolíticos [85].

Por otro lado, la gran diferencia genética que existe entre las poblaciones del norte de África y las poblaciones de África subsahariana queda manifestada por la gran distancia que existe entre ellas.

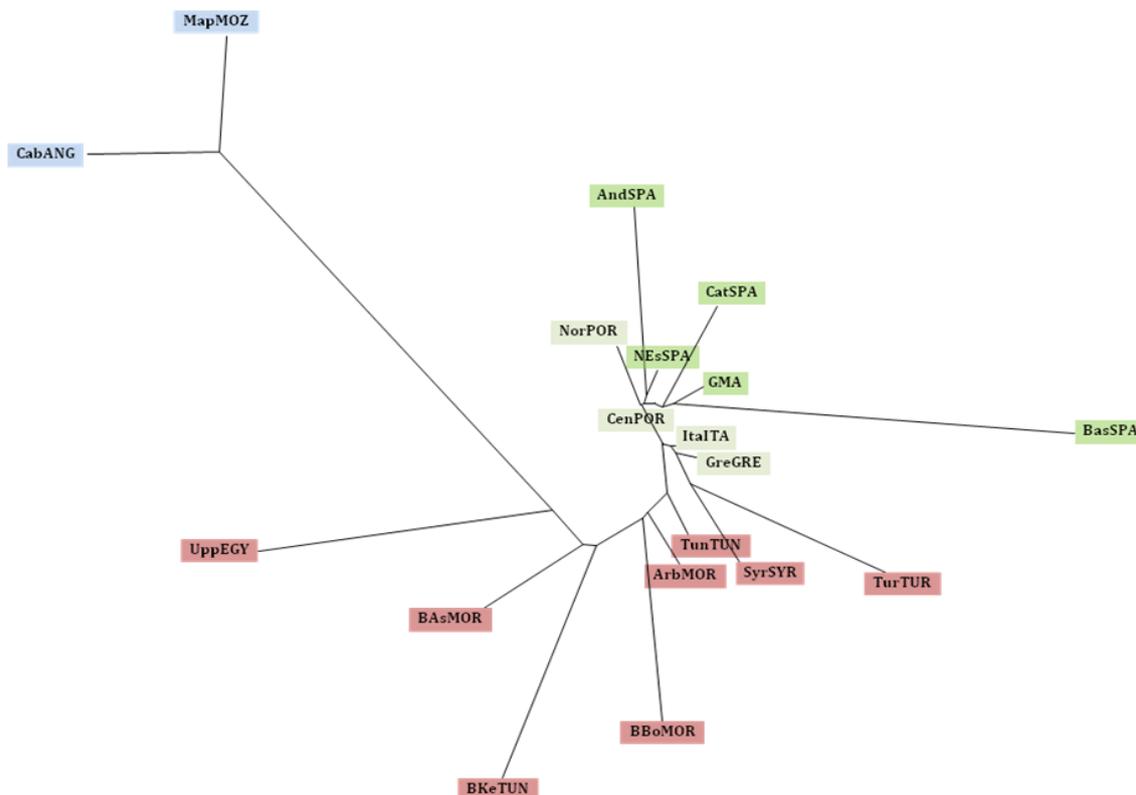


Figura 3.5: Dendrograma de las poblaciones europeas y africanas obtenido a partir de las distancias genéticas de Reynolds; calculado con *Treewiew*.

Además, para una mejor visualización de los resultados, las distancias genéticas calculadas (Nei, Reynold y Cavalli-Sforza) se representaron mediante escalamiento multidimensional no métrico [57] con *IBM SPSS Statistics 20* (Figura 3.6). La dimensión 1 separa claramente las poblaciones del norte de África (Árabes de Marruecos (ArbMOR), Bereberes de Bouhria (BBoMOR), Bereberes de Asni (BAsMOR), Bereberes de Kesra (BKeTUN), Túnez (TunTUN), Alto Egipto (UppEGY), Turquía (TurTUR) y Siria (SyrSYR)), localizadas en el área negativa; de las poblaciones Españolas (País Vasco (BasSPA), Cataluña (CatSPA), Noreste de España (NesSPA), Andalucía (AndSPA) y GMA), localizadas en el área positiva.

Podemos observar claramente dos clústeres que coinciden con la distribución geográfica de los individuos. La población GMA se localiza con el resto de poblaciones españolas.

A diferencia de lo detectado en el análisis de frecuencias, a parte de las poblaciones vasca y berebere de Kesra; se distancian las poblaciones turca, bereberes marroquí y Sur de Egipto. Debido al tipo de análisis de distancias y a la representación en PROXSCAL, las similitudes y proximidades se potencian con respecto a los análisis de frecuencias.

Basado en el estudio de 15 STRs autosómicos incluidos en este estudio, los análisis de distancias genéticas revelan que las poblaciones del norte de África han tenido muy poca influencia en la población GMA.

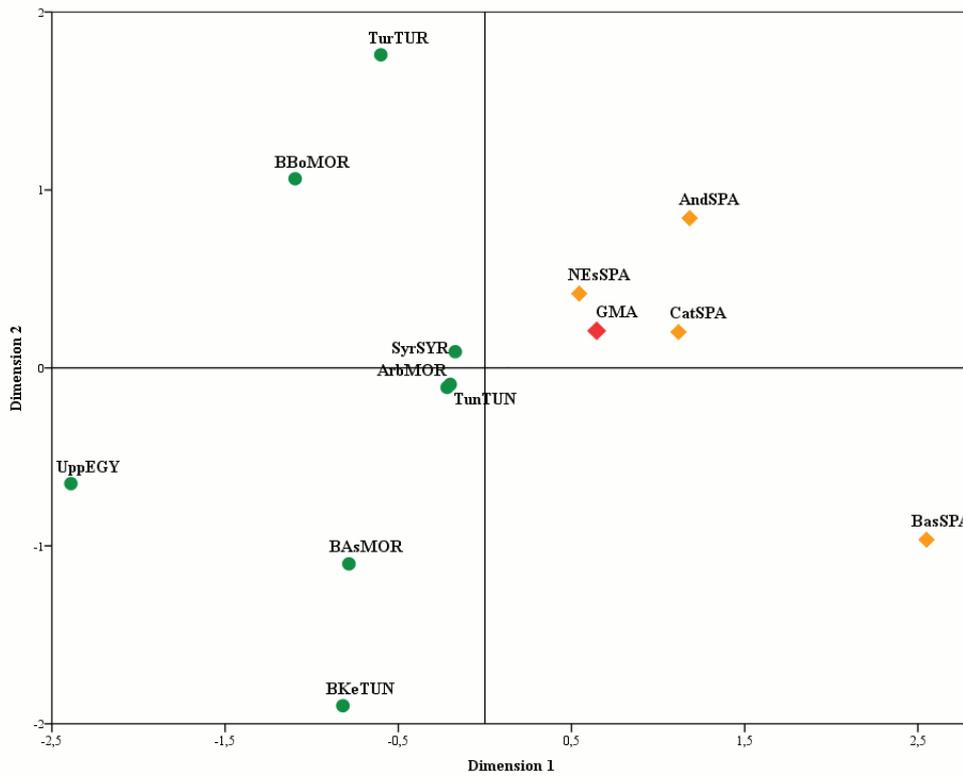


Figura 3.6:
MDS calculado a partir de la distancia de Nei; valor de stress 0,19979.

3.5 Estudio de los Componentes Ancestrales

Para complementar los análisis de frecuencias alélicas y distancias genéticas, se estudió la estructura de la población GMA comparándola con otras poblaciones tomadas de la bibliografía a través del programa STRUCTURE. Las características de las poblaciones empleadas quedan resumidas en la tabla 3.5.

El modelo que mejor explica las diferencias existentes entre las poblaciones estudiadas es el que asume 5 componentes ancestrales ($K=5$ ln P[D]= -94172,89).

En este análisis, mientras que la mayoría de los individuos muestran pertenencia a un solo clúster predominante correspondiente a su afiliación geográfica (Figura 3.7); la población marroquí presenta contribuciones de varios clústeres. Todas las poblaciones europeas muestran un clúster "europeo" distintivo (color rojo), la población somalí un componente "Este de África" (color verde) y los individuos sudafricanos están representados por el componente "Este de África" y otro componente "Sur de África" (color morado). Se puede ver que Libia representa su propio clúster (color azul claro) como consecuencia del alto ratio de consanguinidad que se han descrito en esta población [76].

Sin embargo, la correspondencia observada entre clústeres y poblaciones no queda patente en la muestra de Marruecos. Los individuos muestran pertenencia parcial a múltiples clústeres (europeo y subsahariano) como resultado de la mezcla de grupos étnicos que conforman esta población (árabes, bereberes y saharauis) [75]. Estos resultados sugieren identidades que no corresponden con la afiliación geográfica, lingüística o étnica.

La población GMA pertenece de forma inequívoca al clúster europeo sin mostrar ningún componente de ninguno de los clústeres africanos. Estos resultados pueden explicarse por la falta de interacción genética entre la población musulmana que habitaba el territorio y los conquistadores españoles. Los datos históricos señalan que los individuos musulmanes que habitaban el Reino de Granada fueron expulsados o aislados, y muy pocos fueron cristianizados y permanecieron en la región. Además, tras la ocupación de la ciudad de Granada por los españoles en 1492, se tomó gente del norte de la Península para repoblar la región, aislando aún más a los musulmanes que permanecieron en el territorio.

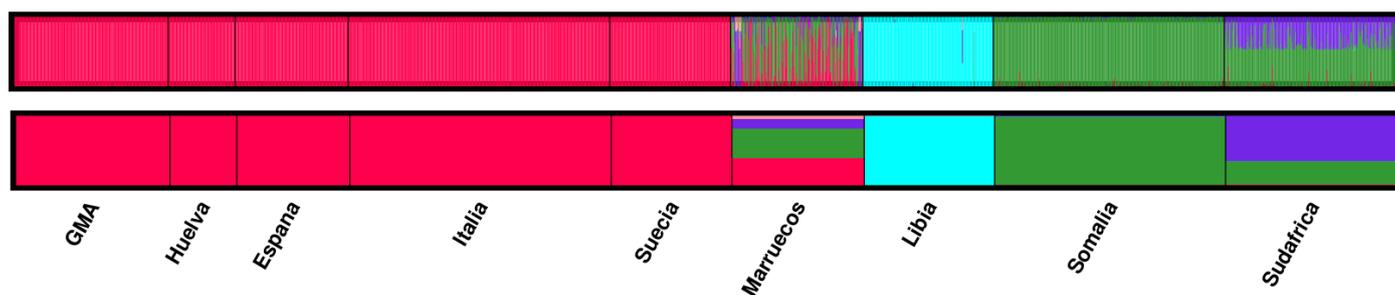


Figura 3.7: Análisis estructural de $K= 5$ para 9 poblaciones. Asignación media de cada individuo a los clústeres resultantes del análisis de estructura. Cada individuo se representa con una fina línea vertical, que está dividida en los colores asignados a cada K representando la pertenencia estimada de cada individuo a cada clúster K (arriba). Asignación media de cada población a los clústeres resultantes del análisis de estructura (abajo).

En conclusión, los datos genéticos obtenidos a través del estudio de STRs autosómicos revelan que ningún componente africano ha permanecido en la población del Sur de la Península Ibérica a pesar de haber estado ocupada por éstos durante 800 años. La expulsión de los pueblos del norte de África y la repoblación de la zona con gente del resto de la Península fue tan efectiva que es difícil identificar algún trazo del legado genético dejado por los antiguos pobladores. Las similitudes entre la población GMA y la población europea, o más concretamente otras poblaciones españolas, hace que las posibles diferencias existentes sean muy difíciles de identificar.

4. BIBLIOGRAFÍA

- [1] J.M. Butler, C.R. Hill, Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis, *Forensic Sci Rev.* 24 (2012) 15–26.
- [2] G. Filoglu, O. Bulbul, G. Rayimoglu, F.E. Yediay, T. Zorlu, S. Ongoren, et al., Evaluation of reliability on STR typing at leukemic patients used for forensic purposes., *Mol. Biol. Rep.* 41 (2014) 3961–72. doi:10.1007/s11033-014-3264-9.
- [3] J. Weber, M. P., Abundant class of human DNA polymorphism which can be typed using then polymerase chain reaction., *Am. J. Hum. Genet.* 44 (1989) 338–396.
- [4] E.P. Economou, A.W. Bergen, A.C. Warren, S.E. Antonarakis, The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome, *Proc. Natl. Acad. Sci. U. S. A.* 87 (1990) 2951–2954.
- [5] J.M. Butler, *Forensic DNA Typing-Biology, Technology, and Genetics of STR Markers DNA*, 3rd editio (2010) 660.
- [6] H. Ellegren, Microsatellites: simple sequences with complex evolution., *Nat. Rev. Genet.* 5 (2004) 435–445. doi:10.1038/nrg1348.
- [7] P. Vasquez, M. Garde, B. Martínez, Amplificación y tipado de marcadores genéticos moleculares de tipo microsatélites (STR's) autosómicos y del cromosoma Y, *Curso Online Genética Forense. Univ. Zaragoza.* (n.d.).
- [8] E. Buschiazzo, N.J. Gemmell, The rise, fall and renaissance of microsatellites in eukaryotic genomes, *BioEssays.* 28 (2006) 1040–1050. doi:10.1002/bies.20470.
- [9] P. Madesis, I. Ganapoulos, A. Tsaftaris, Microsatellites: Evolution and Contribution, in: S.K. Kantartzi (Ed.), *Microsatellites Methods Protoc.*, 1st ed., Springer Science, 2013: pp. 1–10.
- [10] Y.-C. Li, A.B. Korol, T. Fahima, A. Beiles, E. Nevo, Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review, *Mol. Ecol.* 11 (2002) 2453–2465. doi:10.1046/j.1365-294X.2002.01643.x.
- [11] A.M. Leopoldino, S.D.J. Pena, The mutational spectrum of human autosomal tetranucleotide microsatellites, *Hum. Mutat.* 21 (2003) 71–79. doi:10.1002/humu.10153.
- [12] X. Xu, M. Peng, Z. Fang, The direction of microsatellite mutations is dependent upon allele length., *Nat. Genet.* 24 (2000) 396–399. doi:10.1038/74238.
- [13] M. Kayser, R. Kittler, A. Erler, M. Hedman, A.C. Lee, A. Mohyuddin, et al., A comprehensive survey of human Y-chromosomal microsatellites, *Am. J. Hum. Genet.* 74 (2004) 1183–1197.
- [14] A. Relationship Testing Program Unit, Annual report summary for testing, 2008.
- [15] Q.-Y. Huang, F.-H. Xu, H. Shen, H.-Y. Deng, Y.-J. Liu, Y.-Z. Liu, et al., Mutation patterns at dinucleotide microsatellite loci in humans., *Am. J. Hum. Genet.* 70 (2002) 625–634. doi:10.1086/338997.
- [16] I.W. Evett, C. Buffery, G. Willott, A guide to interpreting single locus profiles of DNA mixtures in forensic cases, *J. Forensic Sci. Soc.* 31 (1991) 41–47.
- [17] J.-P. Han, F. Yang, C. Xu, Y.-L. Wei, X.-C. Zhao, L. Hu, et al., A new strategy for sperm isolation and STR typing from multi-donor sperm mixtures., *Forensic Sci. Int. Genet.* 13 (2014) 239–46. doi:10.1016/j.fsigen.2014.08.012.

- [18] J.-A. Bright, S.F. Petricevic, Recovery of trace DNA and its application to DNA profiling of shoe insoles., *Forensic Sci. Int.* 145 (2004) 7–12. doi:10.1016/j.forsciint.2004.03.016.
- [19] K.M. Horsman-Hall, Y. Orihuela, S.L. Karczynski, A.L. Davis, J.D. Ban, S. a Greenspoon, Development of STR profiles from firearms and fired cartridge cases., *Forensic Sci. Int. Genet.* 3 (2009) 242–50. doi:10.1016/j.fsigen.2009.02.007.
- [20] P. Dieltjes, R. Mieremet, S. Zuniga, T. Kraaijenbrink, J. Pijpe, P. de Knijff, A sensitive method to extract DNA from biological traces present on ammunition for the purpose of genetic profiling., *Int. J. Legal Med.* 125 (2011) 597–602. doi:10.1007/s00414-010-0454-4.
- [21] A. Miloš, A. Selmanović, L. Smajlović, R.L.M. Huel, C. Katzmarzyk, A. Rizvić, et al., Success rates of nuclear short tandem repeat typing from different skeletal elements, *Croat. Med. J.* 48 (2007) 486.
- [22] S. Amory, C. Keyser, E. Crubézy, B. Ludes, STR typing of ancient DNA extracted from hair shafts of Siberian mummies, *Forensic Sci. Int.* 166 (2007) 218–229.
- [23] J. Burger, S. Hummel, B. Hermann, DNA preservation: A microsatellite DNA based study on ancient skeletal remains, *Electrophoresis.* 20 (1999) 1728–1729.
- [24] I. Zupanič Pajnič, B. Gornjak Pogorelc, J. Balažic, T. Zupanc, B. Štefanič, Highly efficient nuclear DNA typing of the World War II skeletal remains using three new autosomal short tandem repeat amplification kits with the extended European Standard Set of loci, *Croat. Med. J.* 53 (2012) 17–23. doi:10.3325/cmj.2012.53.17.
- [25] C.S. Oh, S.J. Lee, J.B. Park, S.D. Lee, S.B. Seo, H.Y. Kim, et al., Autosomal short tandem repeat analysis of ancient DNA by coupled use of mini- and conventional STR kits., *J. Forensic Sci.* 57 (2012) 820–5. doi:10.1111/j.1556-4029.2011.02044.x.
- [26] M.M. Holland, C.A. Cave, C.A. Holland, T.W. Bille, Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center attacks, *Croat. Med. J.* 44 (2003) 264–272.
- [27] J.L. Prieto, C. Tortosa, A. Bedate, L. Segura, J.M. Abenza, M.C. Mariscal De Gante, et al., The 11 March 2004 Madrid terrorist attacks: The importance of the mortuary organisation for identification of victims. A critical review, *Int. J. Legal Med.* 121 (2007) 517–522.
- [28] Z.M. Zhao, Y. Liu, Y. Lin, Mutations of short tandem repeat loci in Identifiler system, *Fa Yi Xue Za Zhi.* 23 (2007) 290–291,294.
- [29] J.P. Du, P.Y. Chen, Y. Huang, J. Zhang, C.T. Li, Y.B. Li, Potential forensic application of closely linked autosomal STR haplotype in complex kinship testing, *Forensic Sci. Int. Genet. Suppl. Ser.* 4 (2013) e137–e139. doi:10.1016/j.fsigss.2013.10.070.
- [30] E. Ramírez-Flores, M. Saiz, D. Villegas-Carmona, M.J. Alvarez-Cubero, J.C. Alvarez, L. Vega-Navarrete, Genetic variation of 24 STR loci in a Mexican Mestizo population from Mexico DF, *Forensic Sci.* 10 (2014) e4–e6. doi:10.1016/j.fsigen.2014.02.001.
- [31] L. de Assis Poiaras, P. de Sá Osorio, F.A. Spanhol, S.C. Coltre, R. Rodenbusch, L. Gusmão, et al., Allele frequencies of 15 STRs in a representative sample of the Brazilian population, *Forensic Sci. Int. Genet.* 4 (2010) e61–e63.
- [32] I. Soltyszewski, W. Pepinski, P. Wolanska-Nowak, A. Maciejewska, R. Paszkowska, M. Abreu-Glowacka, et al., Polish population data on 15 autosomal STRs of

- AmpFLSTR NGM PCR kit., *Forensic Sci. Int. Genet.* 9 (2014) 142–9. doi:10.1016/j.fsigen.2013.12.002.
- [33] M. Sadam, G. Tasa, A. Tiidla, A. Lang, E.P. Axelsson, I.Z. Pajnič, Population data for 22 autosomal STR loci from Estonia., *Int. J. Legal Med.* (2014). doi:10.1007/s00414-014-1089-7.
- [34] S. Cardoso, M.A. Alfonso-Sánchez, L. Valverde, D. Sánchez, M.T. Zarrabeitia, A. Odriozola, et al., Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon, *Heredity* (Edinb). (2012).
- [35] N.N. Gazi, M. Eaaswarkhanth, A. Ferdous, S.S. Qadri, L. Singh, K. Thangaraj, Genetic polymorphisms of 15 autosomal STR loci in three isolated tribal populations of Bangladesh, *Forensic Sci. Int. Genet.* 4 (2010) 265–266. doi:DOI: 10.1016/j.fsigen.2009.08.003.
- [36] G. Martinez-Cortes, I. Nuno-Arana, R. Rubi-Castellanos, G. Vilchis-Dorantes, A. Luna-Vázquez, R.M. Coral-Vázquez, et al., Origin and genetic differentiation of three Native Mexican groups (Purépechas, Triquis and Mayas): contribution of CODIS-STRs to the history of human populations of Mesoamerica, *Ann. Hum. Biol.* 37 (2010) 801–819.
- [37] T. Kraaijenbrink, K.J. van der Gaag, S.B. Zuniga, Y. Xue, D.R. Carvalho-Silva, C. Tyler-Smith, et al., A linguistically informed autosomal STR survey of human populations residing in the greater Himalayan region., *PLoS One.* 9 (2014) e91534. doi:10.1371/journal.pone.0091534.
- [38] H. Sun, C. Zhou, X. Huang, S. Liu, K. Lin, L. Yu, et al., Correlation between the linguistic affinity and genetic diversity of Chinese ethnic groups., *J. Hum. Genet.* 58 (2013) 686–93. doi:10.1038/jhg.2013.79.
- [39] K.B. Schroeder, T.G. Schurr, J.C. Long, N.A. Rosenberg, M.H. Crawford, L.A. Tarskaia, et al., A private allele ubiquitous in the Americas., *Biol. Lett.* 3 (2007) 218–223. doi:10.1098/rsbl.2006.0609.
- [40] M.H. Crawford, K.G. Beaty, DNA fingerprinting in anthropological genetics: past, present, future., *Investig. Genet.* 4 (2013) 23. doi:10.1186/2041-2223-4-23.
- [41] L. Kerrigan, R.W. Nims, Authentication of human cell-based products: the role of a new consensus standard, *Regen. Med.* 6 (2011) 255–260. doi:10.2217/rme.11.5.
- [42] X. Zhao, S. Hu, L. Wang, Q. Zhang, X. Zhu, H. Zhao, et al., Functional short tandem repeat polymorphism of PTPN11 and susceptibility to hepatocellular carcinoma in Chinese populations., *PLoS One.* 9 (2014) e106841. doi:10.1371/journal.pone.0106841.
- [43] H.T. Orr, H.Y. Zoghbi, Trinucleotide repeat disorders., *Annu. Rev. Neurosci.* 30 (2007) 575–621. doi:10.1146/annurev.neuro.29.051605.113042.
- [44] C. Röthlein, M.S. Miettinen, T. Borwankar, J. Bürger, T. Mielke, M.U. Kumke, et al., Architecture of polyglutamine-containing fibrils from time-resolved fluorescence decay., *J. Biol. Chem.* 289 (2014) 26817–28. doi:10.1074/jbc.M114.581991.
- [45] T. Klockgether, The clinical diagnosis of autosomal dominant spinocerebellar ataxias., *Cerebellum.* 7 (2008) 101–5. doi:10.1007/s12311-008-0023-2.
- [46] H. Vauhkonen, M. Hedman, M. Vauhkonen, M. Kataja, P. Sipponen, A. Sajantila, Evaluation of gastrointestinal cancer tissues as a source of genetic information for forensic investigations by using STRs, *Forensic Sci. Int.* 139 (2004) 159–167. doi:10.1016/j.forsciint.2003.10.016.

- [47] Y. Liu, L. Li, C. Li, Z. Zhao, Allelic alterations of STRs in archival paraffin embedded tissue as DNA source for paternity testing, *Forensic Sci. Int. Genet. Suppl. Ser. 2* (2009) 12–13. doi:10.1016/j.fsigss.2009.08.019.
- [48] K. Mullis, F. Faloona, S. Scharf, Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction, *Cold Spring Harb. Symp. Quant. Biol.* 51 (1986) 263–273.
- [49] AmpFLSTR® Identifiler® PCR Amplification Kit. User's Manual Part Number 4323291 Rev. D 08/2006© Copyright 2006, Applied Biosystems. All rights reserved. Printed in the U.S.A., (n.d.).
- [50] A. Biosystems, AmpFLSTR Identifiler PCR Amplification Kit, User's Manual, (2006).
- [51] W. Bär, B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, P. Lincoln, et al., DNA recommendations. Further report of the DNA Commission of the ISFG regarding the use of short tandem repeat systems, *Forensic Sci Int.* 87 (1997) 179–184.
- [52] L. Excoffier, H. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [53] A. Tereba, Tools for Analysis of Population Statistics, Profiles DNA. (2001).
- [54] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies, *Genetics.* 164 (2003) 1567–1587.
- [55] M.J. Hubisz, D. Falush, M. Stephens, J.K. Pritchard, Inferring weak population structure with the assistance of sample group information, *Mol. Ecol. Resour.* 9 (2009) 1322–1332.
- [56] J. Felsenstein, PHYLIP (Phylogeny Inference Package), version 3.62 [computer program], Dep. Genome Sci. Univ. Washingt. (2004).
- [57] J.B. Kruskal, Nonmetric multidimensional scaling: A numerical method, *Psychometrika.* 29 (1964) 115–129.
- [58] R.D. Page, Visualizing phylogenetic trees using TreeView., *Curr. Protoc. Bioinforma.* / Ed. Board, Andreas D.Baxeavanis ...[et Al.]. Chapter 6 (2002).
- [59] M. Jakobsson, N.A. Rosenberg, CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure, *Bioinformatics.* 23 (2007) 1801–1806.
- [60] N.A. Rosenberg, DISTRUCT: a program for the graphical display of population structure, *Mol. Ecol. Notes.* 4 (2004) 137–138.
- [61] C. Coudray, E. Guitard, C. Keyser-Tracqui, M. Melhaoui, M. Cherkaoui, G. Larrouy, et al., Population genetic data of 15 tetrameric short tandem repeats (STRs) in Berbers from Morocco, *Forensic Sci. Int.* 167 (2007) 81–86.
- [62] L. Abdin, I. Shimada, B. Brinkmann, C. Hohoff, Analysis of 15 short tandem repeats reveals significant differences between the Arabian populations from Morocco and Syria, *Leg. Med.* 5 (2003) 150–155.
- [63] G.A. Omran, G.N. Ruty, M.A. Jobling, Genetic variation of 15 autosomal STR loci in Upper (Southern) Egyptians, *Forensic Sci. Int. Genet.* 3 (2009) e39–e44.
- [64] C. Brandt-Casadevall, M.B. Dhiab, F. Taroni, C. Gehrig, N. Dimo-Simonin, M. Zemni, et al., Tunisian population data on 15 PCR-based loci, *Forensic Sci. Int.* 126 (2002) 272–274.
- [65] L. Cherni, B. Loueslati Yaâcoubi, L. Pereira, C. Alves, H. Khodjet El Kill, A. Ben Ammar El Gaaied, et al., Data for 15 autosomal STR markers (Powerplex 16 System) from

- two Tunisian populations: Kesra (Berber) and Zriba (Arab), *Forensic Sci. Int.* 147 (2005) 101–106. doi:DOI: 10.1016/j.forsciint.2004.04.009.
- [66] B.S. Akbasak, B. Budowle, D.J. Reeder, J. Redman, M.C. Kline, Turkish population data with the CODIS multiplex short tandem repeat loci, *Forensic Sci. Int.* 123 (2001) 227–229.
- [67] A. Pérez-Lezaun, F. Calafell, J. Clarimón, E. Bosch, E. Mateu, L. Gusmão, et al., Allele frequencies of 13 short tandem repeats in population samples from the Iberian Peninsula and Northern Africa, *Int. J. Legal Med.* 113 (2000) 208–214.
- [68] M.F. Pinheiro, L. Cainé, L. Pontes, D. Abrantes, G. Lima, M.J. Pereira, et al., Allele frequencies of sixteen STRs in the population of Northern Portugal, *Forensic Sci. Int.* 148 (2005) 221–223.
- [69] V. Lopes, A. Serra, J. Gamero, L. Sampaio, F. Balsa, C. Oliveira, et al., Allelic frequency distribution of 17 STRs from Identifiler and PowerPlex-16 in Central Portugal area and the Azores archipelago, *Forensic Sci. Int. Genet.* 4 (2009) e1–e7. doi:DOI: 10.1016/j.fsigen.2008.12.001.
- [70] M. Paredes, M. Crespillo, J.A. Luque, J.L. Valverde, STR frequencies for the PowerPlex® 16 System Kit in a population from Northeast Spain, *Forensic Sci. Int.* 135 (2003) 75–78. doi:DOI: 10.1016/S0379-0738(03)00172-5.
- [71] F. Brisighelli, C. Capelli, I. Boschi, P. Garagnani, M. V Lareu, V.L. Pascali, et al., Allele frequencies of fifteen STRs in a representative sample of the Italian population, *Forensic Sci. Int. Genet.* 3 (2009) e29–e30.
- [72] P. Sánchez-Diz, P.G. Menounos, A. Carracedo, I. Skitsa, 16 STR data of a Greek population, *Forensic Sci. Int. Genet.* 2 (2008) e71–e72. doi:DOI: 10.1016/j.fsigen.2008.01.002.
- [73] S. Beleza, C. Alves, F. Reis, A. Amorim, A. Carracedo, L. Gusmão, 17 STR data (AmpF/STR Identifiler and Powerplex 16 System) from Cabinda (Angola), *Forensic Sci. Int.* 141 (2004) 193–196. doi:DOI: 10.1016/j.forsciint.2004.01.008.
- [74] C. Alves, L. Gusmão, A. Damasceno, B. Soares, A. Amorim, Contribution for an African autosomic STR database (AmpF/STR Identifiler and Powerplex 16 System) and a report on genotypic variations, *Forensic Sci. Int.* 139 (2004) 201–205.
- [75] M. Bouabdellah, F. Ouenzar, R. Aboukhalid, M. Elmzibri, D. Squalli, S. Amzazi, STR data for the 15 AmpFI STR Identifiler loci in the Moroccan population, *Prog. Forensic Genet.* 12 - Proc. 22nd Int. ISFG Congr. 1 (2008) 306–308. doi:DOI: 10.1016/j.fsigss.2007.10.123.
- [76] S. Elmrghni, R.A. Dixon, Y.M. Coulson-Thomas, D.R. Williams, Genetic data provided by 15 autosomal STR loci in the Libyan population living in Benghazi, *Forensic Sci. Int. Genet.* 6 (2012) e93–e94.
- [77] A.O. Tillmar, G. Bäckström, K. Montelius, Genetic variation of 15 autosomal STR loci in a Somali population, *Forensic Sci. Int. Genet.* 4 (2009) e19–e20. doi:DOI: 10.1016/j.fsigen.2009.01.004.
- [78] C.M. Schlebusch, H. Soodyall, M. Jakobsson, Genetic variation of 15 autosomal STR loci in various populations from southern Africa, *Forensic Sci. Int. Genet.* 6 (2012) e20–e21.
- [79] C. Coudray, R. Calderon, E. Guitard, B. Ambrosio, A. González-Martín, J.M. Dugoujon, Allele frequencies of 15 tetrameric short tandem repeats (STRs) in Andalusians from Huelva (Spain), *Forensic Sci. Int.* 168 (2007) e21–e24. doi:10.1016/j.forsciint.2007.01.016.

- [80] C. Capelli, V. Onofri, F. Brisighelli, I. Boschi, F. Scarnicci, M. Masullo, et al., Moors and Saracens in Europe: Estimating the medieval North African male legacy in southern Europe, *Eur. J. Hum. Genet.* 17 (2009) 848–852.
- [81] S. Plaza, F. Calafell, A. Helal, N. Bouzerna, G. Lefranc, J. Bertranpetit, et al., Joining the pillars of hercules: mtDNA sequences show multidirectional gene flow in the Western Mediterranean, *Ann. Hum. Genet.* 67 (2003) 312–328.
- [82] S.M. Adams, E. Bosch, P.L. Balaesque, S.J. Ballereau, A.C. Lee, E. Arroyo, et al., The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula., *Am. J. Hum. Genet.* 83 (2008) 725–36. doi:10.1016/j.ajhg.2008.11.007.
- [83] M. Brion, A. Salas, A. González-Neira, M. V Lareu, A. Carracedo, Insights into Iberian population origins through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite, *Am. J. Phys. Anthropol.* 122 (2003) 147–161. doi:10.1002/ajpa.10231.
- [84] J. Bertranpetit, L.L. Cavalli-Sforza, A genetic reconstruction of the history of the population of the Iberian Peninsula, *Ann. Hum. Genet.* 55 (1991) 51–67.
- [85] O. García, I. Yurrebaso, I. Uriarte, J.A. Pérez, R. Peñas, S. Alonso, et al., Distribution of Y-chromosomal haplotypes in the Basque Country autochthonous population using a 17-locus multiplex PCR assay, *Int. Congr. Ser.* 1288 (2006) 319–321.
- [86] E. González-Pérez, E. Esteban, M. Via, M. Gayà-Vidal, G. Athanasiadis, J.M. Dugoujon, et al., Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems), *Am. J. Phys. Anthropol.* 141 (2010) 430–9. doi:10.1002/ajpa.21161.

GENOTIPADO DE INDELs EN CROMOSOMAS AUTOSÓMICOS

AUTOSOMAL INDELs GENOTYPING

4

ABSTRACT

The genetic data of thirty eight autosomal INDELs in 143 unrelated donor residents of the provinces of Granada, Málaga and Almería (GMA) were analyzed in order to determine the genetic structure of this population and the usefulness of these markers in human identification. Observed allelic frequencies were also analyzed. Although mean values for forensic parameters do not reach values over 60%, combined discrimination power and combined exclusion power show high values, $1-3.5496 \times 10^{-15}$ y 99.630617%, respectively. The joint use of aSTR and INDELs may increase the matching probability in human identification, especially when working with compromised samples.

AMOVA and STRUCTURE analysis show that the populations of the three provinces can be treated genetically as a single population. However, the use of AIM INDELs might lead to the identification of possible minor substructure within a population.

1. INTRODUCCIÓN

Los polimorfismos de inserción/delección (INDELs) son polimorfismos de longitud creados por inserciones o deleciones de uno o más nucleótidos en el genoma. Los estudios de Weber *et al.* 2002 [1] y Mills *et al.* 2006 [2] contribuyeron de forma decisiva en el conocimiento sobre este tipo de polimorfismos así como su utilidad en estudios genéticos por su abundancia en el genoma y su fácil análisis.

En 2002, Weber *et al.* identificaron y caracterizaron unos 2000 INDELs con diferentes variaciones de longitud (desde 2pb a 10kb) y calcularon sus frecuencias alélicas en las poblaciones europea, africana, japonesa y nativa americana [1]. Posteriormente, en 2006, se generó un mapa de variación INDEL en el genoma comprendiendo 415 000 polimorfismos [2]. Actualmente, a través de estudios de secuenciación a escala poblacional se han identificado 1,6 millones de INDEL de 1 a 10 000pb [3]. Sin embargo, el mayor grupo de INDELs son aquellos con un pequeño número de nucleótidos [2]; disminuyendo en frecuencia a medida que aumenta el tamaño del INDEL [4].

Estos son los polimorfismos genéticos más abundantes después de los SNPs; el ratio de INDELs/SNPs se ha calculado como 0,19 [3,5], representando aproximadamente un 16% de toda la variabilidad humana y se encuentran presentes en todo el genoma humano, aproximadamente un INDEL cada 7,2kb [2].

El deslizamiento de la polimerasa es el causante del 75% de los INDELs, mientras que el resto se deben a simples deleciones de secuencia completa, movilización de elementos transponibles, errores durante la reparación de las hebras de ADN, entrecruzamientos desiguales... [3,6]. En el estudio del origen genético de los INDEL, se consideró la recombinación como un posible mecanismo de mutagénesis ya que el entrecruzamiento introduce roturas en ambas hebras. Se ha descrito cómo los SNPs cercanos a los puntos calientes de recombinación se encuentran enriquecidos. Sin embargo, no se han descrito fenómenos de enriquecimiento de INDELs en sitios calientes de recombinación. La falta de estas señales implica que la recombinación no tiene un alto efecto mutagénico en INDELs [3].

La tasa de mutación de INDELs de pequeño tamaño depende en gran medida de la presencia de STRs y puede aumentar drásticamente a medida que incrementa el número de unidades de repetición. Se ha observado que la tasa de mutación de los INDELs varía a lo largo del genoma, yendo desde 4×10^{-5} hasta 0,3 [3]. Los INDELs

de gran tamaño son mucho menos frecuentes que los INDELS de tamaño corto pero contribuyen en mayor medida a la evolución del tamaño del genoma [6].

Se ha determinado que los INDELS tienen una menor presencia en secuencias codificantes comparado con otras categorías génicas, un ratio 0,046 INDELS/SNPs (Figura 4.1) [3,5].

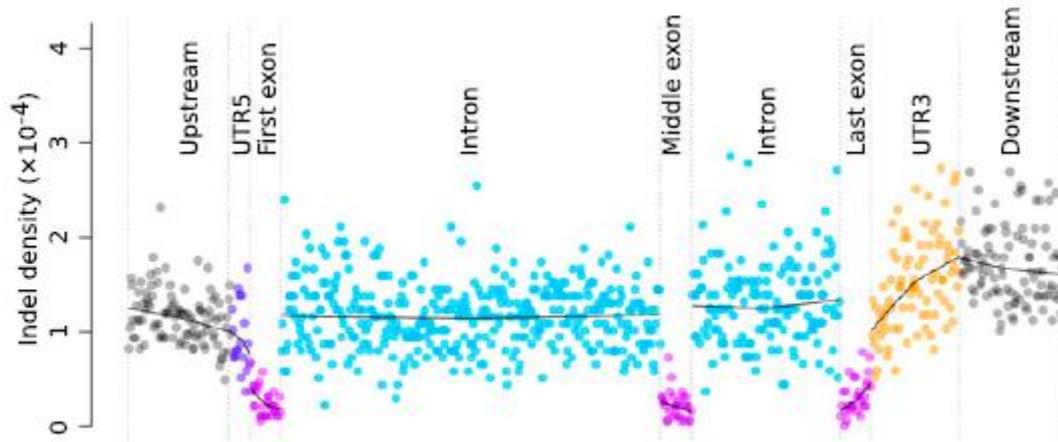


Figura 4.1: Densidad de INDELS en seis regiones génicas. Imagen tomada de Montgomery et al. 2013 [3]

Esto se debe a que los INDELS generan mayores cambios en exones mientras que los SNPs generalmente producen cambios sinónimos con poco impacto en la función del gen. Por ello, se da una fuerte selección purificadora hacia la eliminación de los INDELS localizados en regiones codificantes; tanto porque afectan directamente a la secuencia nucleotídica de un elemento funcional como porque la longitud de la región afectada por el INDEL es funcionalmente importante [6].

Estudios recientes en INDELS localizados en regiones génicas reportan que, en general; las deleciones se segregan en menor frecuencia que las inserciones concluyendo que las deleciones sufren un mayor nivel de selección purificadora en regiones génicas [7]. A pesar de estos procesos de selección purificadora, siguen estando presentes INDELS en regiones codificantes que tienen un impacto sustancial en los humanos [5].

1.1. Origen y Evolución de los Indels

En un primer intento de caracterizar los INDELS presentes en el genoma humano, se realizaron estudios de evolución de éstos y determinación del estado ancestral de los polimorfismos. Los resultados arrojaron que de los 2000 INDELS estudiados [1], prácticamente todos se generaron de forma posterior a la

divergencia del ancestro común entre humanos, chimpancés y gorilas ya que en muy pocos casos se comparten los polimorfismos de longitud entre humanos y simios. Los alelos monomórficos encontrados en chimpancés y gorilas seguramente representan al estado ancestral de estas secuencias [1].

Tras el tipado de los 2000 INDELS en el genoma de los simios se distinguió entre deleciones e inserciones en función de qué alelos coincidía exactamente en longitud con el observado en el chimpancé. Se determinó un ratio deleción:inserción de 4:1 [1]. Estudios realizados a posteriori con aproximadamente 10 000 INDELS revelan que el ratio deleción:inserción es de 2,4:1 para INDELS polimórficos humanos [7]. Otros estudios utilizando la especie *Macaca mulatta* como grupo de referencia, han determinado que el 63% de la variación INDEL entre humanos y chimpancés se debe a inserciones. Asimismo, la mayoría de estas inserciones se debieron a fenómenos de retrotransposición ocurridos en la línea humana tras la divergencia de ambas especies de su ancestro común [8].

Por otro lado, el estudio de los INDELS que compartían tanto humanos como gorilas y chimpancés; determinó que el alelo presente en gorilas era el mismo que el presente en chimpancés. Estos datos favorecen a la hipótesis que los polimorfismos presentes en simios representan el alelo ancestral de los INDELS [1].

A continuación se estudiaron un grupo de INDELS de los cuales o el chimpancé o el gorila presentaban alelos de longitud diferente a los observados en humanos. De los estudiados, en un 43% de los marcadores sólo una de las dos especies tenía alelos que coincidían con los humanos. La presencia de estos alelos se puede explicar mediante una mutación generada en el ancestro común entre gorilas y chimpancés tras haberse separado la línea evolutiva humana. Por otro lado, un 46% de estos polimorfismos tenían la misma longitud tanto en gorilas como en chimpancés pero no coincidían con la longitud de los alelos presentes en humanos. Estos resultados quedan explicados por dos mutaciones separadas de los INDELS en la línea ancestral humana tras la divergencia con el ancestro común. La primera mutación se fijó en la línea humana, y la segunda dio lugar a los polimorfismos encontrados en la actualidad. Por último, un 11% de los polimorfismos presenta alelos de longitudes diferentes entre gorilas y chimpancés, así como con los humanos. Estos casos pueden ser explicados por la aparición de mutaciones en ambas líneas tras la divergencia de éstas [1].

El hecho de que prácticamente todos los INDELS humanos son aparentemente monomórficos en chimpancés y gorilas concuerda con los resultados encontrados previamente en SNPs [1].

1.2. Aplicaciones

1.2.1. INDELS como marcadores genéticos forenses.

Como marcador genético, los INDELS combinan características tanto de SNPs como de STRs. Por un lado, comparten una variación binaria similar a los SNPs: fragmentos de amplificación cortos y tasas de mutación más bajas que los STRs y generalmente derivan de un evento mutacional único no recurrente en el tiempo de evolución de los humanos modernos. Además son abundantes y se distribuyen por todo el genoma [9]. Estas características han hecho que se despierte un interés en las aplicaciones forenses de los INDELS. La más relevante es la posibilidad de analizar amplicones mucho más cortos en comparación con los STRs de uso estándar (Figura 4.2).

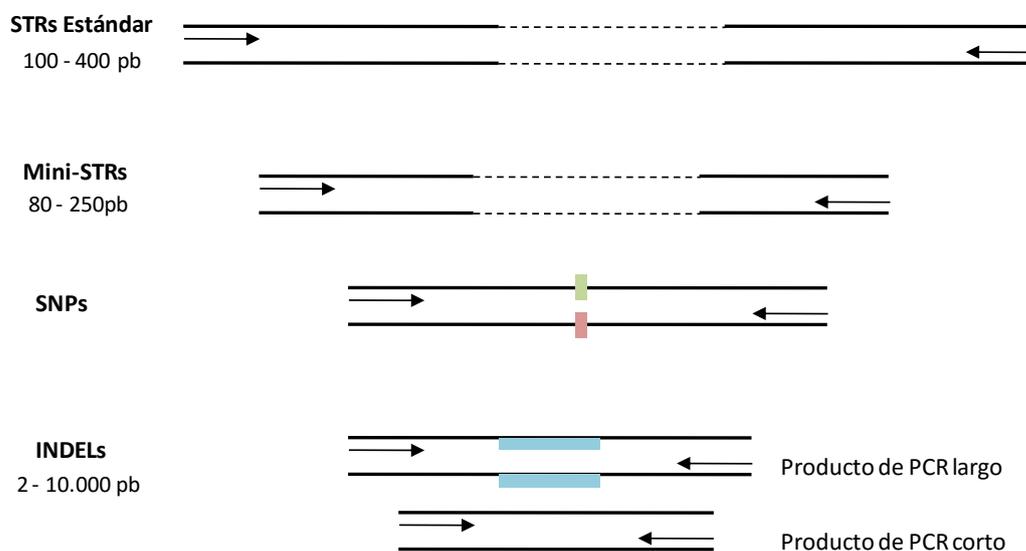


Figura 4.2: Análisis de INDELS en amplicones cortos. Al igual que los SNPs, se pueden analizar en amplicones mucho más cortos que los STRs de uso estándar; incluso que los mini-STRs recientemente diseñados.

El uso de miniINDELS, con muy pequeña variación de longitud entre alelos (2-5pb); y el hecho de que estos marcadores son estudiados mediante técnicas de PCR con primers marcados con fluorescencia [10] frente a los protocolos de tipado de

SNPs mucho más laboriosos y costosos; los convierte en marcadores idóneos para el trabajo con muestras degradadas (Figura 4.3).

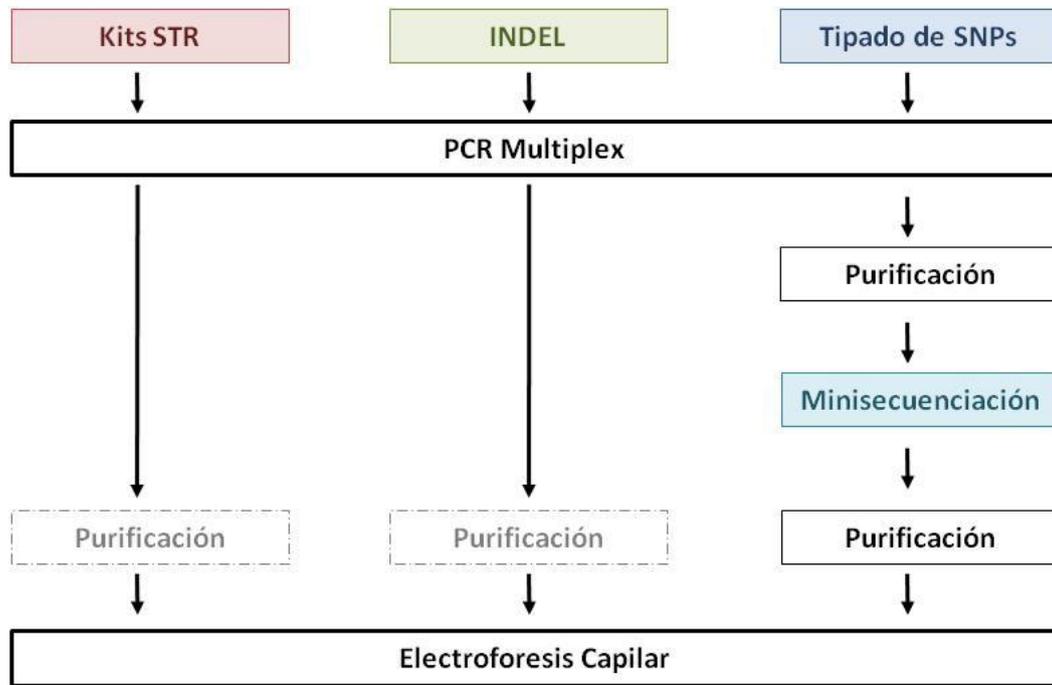


Figura 4.3: Comparativa del flujo de trabajo con los diferentes marcadores forenses.

Además, no sólo se han descubierto INDELS en cromosomas autosómicos sino que también se han descrito a lo largo del cromosoma X [11,12], en las regiones flanqueantes de STRs de cromosoma X [13,14] y en el cromosoma Y [15].

Por último, se ha determinado la utilidad de los INDEL en la identificación de muestras tumorales. Estudios en tejido gastrointestinal tumoral han detectado cuatro tipos diferentes de fenómenos de mutación en STRs (alelo nuevo, alelo adicional, pérdida de la heterocigosidad y pérdida parcial de la heterocigosidad). Sin embargo, sólo se detectan dos tipos de mutaciones en INDELS (pérdida de la heterocigosidad y pérdida parcial de la heterocigosidad) [16]. Mientras que la pérdida de la heterocigosidad y pérdida parcial de la heterocigosidad no afecta al genotipo del individuo, las otras dos generan grandes problemas en la identificación del individuo del que procede la muestra tumoral. Se ha determinado que la tasa de mutación de INDELS en tejidos tumorales es de 0,25% frente al 5,25% que se observa en STRs y se ha demostrado que los fenómenos mutagénicos en ambos marcadores no están correlacionados. Estos resultados indican que los INDELS tienen mayor poder de identificación que los STRs en muestras de tejido tumoral gastrointestinal [16].

1.2.2. INDELS como herramienta complementaria en estudios de filiación.

En estudios de paternidad, así como en otros estudios de establecimiento de relaciones biológicas; se pueden observar incompatibilidades en la transmisión mendeliana de los STRs. Esto hace que sea necesario diferenciar los fenómenos de mutación frente a una exclusión. Sin embargo, tras estudiar estas dudas, las conclusiones no pueden ser del todo claras. Debido a la alta tasa de mutación de los STRs es fácil encontrarse con este tipo de situaciones. En estos casos, la incorporación de un amplio número de INDEL o SNPs que presenten bajas tasas de mutación ayuda a resolver estos problemas. La caracterización de un mayor número de marcadores permite desvelar posibles incompatibilidades de transmisión, favoreciendo la exclusión de la relación de parentesco, o en caso de inclusión, el incremento de los parámetros estadísticos de paternidad [17]. Se ha determinado que serían necesarios unos 100 INDELS para alcanzar un porcentaje de exclusión de 99,8% [18] pero se ha de tener cuidado cuando no se detectan incompatibilidades ya que, en ocasiones la falsa hipótesis de paternidad puede ser reforzada por los valores de LR [18].

1.2.3. Estudio de la estructura de las poblaciones

Como ya se ha comentado anteriormente; los INDELS derivan de un único evento mutacional que se da en muy bajas frecuencias. Debido a esto son genéticamente muy estables y presentan una distribución de sus frecuencias alélicas muy diferente entre poblaciones distantes; por lo que se pueden considerar como AIMs [1,19,20]. Se han desarrollado paneles de INDELS capaces de identificar estratificación en poblaciones mestizas y estimar proporciones ancestrales tanto individuales como globales [20–22]. Además, se han desarrollado recientemente un gran número de estudios donde se usan los INDELS para determinar la estructura genética de las poblaciones así como su uso en poblaciones actuales [23–25].

1.2.4. INDELS en el desarrollo de enfermedades

Se han identificado INDELS que residen en regiones codificantes causando cambios en el marco de lectura o superposición de los límites de los exones; que afectan únicamente al último exón y a todos los transcritos descritos en *Ensembl*. Este tipo de INDELS generan consecuencias que afectan al fenotipo.

En ocasiones esta alteración génica da lugar a enfermedades no letales, de aparición temprana y asociadas a fenotipos que pueden pasar fácilmente desapercibidos. Este es el caso del gen *MYO6*, pérdida progresiva de la audición [26,27] o del *IRAK3*, aparición temprana de asma [28].

Una de las enfermedades genéticas más comunes en humanos, la fibrosis quística, está causada por un INDEL en la región codificante del gen *CFTR* (Cystic fibrosis transmembrane conductance regulator) que causa la eliminación de un aminoácido [29].

Por otro lado, hasta la fecha se han descrito un total de 52 genes implicados en cáncer en humanos que tienen INDELs en regiones codificantes [5].

Se ha descrito un INDEL en la región promotora del gen *BRM* (Brahma homólogo) [30], así como en la región promotora del gen *PDCD6IP* (proteína de interacción de muerte celular programada 6) [31]; en las que los genotipos ins/del e ins/ins confieren un mayor riesgo de carcinoma hepatocelular en la población china. Sin embargo, estos dos mismos genotipos en la región no traducida 3' del gen *SGSM3* confieren una disminución de un 45% del riesgo de padecer este mismo tipo de cáncer [32].

2. MATERIAL Y MÉTODOS

2.1. Población de Estudio

Se han estudiado un total de 143 individuos adultos no relacionados entre sí de la región del sureste de la Península Ibérica [Almería (45), Granada (53), Málaga (45)], ver capítulo 2 páginas 39-40. Todos los donantes fueron informados del propósito de este estudio, aceptaron su participación y firmaron un consentimiento informado según la guía ética de la declaración de Helsinki.

2.2. Análisis Genético Molecular de la Muestra

La toma de muestras así como su posterior análisis están estrictamente registrados por rigurosos controles, tanto intralaboratorio como interlaboratorio.

2.2.1. Extracción

Se tomaron muestras del epitelio bucal de todos los individuos mediante hisopos estériles de algodón. El ADN genómico se obtuvo mediante un protocolo de extracción orgánica con fenol:cloroformo:alcohol isoamílico (25:24:1) (Apéndice I).

2.2.2. Cuantificación

La estimación de la cantidad y calidad de ADN nuclear de las muestras se ha realizado a través de electroforesis en geles de agarosa al 0,8% (Apéndice I).

2.2.3. Amplificación

Las muestras se amplificaron mediante una multiplex de 38 INDELS de acuerdo con Pereira *et al.* 2012 (Tabla 4.1) [17]. Los fluorocromos empleados en el marcaje de los primers forward son 6-FAM™ (azul), VIC® (verde), NED™ (amarillo) y PET® (rojo). La estrategia de amplificación comprende diferentes aspectos de interés forense:

- Screening de estructuras secundarias entre primers para una coamplificación eficiente.

- Distribución balanceada a lo largo de todo el genoma, cubriendo todos los autosomas (Figura 4.4).
- Localización cromosómica evitando desequilibrio de ligamiento entre marcadores situados en el mismo cromosoma.
- Etiquetado con múltiples marcadores y distribución en función del tamaño esperado del amplicón para poder analizarlos en una única electroforesis.

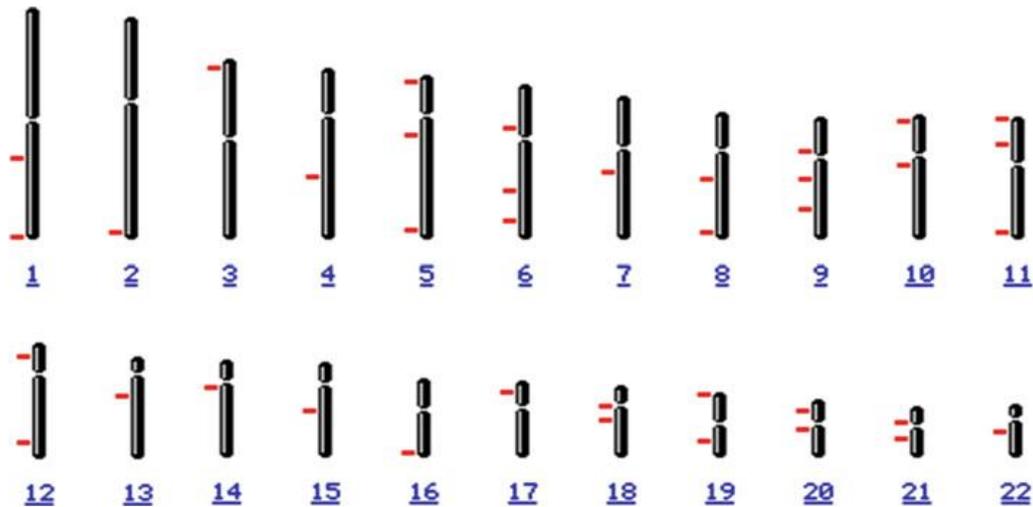


Figura 4.4: Distribución de los 38 INDEL estudiados a lo largo de los 22 autosomas humanos. Imagen tomada de Pereira et al. 2012 [17].

Todos los INDEL empleados son autosómicos y bialélicos; el alelo minoritario presenta una frecuencia superior a 0,25 en poblaciones europeas, africanas y asiáticas; tienen una heterocigosidad media de 0,40; se localizan en regiones no codificantes; y tienen un tamaño alélico entre 2 y 5 pb (Tabla 4.2).

Tabla 4.1: Componentes, composición y volumen de la reacción de amplificación.

Reactivo	Composición	Volumen
Qiagen Multiplex PCR kit (2x)	MgCl ₂ , dNTPs, PCR Buffer, HotStarTaq DNA polimerasa	5µl
Primer Mix (10x)	0,1µl de cada primer excepto rs2308137 (0,2 µl) y rs3047269 (0,3 µl)	1µl
ADN	0,3-5ng/µl	1µl
Agua estéril		3µl

La amplificación de todos los loci se realizó en dos termocicladores GeneAmp® PCR Systems 2720 de Applied Biosystems® (Figura 4.5).

Tabla 4.2: Características de los 38 INDELS estudiados.

Cód.	Interno	MID	rs	Posición (pb)			Alelos	Tamaño amplicón (S-L)
				Crom.	Build 36.3	Build 37.1		
B1	MID-2719	rs34541393	20	30165066	30701405	-/AACT	57-61	
B2	MID-185	rs16624	2	234681130	235016391	-/GT	65-67	
B3	MID-1493	rs2307689	19	48896180	44204340	-/TTC	74-77	
B4	MID-2946	rs35769550	8	76681235	76518680	-/TGAC	89-93	
B5	MID-1504	rs2307700	22	25120901	26790901	-/TCAC	101-105	
B6	MID-520	rs140809	10	6027167	5987163	-/CAA	115-118	
B7	MID-2305	rs3047269	1	161077452	162810828	-/CTGA	126-130	
B8	MID-3221	rs33972805	11	125794082	126288872	-/CT	135-137	
B9	MID-2698	rs33917182	20	11643625	11695625	-/CA	142-144	
B10	MID-116	rs16402	9	38396788	38406788	-/TTAT	150-154	
G1	MID-785	rs1610871	5	171020572	171087970	-/TAGG	61-65	
G2	MID-1151	rs2067238	12	113772931	115288548	-/GCT	71-74	
G3	MID-1209	rs2067294	9	70504241	71314421	-/CTT	80-83	
G4	MID-1514	rs2307710	6	47929222	47821263	-/AGGA	92-96	
G5	MID-2050	rs2308242	3	8591709	8616709	-/CT	106-108	
G6	MID-1384	rs2307580	9	104626014	105586193	-/AATT	120-124	
G7	MID-743	rs1160956	5	65414216	65378460	-/AGA	128-131	
G8	MID-3097	rs34511541	18	34677042	36423040	-/CTCTT	143-148	
G9	MID-1782	rs2307978	7	83121850	83283913	-/GA	156-158	
Y1	MID-2648	rs3051300	17	10076666	10135941	-/GTAT	63-67	
Y2	MID-2890	rs10629077	21	30294208	31372337	-/AT	74-76	
Y3	MID-3277	rs10688868	11	258180	268180	-/CT	81-83	
Y4	MID-1120	rs2067208	16	83139788	84582287	-/GCCAG	93-98	
Y5	MID-1383	rs2307579	1	245878706	247812083	-/ATG	104-107	
Y6	MID-1824	rs2308020	15	51268809	53481517	-/TT	127-129	
Y7	MID-3114	rs3080855	18	21507205	23253207	-/AATT	133-137	
Y8	MID-834	rs1610919	12	14801263	14909996	-/AT	142-144	
Y9	MID-1643	rs2307839	6	117200251	117093558	-/GA	152-154	
R1	MID-1945	rs2308137	6	149655891	149614198	-/GA	61-63	
R2	MID-2592	rs36040336	19	1353662	1402662	-/AT	65-67	
R3	MID-649	rs1160886	10	54112392	54442386	-/ACT	75-78	
R4	MID-1830	rs2308026	4	119404855	119185407	-/CA	83-85	
R5	MID-1330	rs2307526	5	5178112	5125112	-/ACAC	95-99	
R6	MID-3220	rs34811743	11	30134266	30177690	-/TG	108-110	
R7	MID-1997	rs2308189	14	28106508	29036757	-/AACTA	119-124	
R8	MID-3031	rs5895447	8	138489776	138420594	-/CA	128-130	
R9	MID-1979	rs2308171	13	43778155	44880155	-/TCTG	135-139	
R10	MID-2806	rs35605984	21	14556736	15634865	-/TAAAG	151-156	

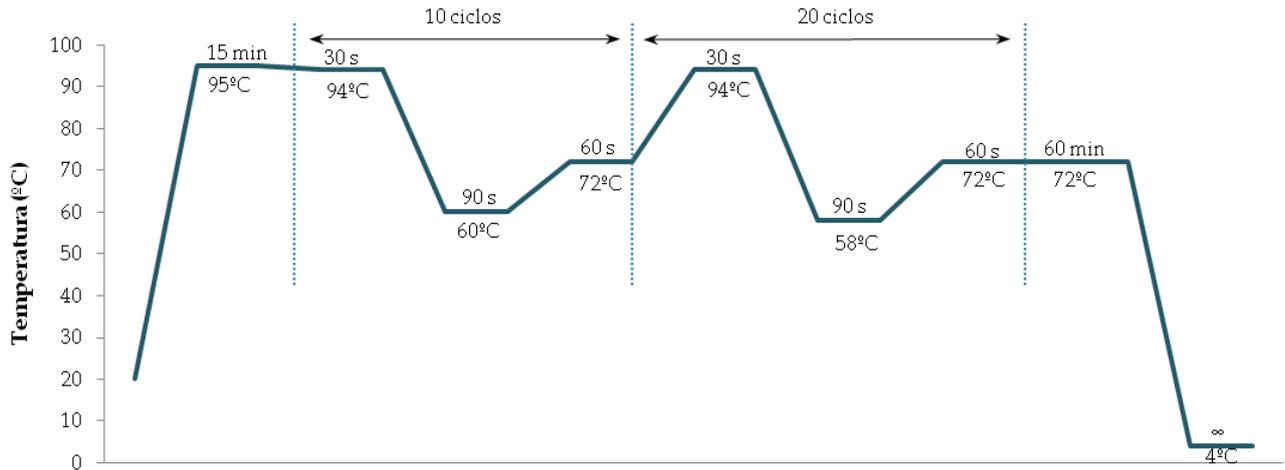


Figura 4.5: Gráfico de las condiciones de amplificación.

En todas las reacciones se han amplificado sistemáticamente dos controles:

- Control positivo: verificar que la reacción de amplificación ha sido correcta.
- Control negativo: descartar una posible contaminación por ADN extraño.

2.2.4. Visualización de los resultados

Las muestras amplificadas fueron visualizadas por electroforesis capilar en un analizador genético ABI Prism 3130 de Applied Biosystems®.

Para la correcta detección de los fragmentos de ADN nuclear amplificados, se han de preparar las muestras antes de separarlas por electroforesis capilar (Apéndice II).

Para preparar las muestras, se añadió en cada pocillo de una placa de 96 pocillos los siguientes volúmenes por muestra:

- 10 µl de formamida desionizada
- 0,3 µl del estándar de tamaño (500pb marcado en naranja- LIZ)
- 1 µl de producto amplificado

La electroforesis se llevó a cabo con Performance Optimized Polymer 7 (POP-7™; Life Technologies) con un filtro G5.

Las condiciones de electroforesis fueron las siguientes:

- Módulo.....FragmentAnalysis36_POP7
- Tiempo de inyección.....16 s
- Voltaje de inyección..... 1,2kV
- Voltaje de electroforesis..... 15kV
- Temperatura de electroforesis.....60°C
- Tiempo de electroforesis20min

Los electroferogramas fueron visualizados mediante GeneMapper ID-X 1.1; que permite determinar el tamaño de los fragmentos analizados. Se emplearon los bin set para INDELs facilitados por Pereira *et al.* 2012 (Figura 4.6) [17].

La calidad del tipado de las muestras así como de la designación de alelos fueron garantizadas por el análisis simultáneo de una muestra control de genotipo conocido.

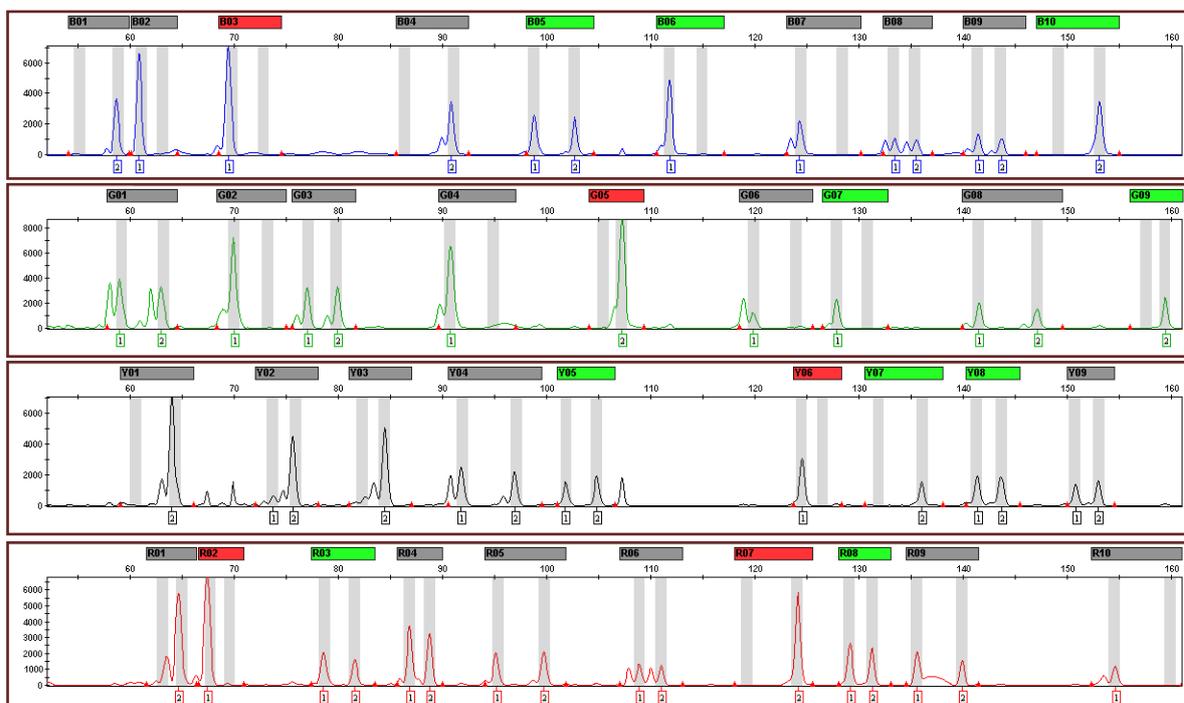


Figura 4.6. Perfil genético obtenido para los 38 INDELs analizados de una de las muestras de este estudio.

2.3. Análisis Estadístico de la Muestra

El cálculo de las frecuencias de las repeticiones encontradas en cada una de las poblaciones se realizó por distintos métodos. En primer lugar, la obtención de las frecuencias alélicas muestrales se determinó por el método clásico del recuento en base a los alelos observados (Li, 1976). En segundo lugar se recalculó este mismo parámetro con dos programas estadísticos, *Arlequin 3.5.1.3* [33] y *PowerStatsv12* (Promega® Corporation) [34]. En todos los casos se han de obtener los mismos resultados.

La heterocigosidad observada y esperada se calcularon mediante estos dos últimos programas *PowerStatsv12* y *Arlequin 3.5.1.3*. Este último también sirvió como herramienta para calcular el equilibrio de Hardy-Wienberg con 1 000 000 pasos en la cadena de Markov y 100 000 pasos de dememorización, así como el

desequilibrio de ligamiento entre loci. Además, el desequilibrio de ligamiento entre loci también fue calculado con *SNPAnalyzer 2* [35].

Todos los parámetros forenses más relevantes se calcularon mediante la hoja de cálculo *PowerStatsv12* obteniendo el poder de discriminación (PD), el poder de exclusión (PE), el índice de contenido polimórfico (PIC), la probabilidad de coincidencia (MP) y el índice típico de paternidad (TPI).

Para determinar la subestructura genética de la población del Sudeste de la Península Ibérica, se testó la hipótesis de una distribución aleatoria de los individuos entre parejas de poblaciones mediante un test exacto de diferenciación de la población y el análisis de la varianza molecular (AMOVA) con 10 000 repeticiones experimentales mediante *Arlequin 3.5.1.3*. Además, se utilizó *STRUCTURE 2.3.4* [36,37] para inferir grupos de individuos a partir de los datos genéticos de loci no ligados. Para todas las simulaciones y cálculos, se asumieron los modelos de mestizaje y no mestizaje con un periodo de *burn-in* de 50 000 iteraciones seguido de 100 000 iteraciones adicionales, estimando de 1 a 4 contribuciones (K). Cada carrera se repitió 10 veces y se calcularon las probabilidades para cada K en cada grupo de carreras. Las 10 réplicas para cada K se evaluaron con *CLUMPP 1.1.2* (<http://www.stanford.edu/group/rosenberglab/clumpp.html>) [38]. Los grupos combinados resultantes se visualizaron mediante *DISTRUCT 1.1* (<http://www.stanford.edu/group/rosenberglab/distruct.html>) [39].

3. RESULTADOS Y DISCUSIÓN

3.1. Estudio Poblacional

Los resultados obtenidos del estudio de frecuencias alélicas de 38 INDELS quedan resumidos en la tabla 4.3.

Tabla 4.3: Parámetros forenses relevantes para 38 INDELS en la población de Granada, Málaga y Almería.

Número rs	Cod.	Alelo corto	Alelo largo	Ho	He	PIC	PD	PE	MP	TPI	p
rs34541393	B01	0,4056	0,5944	0,5175	0,4864	0,3659	0,5980	0,2033	0,4020	1,0362	0,4939
rs16624	B02	0,7448	0,2552	0,3706	0,3600	0,3079	0,5429	0,1048	0,4571	0,8125	0,8172
rs2307689	B03	0,3427	0,6573	0,4476	0,4564	0,3490	0,5976	0,1456	0,4024	0,9051	0,8590
rs35769550	B04	0,3671	0,6329	0,4196	0,4644	0,3567	0,6183	0,1309	0,3817	0,8720	0,2817
rs2307700	B05	0,5769	0,4231	0,4126	0,4919	0,3690	0,6418	0,1309	0,3582	0,8720	0,0573
rs140809	B06	0,2483	0,7517	0,3427	0,3746	0,3036	0,5398	0,0826	0,4602	0,7606	0,3690
rs3047269	B07	0,4476	0,5524	0,5385	0,4955	0,3722	0,5937	0,2305	0,4063	1,1000	0,3189
rs33972805	B08	0,4790	0,5210	0,5385	0,5009	0,3746	0,6027	0,2234	0,3973	1,0833	0,4063
rs33917182	B09	0,6014	0,3986	0,4476	0,4839	0,3645	0,6265	0,1456	0,3735	0,9051	0,3884
rs16402	B10	0,2587	0,7413	0,3636	0,3849	0,3100	0,5489	0,0933	0,4511	0,7857	0,5136
rs1610871	G01	0,5350	0,4650	0,5385	0,4993	0,3738	0,6011	0,2234	0,3989	1,0833	0,4015
rs2067238	G02	0,6434	0,3566	0,4615	0,4605	0,3536	0,6009	0,1560	0,3991	0,9286	1,0000
rs2067294	G03	0,2972	0,7028	0,4196	0,4220	0,3305	0,5750	0,1218	0,4250	0,8512	1,0000
rs2307710	G04	0,3881	0,6119	0,5245	0,4766	0,3622	0,5868	0,2098	0,4132	1,0515	0,2855
rs2308242	G05	0,2203	0,7797	0,2727	0,3447	0,2845	0,5069	0,0582	0,4931	0,7010	0,0277
rs2307580	G06	0,4930	0,5070	0,4685	0,5017	0,3750	0,6392	0,1614	0,3608	0,9408	0,5109
rs1160956	G07	0,8252	0,1748	0,3217	0,2895	0,2469	0,4550	0,0728	0,5450	0,7371	0,2508
rs34511541	G08	0,4545	0,5455	0,4615	0,4976	0,3729	0,6379	0,1560	0,3621	0,9286	0,4054
rs2307978	G09	0,2273	0,7727	0,3427	0,3525	0,2896	0,5178	0,0826	0,4822	0,7606	0,8128
rs3051300	Y01	0,3636	0,6364	0,4126	0,4663	0,3557	0,6183	0,1263	0,3817	0,8614	0,2018
rs10629077	Y02	0,2483	0,7517	0,3566	0,3815	0,3036	0,5391	0,0896	0,4609	0,7772	0,5114
rs10688868	Y03	0,2797	0,7203	0,3217	0,4044	0,3218	0,5694	0,0728	0,4306	0,7371	0,0228
rs2067208	Y04	0,2902	0,7098	0,3566	0,4134	0,3271	0,5778	0,0896	0,4222	0,7772	0,1068
rs2307579	Y05	0,5175	0,4825	0,4755	0,5011	0,3747	0,6295	0,1785	0,3705	0,9795	0,6275
rs2308020	Y06	0,6538	0,3462	0,4546	0,4543	0,3502	0,5973	0,1507	0,4027	0,9167	1,0000
rs3080855	Y07	0,3077	0,6923	0,4266	0,4248	0,3353	0,5776	0,1357	0,4224	0,8827	1,0000
rs1610919	Y08	0,5280	0,4720	0,5175	0,4998	0,3742	0,6180	0,1968	0,3820	1,0214	0,7381
rs2307839	Y09	0,2413	0,7587	0,3636	0,3710	0,2991	0,5307	0,0970	0,4693	0,7944	0,8248
rs2308137	R01	0,2832	0,7168	0,4615	0,4163	0,3236	0,5506	0,1507	0,4494	0,9167	0,2219
rs36040336	R02	0,8077	0,1923	0,2448	0,3117	0,2624	0,4655	0,0432	0,5345	0,6620	0,0147
rs1160886	R03	0,4266	0,5734	0,4476	0,4909	0,3696	0,6363	0,1456	0,3637	0,9051	0,3073
rs2308026	R04	0,4091	0,5909	0,4406	0,4852	0,3666	0,6329	0,1406	0,3671	0,8938	0,3066
rs2307526	R05	0,3147	0,6853	0,3706	0,4302	0,3383	0,5950	0,1009	0,4050	0,8034	0,1187

rs34811743	R06	0,6259	0,3741	0,4825	0,4700	0,3587	0,6016	0,1727	0,3984	0,9662	0,8576
rs2308189	R07	0,3741	0,6259	0,4965	0,4700	0,3587	0,5950	0,1845	0,4050	0,9931	0,5915
rs5895447	R08	0,3392	0,6608	0,4546	0,4498	0,3478	0,5929	0,1507	0,4071	0,9167	1,0000
rs2308171	R09	0,2448	0,7552	0,3846	0,3746	0,3014	0,5334	0,1009	0,4666	0,8034	0,8224
rs35605984	R10	0,3042	0,6958	0,3566	0,4248	0,3337	0,5879	0,0970	0,4121	0,7944	0,0776

Todos los loci estudiados se encuentran en equilibrio Hardy-Weinberg para esta población ($p > 0,05$), excepto 3 (rs36040336, rs10688868 y rs2308242). Tras aplicar la corrección de Bonferroni, todos los loci se encuentran dentro del equilibrio HWE ($p > 0,00132$).

La media de los parámetros forenses calculados es poder de discriminación (PD) = 58,11%; poder de exclusión (PE) = 13,57%, índice de contenido polimórfico (PIC) = 33,85% y heterocigosidad observada (H_o) = 42,19%. Los marcadores rs3047269, rs33972805 y rs1610871 muestran un mayor grado de heterocigosidad (53,85%) mientras que el marcador rs36040336 muestra el menor nivel (24,48%). El poder de discriminación y exclusión combinado es $1-3,5496 \times 10^{-15}$ y 99,630617%, respectivamente. El estudio del desequilibrio de ligamiento no muestra ninguna asociación significativa entre los 38 INDEL estudiados por lo que se pueden tratar como marcadores independientes.

Estudios realizados de este mismo set de marcadores en la población europea, africana y del este de Asia revelan un 99,99999% de PD combinado, niveles informativos satisfactorios para los requerimientos forenses [40]. Además, este mismo estudio indica los 25 INDELs más informativos para cada población serían suficientes para obtener perfiles genéticos únicos. Este hecho resalta el potencial de este grupo de marcadores en estudios de identificación, incluso en aquellos en los que no se obtienen perfiles completos [40].

3.2. Estudio de la Estructura de la Población

Para determinar si el estudio de los INDEL permite establecer una subestructura de la población de estudio en las tres provincias de las que proceden las muestras, se realizaron estudios de AMOVA y STRUCTURE.

El análisis AMOVA no muestra una subestructura genética significativa entre los individuos de las tres provincias de estudio, por lo que podemos agruparlos dentro de la misma población ($p > 0,05$). Sólo un 0,10% de la variación observada corresponde a la encontrada entre las poblaciones (Tabla 4.4). Cabe destacar que el 96,65% de la variación determinada a través del estudio de los INDELs es la que se observa entre los individuos de una misma población. Estos resultados

confirman a los INDELS como unos buenos marcadores forenses capaces de diferenciar individuos de una misma población entre sí de una forma muy eficaz.

Tabla 4.4: Diseño AMOVA y resultados de las tres provincias del Sudeste de la Península Ibérica.

Diseño	Fuente de Variación	d.f	Suma de cuadrados	Componentes de Varianza	Porcentaje de variación
3 Grupos: Grupo 1= Almería Grupo 2= Málaga Grupo 3= Granada	Entre poblaciones	2	18,646	0,00806 Va	0,10
	Entre individuos dentro de las poblaciones	140	1197,997	0,26982 Vb	3,25
	Dentro de las poblaciones	143	1146,500	8,01748 Vc	96,65
	Total	285	2363,143	8,29536	
					Índices de fijación
					FIS: 0,03256
					FST: 0,00097
					FIT: 0,03350

Los resultados obtenidos por el análisis de AMOVA son corroborados con los que se derivan del estudio de distancias genéticas apareadas así como por el test exacto de diferenciación de la población (Tabla 4.5).

Tabla 4.5: Comparación de las distancias genéticas interpopulacionales (coeficientes Fst) entre las poblaciones de Granada, Málaga y Almería.

	Almería	Málaga	Granada
Almería	0	0,37012 ± 0,0140	0,17090 ± 0,0115
Málaga	0,00093	0	0,45117 ± 0,0122
Granada	0,00239	0,00057	0

El triángulo inferior representa las distancias genéticas apareadas; el triángulo superior representa los p-valores de FST.

El análisis con STRUCTURE revela que no hay evidencias de la presencia de una estructura genética significativa en la población del Sureste de la Península Ibérica en relación a los grupos generados por la división geográfica de las muestras en función de las provincias de procedencia. El modelo con mayor probabilidad asociada es el modelo K=1 (ln P(D)=-6822,76). Esto nos indica que los datos genéticos favorecen un único grupo que engloba a todo el sureste de la Península. Como se puede ver en la figura 4.7; todos los individuos de las distintas provincias

se agrupan en el centro de la figura y se distribuyen equitativamente sin distinción de su origen.

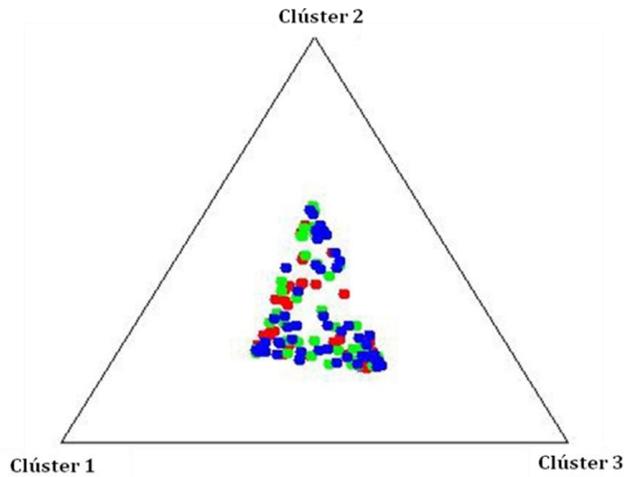


Figura 4.7: Gráfico triangular generado con el programa STRUCTURE para la agrupación por homología genética de los individuos de las tres provincias a las que pertenecen las muestras de este estudio. Los tres componentes de ancestralidad (1, 2 y 3) se representan en cada vértice del triángulo y cada una de las provincias de las que proceden las muestras se representa por un color.

Estudios de comparación genética realizados con otras poblaciones muestran como el uso de cierto grupo de INDELS no permite diferenciar los componentes ancestrales entre poblaciones dentro de un mismo continente u origen genético, aún cuanto se tratan de poblaciones genéticamente muy conservadas como es el caso del País Vasco. Sin embargo, sí que permiten la diferenciación genética entre poblaciones de diferentes continentes y origen [41]. No obstante, el estudio de un panel de INDELS de ancestralidad sí permite la identificación de una subestructura marcada presente entre tres poblaciones mezcladas entre ellas así como la subestructura leve encontrada dentro de una misma población [42].

En general, los INDELS presentan un tamaño pequeño de amplicón, ausencia de *stutter bands* y son fácilmente implementados en laboratorios forenses, por lo que son de gran interés para suplementar los resultados obtenidos con el estudio de STRs autosómicos en el trabajo forense [43,44], sobre todo en aquellos casos en los que se trabaja con muestras altamente degradadas. Diferentes estudios revelan que el índice de paternidad alcanza valores de 5 000 000 cuando se combina el uso de miniSTRs e INDELS en estudios de paternidad con muestras postmortem [43]. Además, en establecimiento de la paternidad, la presencia de incompatibilidades mendelianas entre padre e hijo es un problema. Debido a la alta tasa de mutación de los STR autosómicos no es una situación poco frecuente que se puede solventar con el uso de un gran número de marcadores con baja tasa de mutación como los SNPs e INDELS [40].

4. BIBLIOGRAFÍA

- [1] J.L. Weber, D. David, J. Heil, Y. Fan, C. Zhao, G. Marth, Human diallelic insertion/deletion polymorphisms, *Am. J. Hum. Genet.* 71 (2002) 854–862.
- [2] R.E. Mills, C.T. Luttig, C.E. Larkins, A. Beauchamp, C. Tsui, W.S. Pittard, et al., An initial map of insertion and deletion (INDEL) variation in the human genome, (2006) 1182–1190. doi:10.1101/gr.4565806.1.
- [3] S.B. Montgomery, D.L. Goode, E. Kvikstad, C.A. Albers, Z.D. Zhang, X.J. Mu, et al., The origin, evolution, and functional impact of short insertion–deletion variants identified in 179 human genomes, *Genome Res.* 23 (2013) 749–761. doi:10.1101/gr.148718.112.
- [4] T.R. Bhangale, M.J. Rieder, R.J. Livingston, D. a. Nickerson, Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes, *Hum. Mol. Genet.* 14 (2005) 59–69. doi:10.1093/hmg/ddi006.
- [5] R.E. Mills, W.S. Pittard, J.M. Mullaney, U. Farooq, T.H. Creasy, A. a Mahurkar, et al., Natural genetic variation caused by small insertions and deletions in the human genome., *Genome Res.* 21 (2011) 830–9. doi:10.1101/gr.115907.110.
- [6] E.M. Kvikstad, L. Duret, Strong heterogeneity in mutation rate causes misleading hallmarks of natural selection on indel mutations in the human genome., *Mol. Biol. Evol.* 31 (2014) 23–36. doi:10.1093/molbev/mst185.
- [7] P. Sjödin, T. Bataillon, M.H. Schierup, Insertion and deletion processes in recent human history, *PLoS One.* 5 (2010). doi:10.1371/journal.pone.0008650.
- [8] N. Polavarapu, G. Arora, V.K. Mittal, J.F. McDonald, Characterization and potential functional significance of human-chimpanzee large INDEL variation., *Mob. DNA.* 2 (2011) 13. doi:10.1186/1759-8753-2-13.
- [9] R. Pereira, L. Gusmão, Capillary electrophoresis of 38 noncoding biallelic mini-indels for degraded samples and as complementary tool in paternity testing, *Methods Mol. Biol.* 830 (2012) 141–157.
- [10] J. Huang, H. Luo, W. Wei, Y. Hou, A novel method for the analysis of 20 multi-Indel polymorphisms and its forensic application., *Electrophoresis.* 35 (2014) 487–93. doi:10.1002/elps.201300346.
- [11] J. Edelmann, S. Hering, C. Augustin, R. Szibor, Indel polymorphisms—An additional set of markers on the X-chromosome, *Forensic Sci. Int. Genet. Suppl. Ser. 2* (2009) 510–512. doi:10.1016/j.fsigss.2009.08.148.
- [12] K. Sun, S.M. Zhao, S.H. Zhang, C.T. Li, Development of an 18 X-InDel multiplex PCR system, *J. Forensic Med.* 30 (2014) 101–109. doi:10.3969/j.issn.1004-5619.2014.02.006.
- [13] A. Nagai, M. Hara, T. Ishihara, A. Tamura, A. Kido, Y. Bunai, INDEL polymorphisms at the DXS10146 flanking region in four racial populations, *Forensic Sci. Int. Genet. Suppl. Ser. 4* (2013) e318–e319. doi:10.1016/j.fsigss.2013.10.162.
- [14] A. Nagai, Y. Bunai, Insertion/deletion polymorphisms at the X-STR DXS10146 and DXS10147 flanking regions, *Forensic Sci. Int. Genet. Suppl. Ser. 3* (2011) e129–e130. doi:10.1016/j.fsigss.2011.08.064.

- [15] J. Damas, A. Amorim, L. Gusmão, InDels in Y chromosome haplogroup definition, *Forensic Sci. Int. Genet. Suppl. Ser.* 3 (2011) e178–e179. doi:10.1016/j.fsigss.2011.08.089.
- [16] S. Zhao, S. Zhang, T. Que, Z. Zhao, C. Li, Application of insertion/deletion polymorphisms in human gastrointestinal tumour tissues for identification purpose, *Forensic Sci. Int. Genet. Suppl. Ser.* 3 (2011) e297–e298. doi:10.1016/j.fsigss.2011.09.011.
- [17] R. Pereira, L. Gusmão, Capillary Electrophoresis of 38 Noncoding Biallelic Mini-Indels for Degraded Samples and as Complementary, in: A. Alonso (Ed.), *DNA Electrophor. Protoc. Forensic Genet., Methods in Molecular Biology*, 2012: pp. 141–157. doi:10.1007/978-1-61779-461-2.
- [18] N. Pinto, M. Magalhães, E. Conde-Sousa, C. Gomes, R. Pereira, C. Alves, et al., Assessing paternities with inconclusive STR results: The suitability of bi-allelic markers., *Forensic Sci. Int. Genet.* 7 (2013) 16–21. doi:10.1016/j.fsigen.2012.05.002.
- [19] N. Yang, H. Li, L.A. Criswell, P.K. Gregersen, M.E. Alarcon-Riquelme, R. Kittles, et al., Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: Application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine, *Hum. Genet.* 118 (2005) 382–392. doi:10.1007/s00439-005-0012-1.
- [20] R. Pereira, C. Phillips, N. Pinto, C. Santos, S.E.B. dos Santos, A. Amorim, et al., Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing., *PLoS One.* 7 (2012) e29684. doi:10.1371/journal.pone.0029684.
- [21] N.P.C. Santos, E.M. Ribeiro-Rodrigues, A.K.C. Ribeiro-Dos-Santos, R. Pereira, L. Gusmão, A. Amorim, et al., Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel., *Hum. Mutat.* 31 (2010) 184–90. doi:10.1002/humu.21159.
- [22] L. Bastos-Rodrigues, J.R. Pimenta, S.D.J. Pena, The genetic structure of human populations studied through short insertion-deletion polymorphisms., *Ann. Hum. Genet.* 70 (2006) 658–65. doi:10.1111/j.1469-1809.2006.00287.x.
- [23] E.H. Kim, H.Y. Lee, I.S. Yang, W.I. Yang, K.-J. Shin, Population data for 30 insertion-deletion markers in a Korean population., *Int. J. Legal Med.* 128 (2014) 51–2. doi:10.1007/s00414-013-0851-6.
- [24] S.R.R. Torres, C.J.S. Uehara, A.F. Sutter-Latorre, B.S. de Almeida, T.S. Sauerbier, Y.C.N. Muniz, et al., Population genetic data and forensic parameters of 30 autosomal InDel markers in Santa Catarina State population, Southern Brazil., *Mol. Biol. Rep.* 41 (2014) 5429–33. doi:10.1007/s11033-014-3414-0.
- [25] A. Zidkova, A. Horinek, V. Kebrdlova, M. Korabecna, Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population, *Int. J. Legal Med.* 127 (2013) 7–10. doi:10.1007/s00414-011-0649-3.
- [26] K.M. Sanggaard, K.W. Kjaer, H. Eiberg, G. Nürnberg, P. Nürnberg, K. Hoffman, et al., A novel nonsense mutation in MYO6 is associated with progressive nonsyndromic hearing loss in a Danish DFNA22 family, *Am. J. Med. Genet. Part A.* 146 (2008) 1017–1025. doi:10.1002/ajmg.a.32174.

- [27] N. Hilgert, V. Topsakal, J. van Dinther, E. Offeciens, P. Van de Heyning, G. Van Camp, A splice-site mutation and overexpression of MYO6 cause a similar phenotype in two families with autosomal dominant hearing loss., *Eur. J. Hum. Genet.* 16 (2008) 593–602. doi:10.1038/sj.ejhg.5202000.
- [28] L. Balaci, M.C. Spada, N. Olla, G. Sole, L. Loddo, F. Anedda, et al., IRAK-M is involved in the pathogenesis of early-onset persistent asthma., *Am. J. Hum. Genet.* 80 (2007) 1103–1114. doi:10.1086/518259.
- [29] F.S. Collins, M.L. Drumm, J.L. Cole, W.K. Lockwood, G.F. Vande Woude, M.C. Iannuzzi, Construction of a general human chromosome jumping library, with application to cystic fibrosis., *Science.* 235 (1987) 1046–1049. doi:10.1126/science.2950591.
- [30] X. Gao, M. Huang, L. Liu, Y. He, Q. Yu, H. Zhao, et al., Insertion/deletion polymorphisms in the promoter region of BRM contribute to risk of hepatocellular carcinoma in Chinese populations., *PLoS One.* 8 (2013) e55169. doi:10.1371/journal.pone.0055169.
- [31] Q. Yu, C. Zhou, J. Wang, L. Chen, S. Zheng, J. Zhang, A functional insertion/deletion polymorphism in the promoter of PDCD6IP is associated with the susceptibility of hepatocellular carcinoma in a Chinese population., *DNA Cell Biol.* 32 (2013) 451–7. doi:10.1089/dna.2013.2061.
- [32] C. Wang, H. Zhao, X. Zhao, J. Wan, D. Wang, W. Bi, et al., Association between an insertion/deletion polymorphism within 3'UTR of SGSM3 and risk of hepatocellular carcinoma., *Tumour Biol.* 35 (2014) 295–301. doi:10.1007/s13277-013-1039-x.
- [33] L. Excoffier, H. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows , *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [34] A. Tereba, Tools for Analysis of Population Statistics, Profiles DNA. (2001).
- [35] J. Yoo, Y. Lee, Y. Kim, S.Y. Rha, SNPAnalyzer 2.0: A web-based integrated workbench for linkage disequilibrium analysis and association analysis, *BMC Bioinformatics.* 9 (2008) 290.
- [36] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies, *Genetics.* 164 (2003) 1567–1587.
- [37] M.J. Hubisz, D. Falush, M. Stephens, J.K. Pritchard, Inferring weak population structure with the assistance of sample group information, *Mol. Ecol. Resour.* 9 (2009) 1322–1332.
- [38] M. Jakobsson, N.A. Rosenberg, CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure, *Bioinformatics.* 23 (2007) 1801–1806.
- [39] N.A. Rosenberg, DISTRUCT: a program for the graphical display of population structure, *Mol. Ecol. Notes.* 4 (2004) 137–138.
- [40] R. Pereira, C. Phillips, C. Alves, A.A. Amorim, A. Carracedo, L. Gusmão, et al., A new multiplex for human identification using insertion/deletion polymorphisms, *Electrophoresis.* 30 (2009) 3682–3690. doi:10.1002/elps.200900274.
- [41] P. Martín, O. García, B. Heinrichs, I. Yurrebaso, A. Aguirre, A. Alonso, Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations., *Forensic Sci. Int. Genet.* 7 (2013) e27–30. doi:10.1016/j.fsigen.2012.10.003.

- [42] P.A.D.C. Francez, E.M. Ribeiro-Rodrigues, S.E.B. dos Santos, Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon., *Forensic Sci. Int. Genet.* 6 (2012) 132–5. doi:10.1016/j.fsigen.2011.04.002.
- [43] F. Manta, A. Caiafa, R. Pereira, D. Silva, A. Amorim, E.F. Carvalho, et al., Indel markers: Genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material, *Forensic Sci. Int. Genet.* 6 (2012) 1–4. doi:10.1016/j.fsigen.2011.12.008.
- [44] C.V. da Silva, S. Matos, H.A. Costa, P. Morais, R.M. Dos Santos, R. Espinheira, et al., Genetic portrait of south Portugal population with InDel markers., *Forensic Sci. Int. Genet.* 7 (2013) e101–3. doi:10.1016/j.fsigen.2013.03.009.

ABSTRACT

The genetic data of seventeen Y chromosome short tandem repeats in 146 unrelated donor residents of the provinces of Granada, Málaga and Almería (GMA) were analyzed in order to determine the genetic legacy left by the male inhabitants of the former Kingdom of Granada. A total of 139 unique haplotypes have been described. Observed-allelic frequencies and haplogroup frequencies were also analyzed. AMOVA and STRUCTURE analysis show that the populations of the three provinces can be treated genetically as a single population. The most frequent haplogroup described is R1b1b2 (58.22%). Network analyses based in all individual reveals a net distribution according to haplogroup assignment.

For better characterization, the population GMA was compared with North African, Iberian Peninsula and South European populations. Analysis of allelic frequencies and genetic distances show that GMA population is located within the Spanish populations group. Furthermore, STRUCTURE analyses demonstrate that none African component is present in GMA population. Ancestral components are distributed over the populations according to haplogroup designation instead of geographical location. The most important components in GMA population are those that identify European haplogroups. The analytical results affirm that, owing to the studied genetic markers, the GMA population does not show any characteristics that reflect male genetic influence of North African people. The presence of African haplogroups in GMA population is not relevant when their frequency is compared to those in other European populations.

1. INTRODUCCIÓN

Los cromosomas sexuales tienen morfologías muy diferentes entre ellos. Sin embargo, inicialmente fueron un par de cromosomas homólogos. Los procesos de divergencia iniciados al adquirir uno de ellos funciones de determinación del sexo en fases iniciales de la evolución de los mamíferos vinieron seguidos de una represión de la recombinación. Esto ha llevado a una expansión de la región no recombinante confinando las regiones pseudoautosómicas a los extremos [1].

El cromosoma Y humano es el segundo cromosoma más pequeño con una longitud aproximada de 60 Mb. Se trata de un cromosoma específico masculino, haploide y con una herencia de padres a hijos [2].

Debido a su carácter haploide, no posee cromosoma homólogo, por lo que su recombinación se encuentra reducida a ambos extremos del cromosoma Y, llamados *regiones pseudoautosómicas* o *PAR*. Éstas son las únicas regiones donde se produce recombinación con las regiones homólogas del cromosoma X. PAR1 se localiza en el extremo del brazo corto del cromosoma y tiene una longitud de aproximadamente 2,7 Mb (Figura 5.1) [3]. La recombinación en PAR1 es obligatoria en la meiosis masculina reflejando el origen ancestral de los cromosomas sexuales como un par de cromosomas homólogos. Debido a su tamaño, PAR1 tiene una tasa de recombinación 17 veces mayor a la media de todo el genoma [4,5]. La delección total de PAR1 se asocia con infertilidad total masculina y la recombinación reducida puede dar lugar a espermatozoides aneuploides resultando en una monosomía del cromosoma X (síndrome de Turner) o en XXY (síndrome de Klinefelter) [4]. PAR1 humano presenta homólogos en PARs de otras especies de mamíferos, caballo, perro u oveja [4].

PAR2 se localiza en el extremo del brazo largo del cromosoma Y teniendo un tamaño muchísimo menor, unos 330 Kb (Figura 5.1). Esta región es específica de humanos habiéndose generado por una translocación entre los cromosomas X e Y. PAR2 no presenta prácticamente recombinación.

Se han descrito 29 genes en la región PAR; 24 en PAR1 y 5 en PAR2. De todos, el único asociado a enfermedad es *Short Stature Homeobox* (SHOX), que causa corta estatura y mal desarrollo óseo [6].

El resto del cromosoma ($\approx 95\%$) se conoce como la *región no recombinante del cromosoma Y* o *NR*Y (Figura 5.1). Esta región se mantiene idéntica de padres a hijos siempre y cuando no se den fenómenos de mutación. Se han descrito 78 genes funcionales en esta región [7]. Hay un grupo de 9 genes que tienen un homólogo en

el cromosoma X [8]. Entre ellos encontramos el gen DBY que codifica para una ARN helicasa dependiente de ATP implicada en procesos celulares como iniciación de la traducción o splicing nuclear y mitocondrial, entre otros. Frecuentemente, este gen se encuentra deletado en varones infértiles y su ausencia completa deriva en una disminución significativa de células germinales o incluso una pérdida total de éstas [9]. Por otro lado, se han descrito 11 genes específicos de tejido testicular que presentan localizaciones múltiples [8]. El gen SRY se encuentra dentro de este grupo. Este gen tiene un papel crucial en el desarrollo gonadal de los embriones XY hacia la diferenciación masculina en el desarrollo temprano del embrión [9].

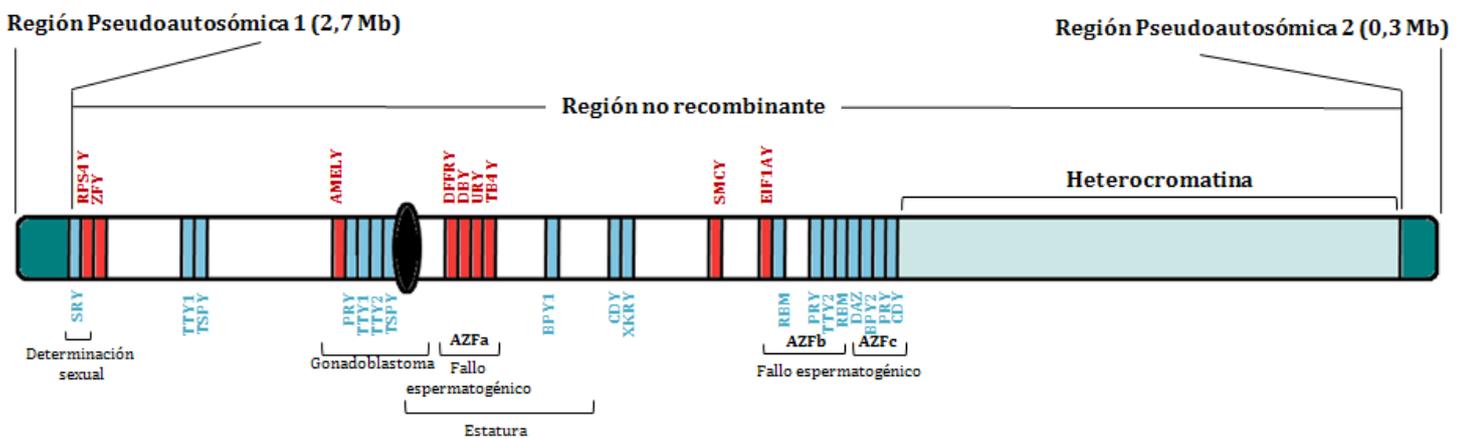


Figura 5.1: Mapa genético del cromosoma Y. En rojo, genes que encuentran su homólogo en el cromosoma X. En azul, genes específicos de tejido testicular. Se indica el fenotipo derivado de fallos en los diferentes genes.

1.1. Origen y Evolución del Cromosoma Y

El cromosoma Y evolucionó de un autósoma y está caracterizado por un declive masivo de genes. Los cromosomas X e Y humanos se originaron hace unos 200 – 300 millones de años en mamíferos placentarios [10] tras la separación de los monotremas y evolucionaron independientemente en aves y reptiles [7]. Tal y como pasa en el cromosoma Y humano, en múltiples especies éste ha sufrido un proceso de degeneración donde pierde la gran mayoría de sus genes originales a lo largo del tiempo evolutivo.

El primer paso en la evolución del cromosoma Y se dio con la adquisición de un gen determinante del sexo en uno de los miembros de una pareja de autosomas. La presencia de este tipo de genes ya se observa en anfibios, peces y múltiples invertebrados. Posteriormente, tuvo lugar una inhibición de la recombinación entre ambos cromosomas a través de mutaciones sexualmente antagonistas, que dio lugar a la generación de cromosomas heteromórficos [11]. Existen pruebas en

genoma mamífero que señalan las inversiones cromosómicas como fuente de la inhibición de la recombinación cromosómica. Estudios sobre los genes localizados en la región degenerada del X en el cromosoma Y, revelan que la inhibición se dio en múltiples pasos [12]. El hecho que el cromosoma Y se transmita exclusivamente de padres a hijos hace que este cromosoma sea una parte ideal del genoma para albergar los genes que potencian la aptitud masculina. Es por ello que la gran mayoría de genes ancestrales se han perdido y los pocos que se localizan en éste tengan funciones específicas masculinas [13].

Estudios de genética comparativa muestran como *Rhesus macaca* tienen prácticamente el mismo cromosoma Y que los humanos anatómicamente modernos. Eso implica que el contenido génico del cromosoma Y se ha mantenido estable desde hace 30 millones de años, momento en el que *Rhesus macaca* se separó de la línea evolutiva [7].

1.2. Marcadores Genéticos en el Cromosoma Y

La naturaleza haploide del cromosoma Y le confiere características especiales como marcador genético. Tiene un tamaño poblacional efectivo mucho menor que los autosomas que da lugar a una baja diversidad genética. A causa del pequeño tamaño poblacional que presenta, los polimorfismos de cromosoma Y pueden estar fuertemente afectados por fenómenos de deriva genética o eventos a nivel poblacional como cuellos de botella o efecto fundador. Además, la dispersión asimétrica de los distintos polimorfismos se ve influida por la transmisión patrilineal de este cromosoma presentada por diversas culturas como la patrilocalización o la poliginia. Estos hechos explican, en parte; la información geográfica derivada del estudio de algunos marcadores de cromosoma Y [14].

Por otro lado, la falta de recombinación en la región eucromática del cromosoma Y hace que los marcadores genéticos no se puedan considerar como independientes y se hayan de construir haplotipos para caracterizar individuos [15].

A finales de la década de los 80, se empezaron a usar secuencias de ADN específicas del cromosoma Y para la determinación del sexo de restos biológicos [16,17].

1.2.1. STRs de cromosoma Y

La realización, en la década de los 80, de estudios sobre la existencia de regiones repetitivas en el cromosoma Y con variabilidad similar a la de los STRs

orientadas en sentido opuesto flanqueadas por zona iguales, por lo que el producto obtenido es el de la amplificación de ambas zonas independientemente, pudiendo ser igual o diferente. Se cree que esta duplicación tuvo lugar anterior a la emergencia del hombre anatómicamente moderno ya que se ha observado en todas las poblaciones humanas estudiadas [24,25].

Se han descrito más de 500 STRs en el cromosoma Y pero no todos son específicos de varones o lo suficiente polimórficos como para ser usados con fines forenses o antropológicos (Figura 5.3) [26,27]. Diversos kits comerciales incorporan los 17 loci más usados. Todos tienen unas características comunes: la mayoría son tetranucleótidos con lo que se minimiza la formación de artefactos *stutter*; todos tienen una tasa de mutación baja, entre 10^{-3} y 10^{-4} asegurando que las mutaciones entre padre e hijo son infrecuentes y presentan muy alta resolución haplotípica entre las poblaciones mundiales [28]. Sin embargo, no todas las poblaciones muestran un alto grado de resolución con los 17 marcadores empleados de rutina. Además, presentan otra gran limitación; no se pueden excluir a los varones familiares del sospechoso de un caso forense, se ha visto que familiares separados hasta por 20 generaciones presentan el mismo perfil para los 17 Y-STRs. Es por ello que recientemente se han incorporado nuevos Y-STRs a los kits comerciales así como se han desarrollado reacciones multiplex con estos nuevos marcadores (Tabla 5.1).

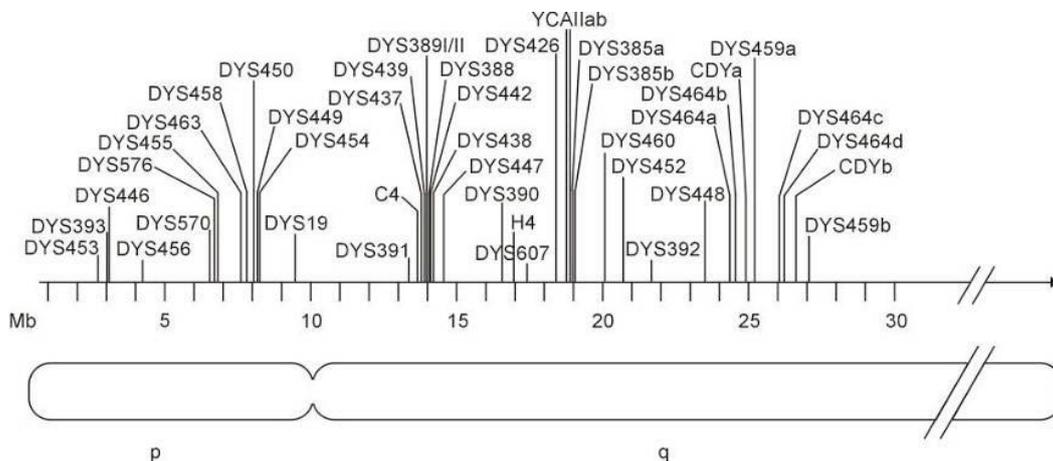


Figura 5.3: Localización cromosómica aproximada de los STR más empleados en cromosoma Y. Imagen tomada de Hammer et al. 2006 [155].

Se ha desarrollado una nueva reacción multiplex que incluye 13 Y-STRs altamente mutables (RM) en la que todos los marcadores incluidos poseen una tasa de mutación por encima de 10^{-2} , dando una diferenciación esperada padre-hijo de 19,5% [29,30]. Estudios en poblaciones humanas han demostrado que

estos 13 marcadores poseen una alta diversidad haplotípica, incrementando 0,6% los resultados obtenidos con los 17 Y-STRs empleados de rutina [30]. Sin embargo, la alta tasa de mutación de estos nuevos marcadores los convierte en malos marcadores para el estudio de relaciones familiares. Por ello, se han descrito un grupo de marcadores con tasas de mutación por debajo de 10^{-2} , que evitarían la necesidad de contemplar estadísticamente fenómenos de mutación [31]. El principal problema que entrañan estos marcadores es su baja diversidad genética, por lo que se han de emplear un mayor número de marcadores con baja tasa de mutación para establecer relaciones familiares padre/hijo [28].

Tabla 5.1: Características de los diferentes loci del cromosoma Y empleados [28,32].

<i>Locus</i>	<i>Categoría</i>	<i>Motivo de repetición</i>	<i>Rango alélico</i>	<i>Tasa de mutación</i>
DYS19	Copia única, complejo	[TAGA] ₃ TAGG[TAGA] ₆₋₁₆	9-20	$2,24 \times 10^{-3}$
DYS389I	Copia única, complejo	[TCTG] ₃ [TCTA] ₆₋₁₄	9-17	$2,93 \times 10^{-3}$
DYS389II	Copia única, complejo	[TCTG] ₄₋₅ [TCTA] ₁₀₋₁₄ N ₂₈ [TCTG] ₃ [TCTA] ₆₋₁₄	24-36	$4,12 \times 10^{-3}$
DYS390	Copia única, complejo	[TCTG] ₈ [TCTA] ₉₋₁₄ [TCTG] ₁ [TCTA] ₄	17-30	$2,11 \times 10^{-3}$
DYS391	Copia única, complejo	TCTA] ₃ [TCTA] ₆₋₁₅	5-16	$2,45 \times 10^{-3}$
DYS392	Copia única, simple	[TAT] ₄₋₂₀	5-20	$5,19 \times 10^{-4}$
DYS393	Copia única, simple	[AGAT] ₇₋₁₈	7-18	$1,05 \times 10^{-3}$
DYS385a/b	Copias múltiples (2), complejo	(AAGG) ₄ N ₁₄ (AAAG) ₃ N ₁₂ (AAAG) ₃ N ₂₉ (AAGG) ₆₋₇ [GAAA] ₇₋₂₃	6-28	$2,45 \times 10^{-3}$
DYS438	Copia única, simple	[TTTTTC] ₇₋₁₆	7-18	$3,75 \times 10^{-4}$
DYS439	Copia única, complejo	[GATA] ₃ N ₃₂ [GATA] ₅₋₁₉	5-19	$5,45 \times 10^{-3}$
DYS437	Copia única, complejo	[TCTA] ₄₋₁₂ [TCTG] ₂ [TCTA] ₄	10-18	$1,22 \times 10^{-3}$
DYS448	Copia única, complejo	[AGAGAT] ₁₁₋₁₃ N ₄₂ [AGAGAT] ₈₋₉	12.2-25	$1,52 \times 10^{-3}$
DYS456	Copia única, simple	[AGAT] ₁₁₋₂₃	5-23	$4,29 \times 10^{-3}$
DYS458	Copia única, simple	[GAAA] ₁₁₋₂₄	11-24	$6,36 \times 10^{-3}$
DYS635	Copia única, complejo	[TCTA] ₄ (TGTA) ₂ [TCTA] ₂ (TGTA) ₂ [TCTA] ₂ (TGTA) ₀₋₂ [TCTA] ₄₋₁₇	13-30	$4,33 \times 10^{-3}$
Y GATA H4	Copia única, complejo	[TAGA] ₃ N ₁₂ (TAGG) ₃ [TAGA] ₈₋₁₅ N ₂₂ [TAGA]	6-18	$3,03 \times 10^{-3}$
DYS481	Copia única, simple	(CTT) ₂₂₋₃₂	16-34	$6,71 \times 10^{-3}$
DYS533	Copia única, simple	(TATC) ₉₋₁₄	7-16.3	$5,01 \times 10^{-3}$
DYS570	Copia única, simple	(TTTC) ₁₄₋₂₄	10-26	$1,79 \times 10^{-2}$
DYS576	Copia única, simple	(AAAG) ₁₃₋₂₂	10-24	$1,43 \times 10^{-2}$
DYS627	Copia única, complejo	(AGAA) ₃ N ₁₆ (AGAG) ₃ (AAAG) ₁₂₋₂₄ N ₈₁ (AAGG) ₃	14-26	$1,23 \times 10^{-2}$
DYS460	Copia única, simple	[TAGA] ₈₋₁₃	7-13	$6,22 \times 10^{-3}$
DYS518	Copia única, complejo	(AAAG) ₃ (GAAG) ₁ (AAAG) ₁₄₋₂₂ (GGAG) ₁ (AAAG) ₄ N ₆ (AAAG) ₁₁₋₁₉ N ₂₇ (AAGG) ₄	32-46	$1,84 \times 10^{-2}$
DYS449	Copia única, complejo	(TTCT) ₁₃₋₁₉ N ₂₂ (TTCT) ₃ N ₁₂ (TTCT) ₁₃₋₁₉	23-37	$1,22 \times 10^{-2}$
DYF387S1	Copias múltiples (2), complejo	(AAAG) ₃ (GTAG) ₁ (GAAG) ₄ N ₁₆ (GAAG) ₉ (AAAG) ₁₃	32-44.2	$1,59 \times 10^{-2}$

1.2.2. SNPs de cromosoma Y

El análisis de los SNPs localizados en la región específica masculina del cromosoma Y es ampliamente empleado en la antropología molecular como herramienta única en estudios evolutivos [33]. Los SNPs de cromosoma Y juegan un papel fundamental en estudios sobre las migraciones humanas ya que permiten una evaluación efectiva de las diferencias principales entre grupos poblacionales. Los alelos de los SNPs de cromosoma Y se designan como ancestral o derivado; determinándose el ancestral por comparación con la secuencia de ADN del chimpancé [2].

Inicialmente, los Y-SNPs se analizaban mediante minisequenciación, lo que dio lugar a la identificación de unos 850 SNPs [34]. Sin embargo, el auge de las tecnologías de nueva secuenciación ha llevado a la caracterización correcta de más de 16 000 SNPs en el cromosoma Y [35].

El análisis de los haplogrupos de cromosoma Y definidos a través del estudio de marcadores binarios es la aproximación empleada para el estudio del origen de las poblaciones humanas modernas así como para determinar la variabilidad existente entre ellas. Debido a la distribución específica de los haplogrupos de cromosoma Y entre las poblaciones, el tipado de Y-SNPs permite la inferencia del origen, la evolución y la historia de las poblaciones humanas mediante la identificación de patrones migratorios masculinos [33] ya que presentan una distribución específica tanto continental como intracontinental muy marcada.

El uso de Y-SNPs en genética forense puede arrojar información sobre el origen étnico de un individuo [36]. Combinados con marcadores convencionales, los Y-SNPs pueden ser herramientas de ayuda en desastres en masa como accidentes aéreos, tsunamis o ataques terroristas donde personas de diferentes áreas geográficas se encuentran implicadas [33].

1.3. Haplogrupos de Cromosoma Y

Se denominan haplogrupos o clados la descripción del cromosoma Y a través de polimorfismos binarios; cuando se emplean únicamente STRs se conoce como haplotipo y cuando se emplea la combinación de ambos tipos de marcadores se denominan linajes [37].

En 2002, el *Y Chromosome Consortium* publicó un árbol filogenético de 153 haplogrupos basado en 243 polimorfismos únicos genotipados en una muestra representativa de la población global. En 2008, se publicó una revisión de este árbol

que contiene 311 haplogrupos e incorporando 600 polimorfismos binarios [38]. Desde entonces, un gran número de nuevos SNPs han sido descritos en el cromosoma Y, dando lugar a cambios en el árbol de haplogrupos de cromosoma Y. Recientemente, se han unificado todos estos resultados dando lugar a un árbol que incluye 359 linajes y 721 SNPs de cromosoma Y [35]. Actualmente, se emplea una nomenclatura basada en mutación, donde se indica la letra mayúscula del clado principal seguida del nombre de la mutación terminal que define el haplogrupo resultante [39].

Estudios basados en la secuenciación de regiones del cromosoma Y y la comparación de regiones homólogas con especies relacionadas, han identificado un “Adán”, ancestro común de todos los linajes de cromosoma Y; que se generó en África central-noroeste hace unos 140 000 años (Figura 5.4) [40]. Este haplogrupo se identifica con la letra A y aparece únicamente en África. Igualmente, el haplogrupo B está prácticamente restringido a África Subsahariana.

Se ha sugerido que el haplogrupo C se originó en Asia después de la salida de África de los humanos anatómicamente modernos. Se han observado algunos subhaplogrupos que se encuentran restringidos a poblaciones nativas americanas; se han observado frecuencias moderadas del haplogrupo C3* en las poblaciones Kicwha y Waorani de Ecuador [41]. Sin embargo, se ha confirmado que el haplogrupo D se encuentra confinado a Asia, siendo su lugar de origen en Asia central y Japón.

El haplogrupo E se encuentra en alta frecuencia en África, moderada en el Medio Este y el sur de Europa y con una ocurrencia ocasional en el centro y sur de Asia. El haplogrupo F es del que derivan la mayor parte de los haplogrupos.

El clado G se distribuye en Oriente Medio, el Mediterráneo y las montañas del Cáucaso. La distribución geográfica del haplogrupo H está prácticamente restringida al subcontinente Indio.

El haplogrupo I, junto con el R; representa uno de los mayores haplogrupos Europeos. A excepción del R, este clado está muy extendido en Europa y ausente en el resto del mundo. El haplogrupo J se extiende a lo largo del Medio Este, el norte de África, Europa, Asia central, Pakistán e India; siendo el haplogrupo J-M172 el más común en Europa.

El haplogrupo K presenta cuatro linajes de los cuales, el haplogrupo K2 se renombró como el nuevo haplogrupo T. La mayoría del haplogrupo L se encuentra en el subcontinente indio, aunque también está presente en Oriente Medio, Asia central, el norte de África y la costa mediterránea europea.

El haplogrupo M está restringido Oceanía y el este de Indonesia. Por otro lado, el haplogrupo N se localiza en el norte euroasiático exclusivamente. El haplogrupo O es el mayoritario en el Medio Oriente.

El haplogrupo P consiste en los haplogrupo Q y R. El haplogrupo Q se encuentra en altas frecuencias en grupos siberianos. Además, es el mayor linaje entre los nativos americanos, siendo el Q-M3 exclusivo del continente americano.

El haplogrupo R está dividido en dos ramas R1-M173 y R2-M479. Dentro del R1-M173, las variantes más extendidas en Eurasia son R1a-M420 y R1b-M343. El haplogrupo R1a-M420 es el más común en el este de Europa mientras que el R1b-M343 se encuentra principalmente en el oeste del continente. Se ha sugerido que la división de estos subclados refleja las expansiones de población que tuvieron lugar tras el periodo postglacial [42].

Por último, el haplogrupo S se encuentra principalmente en Oceanía e Indonesia.

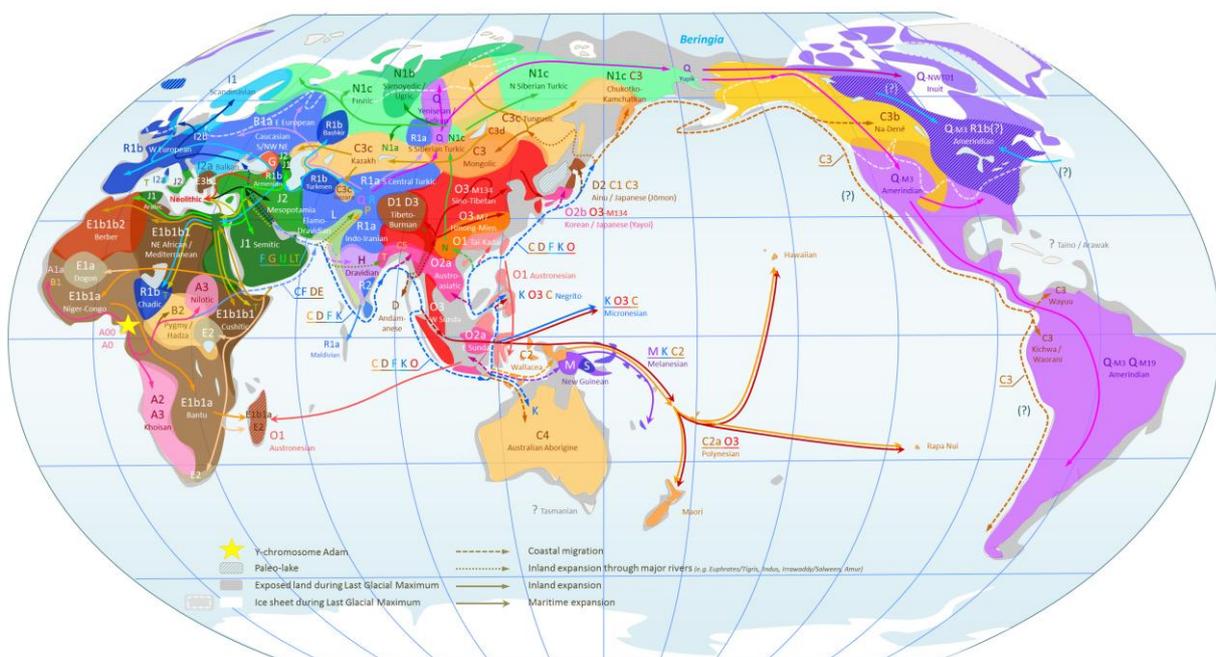


Figura 5.4: Expansión de las poblaciones humanas rastreado la distribución de haplogrupos de cromosoma Y.

1.4. Aplicaciones

El hecho de que el cromosoma Y sea haploide y en su mayoría no recombinante, ha hecho que sea un marcador molecular ampliamente usado en campos de investigación muy diversos [43,44].

1.4.1. Y-STRs como herramienta en estudios de filiación

En estudios de paternidad el análisis de los STR del cromosoma Y nos permite establecer o excluir paternidades incluso en casos en los que el individuo masculino no está disponible. Todos los parientes masculinos pertenecientes a línea paterna del individuo tendrán el mismo haplotipo de cromosoma Y ya que se transmite directamente y sin modificaciones de padres a hijos (Figura 5.5).

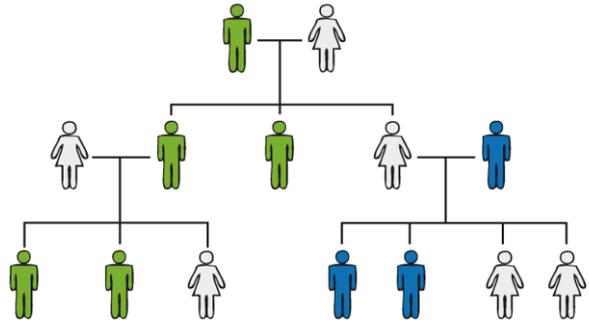


Figura 5.5. Herencia del cromosoma Y.

Actualmente, existen estudios donde se han determinado el perfil de marcadores NRY de fetos en plasma materno. Además, se ha demostrado que a través de esta técnica se pueden realizar estudios de filiación del feto [45].

1.4.2. Y-STRs como herramienta en estudios de criminalística

El estudio de polimorfismos de cromosoma Y es muy útil en delitos contra la libertad sexual, donde las muestras forenses presentan mezcla celular (víctima y agresor). Debido a su exclusividad en hombres, no es necesario separar las células espermáticas de una mezcla [46]. Se ha demostrado en numerosos estudios la eficacia de los marcadores NRY en múltiples casos, incluso en aquellos donde hay baja o incluso nula concentración espermática gracias al análisis de las células epiteliales masculinas [47]. Por otro lado, se han podido obtener perfiles de marcadores NRY completos o parciales pero probatorios en muestras cervico-vaginales tomadas hasta 9 días tras la agresión sexual [48]. Otra aplicación importante de los STRs del cromosoma Y es el estudio del componente minoritario de una mezcla de ADN si éste es masculino [49,50].

1.4.3. Estudio de la antropología evolutiva y la historia de las poblaciones

Los grupos de ligamiento ancestrales que reflejan el origen del cromosoma Y han sido altamente conservados a lo largo de la evolución. Este cromosoma ha adquirido numerosas secuencias procedentes de diversos orígenes, por tanto, es un cromosoma moderno que ha alcanzado su conformación actual a lo largo de la evolución, desde los grandes monos hasta el hombre. Es probable que los estudios

del cromosoma Y contribuyan a una mayor comprensión de los orígenes de los humanos modernos [51,52], así como de su expansión fuera de África [53]. A través del estudio de secuencias de cromosoma Y se han identificado dos cuellos de botella en la expansión de los humanos genéticamente modernos; uno durante la salida del hombre de África hace 47 000 – 52 000 años y otro hace 10 000 años que tuvo lugar debido a cambios culturales que afectaron al éxito reproductivo entre los varones [52].

Dependiendo de la población, las frecuencias de los marcadores NRY poseen una distribución variable [34]. Desde el descubrimiento de los Y-STR, se han estudiado numerosos grupos poblacionales para validar muchos de los marcadores en el campo forense [54–58]. Estas variaciones también se han aplicado en genética de poblaciones y antropología molecular [59,60]. La alta tasa de mutación de éstos proporciona una información muy adecuada para diferenciar poblaciones cercanas desde el punto de vista filogenético. El estudio de los haplogrupos de cromosoma Y demuestra las características específicas de las diferentes poblaciones de la región del Volga, a pesar de su proximidad geográfica [61]. A través del estudio conjunto de NRY y ADN mitocondrial, se ha identificado el origen norteafricano de la población Tuareg libanesa; posiblemente originada en la segunda mitad del Holoceno. Además, se ha podido identificar una contribución paterna sub-sahariana mucho más reciente [62].

1.4.4. Estudios de genealogía genética

La concordancia entre la herencia del apellido y la de las regiones no recombinantes del cromosoma Y hace de éstas una herramienta idónea para realizar estudios de genealogía [63,64]. Se observa una distribución diferencial de los subhaplogrupos de cromosoma Y en la población de la región de Flanders (Bélgica) con apellidos de procedencia francesa o procedencia flamenca. Sin embargo, los estudios de genealogía no han podido identificarlos [65].

1.4.5. Investigación en biomedicina

Se está generalizando el empleo de estos marcadores por su asociación con numerosas enfermedades. El uso de Y-STRs, X-STRs y ADNmt puede ayudar a asociar las enfermedades con haplogrupos o incluso con haplotipos particulares, para así poder realizar un perfecto diagnóstico de los pacientes a la hora de realizar un panel de marcadores genéticos [66,67]. Se ha asociado el haplogrupo

D2a1 con altos niveles de hormona luteinizante en varones japoneses [68]. Por otro lado, se ha establecido una asociación entre el haplogrupo I y un incremento en el riesgo de padecer enfermedades coronarias en varones del Reino Unido [69]. También se están descubriendo como una herramienta para poder asociar enfermedades hereditarias [70], sobre todo en aquellas con mayor incidencia en varones como puede ser el cáncer de próstata [71,72].

2. MATERIAL Y MÉTODOS

2.1. Población de Estudio

Se han estudiado un total de 146 varones adultos no relacionados entre sí de la región sureste de la Península Ibérica [Almería (34), Granada (70), Málaga (42)], ver capítulo 2 páginas 39-40. Todos los donantes fueron informados del propósito de este estudio, aceptaron su participación y firmaron un consentimiento informado según la guía ética de la declaración de Helsinki.

2.2. Análisis Genético Molecular de la Muestra

La toma de muestras así como su posterior análisis están estrictamente registrados por rigurosos controles, tanto intralaboratorio como interlaboratorio.

2.2.1. Extracción

Se tomaron muestras del epitelio bucal de todos los individuos mediante hisopos estériles de algodón. El ADN genómico se obtuvo mediante un protocolo de extracción orgánica con fenol:cloroformo:alcohol isoamílico (25:24:1) (Apéndice I).

2.2.2. Cuantificación

La estimación de la cantidad y calidad de ADN nuclear de las muestras se ha realizado a través de electroforesis en geles de agarosa al 0,8% (Apéndice I).

2.2.3. Amplificación

Las muestras se amplificaron para 16 marcadores de tipo STR: DYS389I, DYS389II, DYS390, DYS456, DYS19, DYS385, DYS458, DYS391, DYS392, DYS393, DYS439, DYS635, DYS437, DYS438, DYS448, GATA H4.1 incluidos en un kit tipo multiplex preparado para ser detectado por electroforesis capilar con emisión de fluorescencia (Tabla 5.2) [73]. Éste permite el análisis simultáneo de 16 loci tipo STR. Uno de los primers de cada pareja está marcado con un fluorocromo que confiere a cada marcador un color (verde, amarillo, azul y rojo) cuando son analizados en el aparato de electroforesis capilar.

- Azul -FAM: DYS389I, DYS389II, DYS390, DYS456
- Verde-VIC: DYS19, DYS385, DYS458
- Amarillo-NED: DYS391, DYS392, DYS393, DYS439, DYS635
- Rojo-PET: DYS437, DYS438, DYS448, GATA H4.1

Tabla 5.2: Componentes, composición y volumen del Kit AmpFISTR® Yfiler®.

Reactivo	Componentes	Volumen
AmpFISTR® Yfiler® Reaction Mix	MgCl ₂ , dATP, dGTP, dCTP, dTTP, suero de albúmina bovina y 0,05% NaO ₃	4 µl
AmpFISTR® Yfiler® Primer Set	Parejas de primers marcados para los 16 loci Y-STR	2 µl
AmpliTaq Gold® DNA Polymerase	Enzima Polimerasa 5U/µL	0,32 µl (2,5u)
AmpFISTR® Control DNA 007	ADN humano masculino 0,10 ng/µL	4mL

La cantidad de ADN por muestra osciló entre 0,25-0,5ng/µl y el volumen final de reacción fue de 10µl.

La amplificación de todos los loci se realizó en dos termocicladores 2720 Applied Biosystems y un termociclador T3 Biometra® (Figura 5.6).

En todas las reacciones se han amplificado sistemáticamente dos controles:

- Control positivo: verificar que la reacción de amplificación ha sido correcta.
- Control negativo: descartar una posible contaminación por ADN extraño.

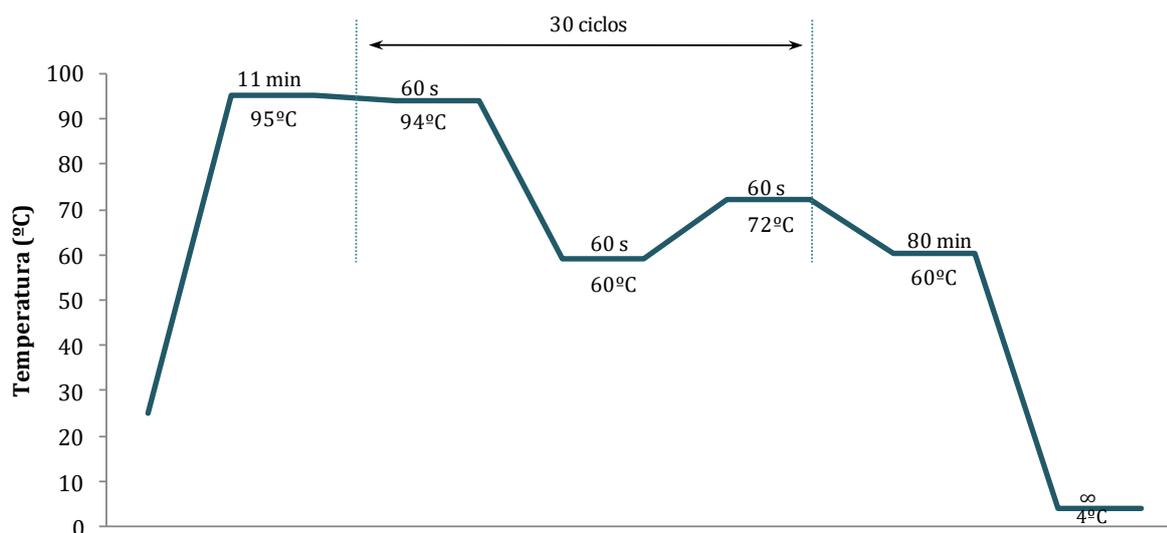


Figura 5.6: Gráfico de las condiciones de amplificación del kit AmpFISTR® Yfiler®.

2.2.4. Visualización de los resultados

Las muestras amplificadas fueron visualizadas por electroforesis capilar en un analizador genético ABI Prism 310 de Applied Biosystems®.

Para la correcta detección de los fragmentos de ADN nuclear amplificados, se han de preparar las muestras antes de separarlas por electroforesis capilar (apéndice II).

Para preparar las muestras, se añadió en cada tubo los siguientes volúmenes por muestra:

- 25 µl de formamida desionizada
- 0,5 µl del estándar de tamaño (500pb marcado en naranja- LIZ)
- 1 µl de producto amplificado

La electroforesis se llevó a cabo con Performance Optimized Polymer 4 (POP-4™; Life Technologies) un filtro G5v2.

Las condiciones de electroforesis fueron las siguientes:

- Tiempo de inyección..... 5s
- Voltaje de inyección..... 15kV
- Voltaje de electroforesis..... 15kV
- Temperatura de electroforesis..... 60°C
- Tiempo de electroforesis28min

Los electroferogramas fueron visualizados mediante GeneMapper ID-X 1.1; que permite determinar el tamaño de los fragmentos analizados. Se emplearon los *bin set* para STRs autosómicos facilitados por la casa comercial. Los alelos se designaron de acuerdo con el número de unidades de repetición basándose en un *ladder* alélico (recomendaciones de ISFG) [74].

Tanto la calidad del tipado de las muestras como la designación de alelos fueron garantizadas por el análisis simultáneo de una muestra control de genotipo conocido.

Los datos se encuentran en la base de datos YHRD (Y-Chromosome Haplotype Reference Database) [75] bajo los números de acceso:

- Granada, Spain [Spanish] YA004154
- Malaga, Spain [Spanish] YA004155
- Almeria, Spain [Spanish] YA004156

2.3. Análisis Estadístico de la Muestra

Los datos de los perfiles genéticos de los individuos que se prestaron para realizar el estudio poblacional fueron introducidos en tablas mediante un código y su perfil genético.

El cálculo de las frecuencias de las repeticiones encontradas en cada una de las poblaciones se realizó por distintos métodos. En primer lugar, la obtención de las frecuencias alélicas muestrales se determinó por el método clásico del recuento en base a los alelos observados (Li, 1976). En segundo lugar, se recalculó este mismo parámetro con dos programas estadísticos, *Arlequin 3.5.1.3* [76] y *PowerStatsv12* (Promega® Corporarion) [77]. En todos los casos se han de obtener los mismos resultados. Mediante el software *Arlequin 3.5.1.3* [76] se calculó la distribución de los genotipos observados en cada grupo mediante el estadístico F.

Todos los parámetros forenses más relevantes se calcularon mediante la hoja de cálculo *PowerStatsv12* obteniendo el poder de discriminación (PD), la probabilidad de coincidencia (MP) y el índice de contenido polimórfico (PIC).

Los haplogrupos de cromosoma Y se predijeron mediante el programa *23-Haplogroup Beta Program (On-line Software)* [78,79] y *YPredictor by Vadim Urasin v15.0*. Mediante estos programas se puede calcular el haplogrupo de cromosoma Y de un individuo partiendo de su perfil de STRs, sin necesidad de realizar el estudio de SNPs. En el caso de *23-Haplogroup Beta Program*, este cálculo es una aproximación, a partir de los datos reales incluidos en una base de datos extendida. Infieren desde su base de perfiles los resultados esperados, <http://www.hprg.com/hapest5/>. En el caso de *YPredictor by Vadim Urasin v15.0* las predicciones se basan en los árboles filogenéticos de cada haplogrupo, calculando una probabilidad que depende de los alelos de cada marcador, la tasa de mutación del marcador y el tiempo de divergencia del nodo parental. Hay que tener cuidado con este tipo de programas ya que al ser un cálculo estadístico ofrece un error [80,81].

Se anotaron los apellidos de los 146 varones estudiados, se analizaron los primeros apellidos de éstos y se compararon con el haplogrupo asociado a cada individuo así como con su haplotipo.

Para determinar la subestructura genética de la población del Sudeste de la Península Ibérica, se testó la hipótesis de una distribución aleatoria de los individuos entre parejas de poblaciones mediante un test exacto de diferenciación de la población y el análisis de la varianza molecular (AMOVA) con 10 000 repeticiones experimentales mediante *Arlequin 3.5.1.3*. Además, se utilizó

STRUCTURE 2.3.4 [82,83] para inferir grupos de individuos a partir de los datos genéticos de loci ligados.

Mediante *Network* 4.6.1.3 se determinó el ancestro común dentro de la población desde donde se originaron el resto de los haplotipos.

Se llevó a cabo un análisis de correspondencias con el programa *Statistica v9.1* (Statsoft Inc. Tulsa USA) a fin de entender la asociación entre las frecuencias alélicas y las poblaciones. Las poblaciones empleadas quedan resumidas en la tabla 5.4.

Además, se empleó la herramienta de cálculo de AMOVA y MDS del YHRD. Todas las poblaciones empleadas se encontraban en la base de datos y quedan resumidas en la tabla 5.4. Para una mejor visualización de los resultados, las distancias genéticas calculadas (Rst, donde la distancia entre dos haplotipos es la suma de sus distancias geométricas) se representaron mediante escalamiento multidimensional no métrico [84] con *IBM SPSS Statistics 20*.

Para estimar la proporción individual de ancestralidad se utilizó *STRUCTURE* 2.3.4 [82,83] Para todas las simulaciones y cálculos, se asumieron los modelos de mestizaje y no mestizaje, incluyendo información poblacional y asumiendo un modelo de frecuencias alélicas independientes. Se estableció un periodo de *burn-in* de 50 000 iteraciones seguido de 100 000 iteraciones adicionales, estimando de 1 a 10 contribuciones (*K*). Cada carrera se repitió 10 veces y se calcularon las probabilidades para cada *K* en cada grupo de carreras. Las 10 réplicas para cada *K* se evaluaron con *CLUMPP 1.1.2* (<http://www.stanford.edu/group/rosenberglab/clumpp.html>) [85]. Los grupos combinados resultantes se visualizaron mediante *DISTRUCT 1.1* (<http://www.stanford.edu/group/rosenberglab/distruct.html>) [86].

Tabla 5.4: Poblaciones, abreviatura, situación geográfica, marcadores estudiados y autor del estudio publicado.

POBLACIÓN	Código	PAÍS	Y-STRs	N	Bibliografía
Granada_Málaga_Almería	GMA	España	17	146	Presente Estudio Granada YA004154, Málaga YA004155, Almería YA004156
ANÁLISIS DE FRECUENCIAS ALÉLICAS					
Marruecos	MorMOR	Marruecos	17	257	Aboukhalid et al. 2010 [87]
Figuig Oasis	FigMOR	Marruecos	17	96	Palet et al. 2010 [88]
Área de Orán	AlgALG	Argelia	17	102	Robino et al. 2008 [89]
Bereberes de Sejenane	BSejTUN	Túnez	17	47	Frigi et al. 2006 [90]
Túnez	TunTUN	Túnez	15	134	Arredi et al. 2004 [91]
Egipto	EgyEGY	Egipto	15	76	Arredi et al. 2004 [91]
Región de Trípoli	TriLIB	Libia	17	176	Triki-Fendri et al. 2013 [92]
Anatolia Central	CAnTUR	Turquía	17	140	Alakoc et al. 2010 [93]
País Vasco	BasSPA	España	23	105	Nuñez et al. 2015 [94]
Cantabria	CanSPA	España	23	98	Nuñez et al. 2015 [94]
Barcelona	BarSPA	España	17	150	Sánchez et al. 2007 [95]
España	SpaSPA	España	17	295	Martínez-González. Datos no publicados.
Norte de Portugal	NorPOR	Portugal	17	175	Pontes et al. 2007[96]
Italia	ItaITA	Italia	15	154	Turrina et al. 2006 [97]
Grecia	GreGRE	Grecia	11	200	Katsaloulis et al. 2013 [98]
Croacia	CroCRO	Croacia	17	1100	Mršić et al. 2012 [99]
Vojvodina	VojSER	Serbia	17	185	Veselinovic et al. 2008 [100]
Holanda	HolHOL	Holanda	36	2085	Westen et al. 2015 [101]
Guinea Ecuatorial	EqGEQG	Guinea Ecuatorial	16	100	Arroyo-Pardo et al. 2005 [102]
Ovambo	OvaNAM	Namibia	28	52	Fujihara et al. 2009 [103]
ANÁLISIS DE DISTANCIAS GENÉTICAS					
Argelia	AlgALG	Argelia	17	102	YA003114 Robino et al. 2008 [89]
Libia	LibLIB	Libia	17	360	YA003680 Elmrghni et al. 2012 [104]; YA003083, YA003752 Triki-Fendri et al. 2013 [92]; YA003669 Ottoni et al. 2011 [62]; YA003690; YA003691
Marruecos	MorMOR	Marruecos	17	445	YA003437, YA003438, YA003439 Aboukhalid et al 2010 [87]; YA003442, YA003443 Palet et al. 2010 [88]; YA003520 Laouina et al. 2011 [105]
Túnez	TunTUN	Túnez	17	153	YA003024, YA003025, YA003026, YA003027 Cherni et al. 2005 [106]; YA003482 Ayadi et al. 2006 [107]; YA003743
Egipto	EgyEGY	Egipto	17	159	YA003080 Manni et al. 2002 [108]; YA003400, YA003401, YA003403, YA003404
Turquía	TurTUR	Turquía	17	296	YA002964 Henke et al. 2001 [109], YA003719, YA003907
Andalucía	AndSPA	España	17	149	YA004020, YA004042
Granada	GraSPA	España	17	180	YA003685 Ambrosio et al. 2012 [110]
Huelva	HueSPA	España	17	167	YA003684 Ambrosio et al. 2012 [110]

País Vasco	BasSPA	España	27	300	YA003184 García et al. 2015 [111]; YA004063
Galicia	GalSPA	España	17	89	YA003791
Cataluña	CatSPA	España	17	1388	YA004023
Italia	ItaITA	Italia	17	3118	YA003991 Robino et al. 2015 [112]; YA003068 Grignani et al. 2000 [113]; YA003067, YA003088, YA003090, YA003093, YA003095 Presciuttini et al. 2001 [114]; YA003891 Robino et al. 2006 [115]; YA003447 Ferri et al. 2008 [116]; YA003295 Cerri et al. 2005 [117]; YA003327 Turrina et al. 2006 [97]; YA003405 Ghiani et al. 2002 [118]; YA003262 Ferri et al. 2009 [119]; YA003069 Onofri et al. 2007 [120]; YA003471 Verzeletti et al. 2009 [121]; YA003490, YA003491 Rodríguez et al. 2009 [122]; YA003721, YA003722, YA003723, YA003724, YA003725 YA003726 YA003744 Brisighelli et al. 2012 [123]; YA003730 Pigionica et al. 2013 [124]; YA003974; YA003786; YA003873; YA003435; YA004045; YA003926; YA003982; YA003983;
Portugal	PorPOR	Portugal	17	534	YA003125 Fernandes et al. 2001 [125]; YA003016 González Neira et al. 2000 [126]; YA003145 & YA003146 Carvalho et al. 2003 [127]; YA003015 & YA003017 Alves et al. 2007 [128]; YA003293
Grecia	GreGRE	Grecia	17	559	YA003127 Parreira et al. 2002 [129]; YA003097, YA003098, YA003099, YA003100, YA003101, YA003102, YA003103, Robino et al. 2004 [130]; YA003325 Bosch et al. 2006 [131]; YA003465 Kovatsi et al. 2009 [132]; YA004096, YA003647 Katsaloulis et al. 2013 [98], YA003862; YA004097
Nigeria	NigNIG	Nigeria	17	135	YA003874; YA003940; YA003973.
Somalia	SomSOM	Somalia	17	172	YA003284 Hallenberg et al. 2005 [133]
ANÁLISIS DE ESTRUCTURAS					
España	SpaSPA	España	17	295	Martínez-González. Datos no publicados.
Italia	ItaITA	Italia	17		Brisighelli et al. 2012 [123]
Holanda	HolHOL	Holanda	36	220	Westen et al. 2015 [101]
Marruecos	MorMOR	Marruecos	17	257	Aboukhalid et al. 2010 [87]
Área de Orán	ArgARG	Argelia	17	102	Robino et al. 2008 [89]
Región de Trípoli	TriLIB	Libia	17	176	Triki-Fendri et al. 2013 [92]
Ovambo	OvaNAM	Namibia	28	52	Fujihara et al. 2009 [103]

3. RESULTADOS Y DISCUSIÓN

3.1 Estudio Poblacional

La distribución de las frecuencias alélicas de los 16 loci STR de cromosoma Y estudiados queda resumida en la tabla 5.6. La tabla 5.5 muestra el poder de discriminación, el índice de contenido polimórfico y la probabilidad de coincidencia.

Los marcadores más informativos son DYS385 y DYS458; mientras que el marcador menos informativo es el DYS393. El poder de discriminación combinado de toda la población es $1 - 6,17168 \cdot 10^{-8}$.

Tabla 5.5: Resumen de los parámetros forenses más relevantes.

	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS393	DYS391
PD	0,635	0,561	0,607	0,676	0,790	0,538	0,441	0,563
PIC	0,573	0,495	0,546	0,619	0,766	0,498	0,398	0,480
MP	0,365	0,439	0,393	0,325	0,210	0,462	0,559	0,437

	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	DYS385 A/B
PD	0,667	0,690	0,589	0,603	0,556	0,624	0,667	0,864
PIC	0,611	0,652	0,518	0,523	0,456	0,570	0,624	0,859
MP	0,333	0,310	0,411	0,397	0,444	0,376	0,333	0,136

Debido a la herencia en bloque del cromosoma Y, no es un buen marcador para estudios de identificación, pero sí puede emplearse como una herramienta que ayude a discernir y aumentar o reducir el campo de búsqueda, siempre acompañado de otros datos. El éxito de discriminación del kit Yfiler es la inclusión de gran cantidad de marcadores, muchos de ellos con tasas de mutación relativamente alta. Sin embargo, la inclusión de marcadores altamente mutables recientemente descritos, hacen que el cromosoma Y sea una herramienta mucho más específica en estudios de identificación [29].

Se han observado 139 haplotipos únicos en toda la población, donde 133 haplotipos sólo se han descrito en una ocasión, 5 haplotipos en dos ocasiones y un haplotipo 3 veces. La diversidad haplotípica del cromosoma Y en la población es de 0,9974 y la capacidad de discriminación del 95,21% $[(N^{\circ} \text{ Haplotipos}/n) \cdot 100]$.

Tabla 5.6: Frecuencias alélicas para los 16 loci STR de cromosoma Y estudiados en la población GMA.

<i>Alelo</i>	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	<i>Alelo</i>	DYS385 A/B
8								0,007						0,007		10,14	0,014
9								0,068				0,014		0,110		11,11	0,034
10								0,548	0,082		0,007	0,034		0,274		11,13	0,041
10,11									0,014							11,14	0,336
11								0,363	0,301		0,336	0,411		0,055		11,15	0,062
12	0,014	0,130					0,171	0,014	0,466		0,062	0,473		0,534		11,16	0,007
13	0,027	0,589				0,144	0,719		0,137		0,541	0,062		0,021		11,18	0,007
14	0,041	0,274		0,048	0,644	0,110					0,041	0,007	0,459			12,12	0,034
15	0,486	0,007		0,103	0,158						0,014		0,479			12,13	0,007
16	0,349			0,144	0,034								0,062			12,14	0,062
16.2				0,007												12,15	0,014
17	0,055			0,370	0,021											12,16	0,007
17.2				0,034												13,13	0,007
18	0,027			0,178											0,166	13,14	0,055
18.2															0,007	13,15	0,021
19				0,068											0,503	13,16	0,048
20				0,048					0,075						0,207	13,18	0,014
21			0,021						0,205						0,090	13,19	0,021
22			0,068						0,116							14,14	0,041
23			0,301						0,486						0,028	14,15	0,014
24			0,541						0,110							14,16	0,034
25			0,068						0,007							14,18	0,007
27				0,021												14,19	0,007
28				0,096												15,15	0,014
29				0,432												15,16	0,007
30				0,349												15,17	0,014
31				0,082												15,18	0,007
32				0,021												15,19	0,007
																16,17	0,014
																16,18	0,007
																17,17	0,014
																17,18	0,014
																17,19	0,007
																18,18	0,007

3.2 Estudio de la Estructura de la Población

Para determinar si el estudio de los marcadores NRY permite establecer una subestructura de la población de estudio en las tres provincias de las que proceden las muestras, se realizaron estudios de AMOVA y STRUCTURE.

El análisis AMOVA no muestra una subestructura genética significativa entre los individuos de las tres provincias de estudio, por lo que podemos agruparlos dentro de la misma población ($p > 0,05$). Cabe destacar que toda la variación determinada a través del estudio de los NRY es la que se observa entre los individuos de una misma población (Tabla 5.7).

Tabla 5.7: Diseño AMOVA y resultados de las tres provincias del Sudeste de la Península Ibérica.

Diseño	Fuente de Variación	d.f	Suma de cuadrados	Componentes de Varianza	Porcentaje de variación
3 Grupos: Grupo 1= Almería Grupo 2= Málaga Grupo 3= Granada	Entre poblaciones	2	1,032	0,00036 Va	0,07
	Dentro de las poblaciones	143	71,413	0,49939 Vb	99,93
	Total	145	72,445	0,49975	
					Índices de fijación
					FST: 0,00071

El estudio de la subestructura genética entre estos tres grupos mediante el programa STRUCTURE corrobora estos resultados. El modelo con mayor valor de probabilidad es K=7 ($\ln P [D] = -1067,04$). Todas las componentes se distribuyen de la misma manera entre las tres poblaciones. Esto indica que los datos genéticos favorecen un único nodo común para las tres poblaciones estudiadas. Estos resultados se ajustan a las expectativas históricas y socioculturales ya que comparten el mismo origen además de una gran proximidad geográfica (Figura 5.7).

El estudio profundo del porcentaje de cada componente revela una asociación entre éstas y los haplogrupos encontrados en la población GMA. Así como la componente de color naranja diferencia a la gran mayoría de los individuos con haplogrupo G junto con algunos individuos del haplogrupo I (8,51%), la componente de color verde recoge a los individuos con haplogrupo J (15,7%). Por otro lado, la componente azul incluye casi todos los individuos con haplogrupo E (12,87%).

Todos los individuos con haplogrupo R se encuentran repartidos entre las componentes celeste y amarillo, con pequeñas influencias de otras componentes presentes en haplogrupos más ancestrales. Tanto la componente roja como rosa diferencian a individuos pertenecientes a diferentes haplogrupos; no siendo ninguna de las dos la predominante en un grupo de muestras sino que se encuentran las dos en proporciones similares.

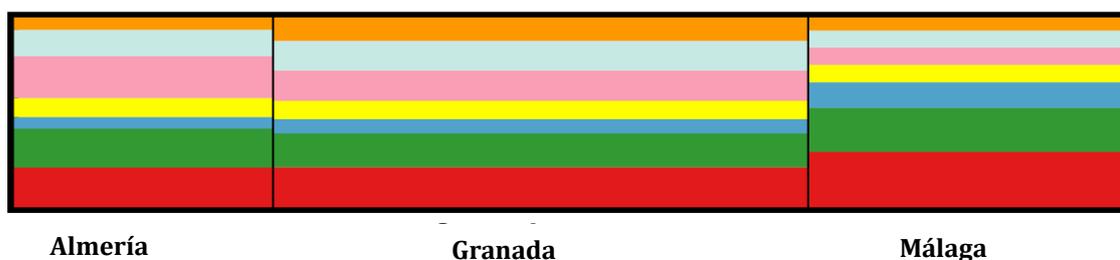


Figura 5.7: Análisis estructural de K= 7 para las 3 poblaciones. Asignación media de cada población a los clústeres resultantes del análisis de estructura.

Se ha descrito previamente que el uso del software STRUCTURE no es adecuado para el análisis de marcadores fuertemente ligados, como los encontrados en el cromosoma Y [134] ya que la distribución de los haplogrupos definen la subestructura poblacional e incluso el nivel de similitud entre poblaciones. Sin embargo, estudios previos han demostrado la capacidad de este software de diferenciar poblaciones ancestrales en población mestiza [135,136].

3.3 Estudio de Haplogrupos

Los haplogrupos predichos mediante el programa *23-Haplogroup Beta Program (On-line Software)* [78,79] y *YPredictor by Vadim Urasin v15.0* se presentan en la tabla 5.8 de forma más expandida y en la figura 5.8 de manera resumida.

El haplogrupo mayoritario presente en la población GMA es el haplogrupo R1b1b2 (58,22%). Estos resultados concuerdan con los observados en la población española donde se observan frecuencias que varían del 55% en la población del oeste de Andalucía al 87,1% en la población vasca [137]. El haplogrupo R1b1b2 haplogrupo es uno de los más frecuentes en el continente europeo con un gradiente de frecuencias que incrementa de este a oeste. Este gradiente indica que el haplogrupo se expandió por Europa a partir de una única fuente proveniente del Oriente Próximo durante el Neolítico [137].

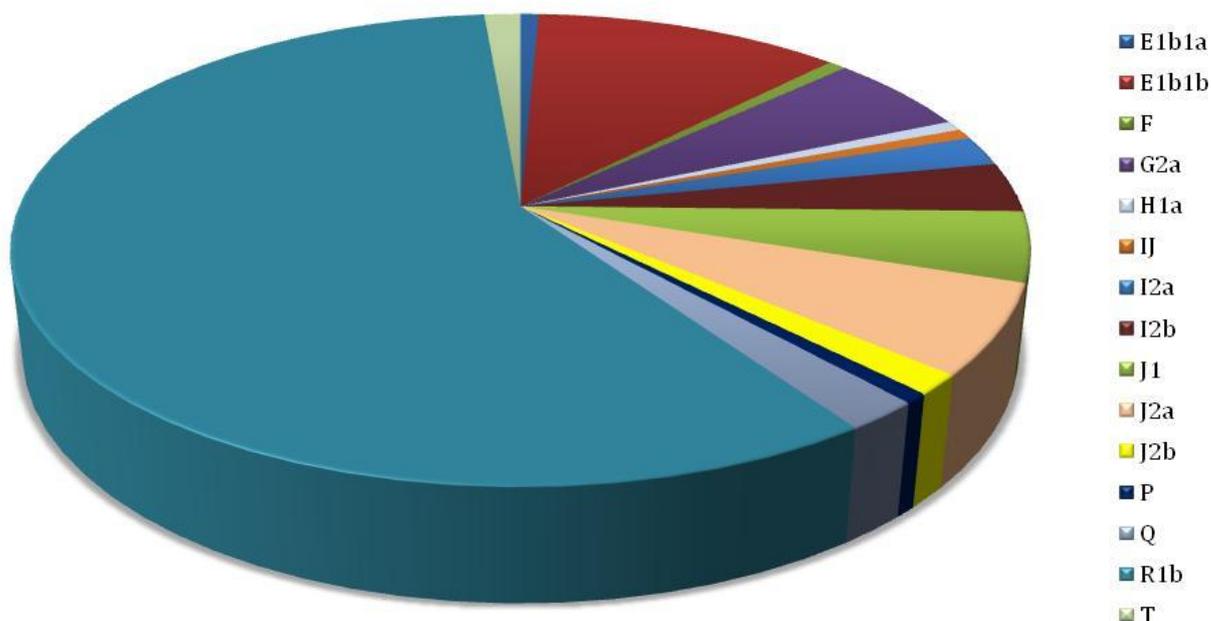


Figura 5.8: Gráfico de sectores de las frecuencias de los haplogrupos en la población GMA.

Tabla 5.8: Frecuencia de los haplogrupos predichos en la población GMA.

_J	E1b1a1g	E1b1b1	E1b1b1a1b	E1b1b1a1c	E1b1b1a1d	E1b1b1b	E1b1b1c	
M429	U175	M35	V13	V22	V65	M81	M123	
0,0068	0,0068	0,0068	0,0205	0,0137	0,0137	0,0479	0,0137	
F-M89	G2a-P15	H1a-M82	I2a-P37.2	I2a2a-M223	I2a2-S154	J1*-M267*	J1c3-P58	
0,0068	0,0548	0,0068	0,0205	0,0274	0,0068	0,0068	0,0411	
J2a4	J2b1	J2b2f-L283	P-M45	Q1a	Q1b	R1b1b2-M269	T1a1a1	T1a2
L26	M205			MEH2	M378		P77	L131
0,0616	0,0068	0,0068	0,0068	0,0137	0,0068	0,5822	0,0068	0,0068

El segundo haplogrupo más frecuente en la población GMA es el E1b1b1, se han descrito un gran número de sus subhaplogrupos (11,64%). Este haplogrupo se originó en el cuerno de África y se encuentra por todo el norte de África y en menores frecuencias por Europa Mediterránea. Entre los subhaplogrupos observados, el E1b1b1b es el más frecuente con un porcentaje de 4,79%. Las frecuencias más altas de este haplogrupo se observan en el norte de África siendo de más del 80% en población berebere marroquí [138]. Sin embargo, este haplogrupo es muy raro en Europa, excepto en la Península Ibérica donde se encuentra en una frecuencia media de 5% [138,139].

El haplogrupo J2a4 presenta una proporción del 6,16%. Otros subhaplogrupos del haplogrupo J encontrados en la población GMA son el J1c3 (4,11%), típico de las poblaciones de la península arábiga (40%-75% de los linajes masculinos) [140]; y J1-M267, J2b1 y J2bf2 en proporciones iguales (0,68%). El J1 se encuentra en bajas frecuencias en subpoblaciones del sur de Europa. Sin embargo, ambos linajes del J2b se observan principalmente en el sud-este europeo.

El haplogrupo G2a-P15 se observa en un 5,48% de los individuos. Las mayores frecuencias de este se han descrito en el Cáucaso y descienden drásticamente hacia Oriente Medio y el Sur de Europa [141]. Este haplogrupo se ha descrito en Italia (10%), Cerdeña (14,1%), Cataluña (8,3%), y Andalucía (2%) [142]. Al igual que E1b1b, G2a se originó durante el Neolítico diseminándose desde Oriente Medio hacia Europa con la aparición de la agricultura [142].

Un 5,47% de los individuos pertenecen a linajes del haplogrupo I2 distribuido principalmente en países de Europa del Este (Bosnia, Serbia, Montenegro, Rumanía...).

Un individuo pertenece al haplogrupo P, del cual derivan el haplogrupo R y el Q y está ampliamente distribuido en Europa, Asia Central y el noreste asiático. Además, se observa un individuo con haplogrupo H1a, predominante en individuos de etnia gitana [143].

También se observan dos individuos pertenecientes a linajes del haplogrupo T, T1a1a1 y T1a2. T1a1a1 se localiza principalmente en Oriente Medio y el oeste europeo. Por otro lado, T1a2 se encuentra mayoritariamente en el norte, este y sudeste de Europa y la península de Anatolia [144].

Por último, 2,05% de la población pertenece a linajes del haplogrupo Q1. Aunque éste es el más extendido en poblaciones nativas americanas [145,146], se encuentra en bajas frecuencias en Europa [147].

3.3.1. Estudio de todos los haplogrupos en la población GMA

Tras la determinación de los haplogrupos presentes en la población GMA, se realizaron estudios de redes Network para determinar aquellos haplotipos más antiguos. Este estudio también permite comprobar si las tres poblaciones difieren lo suficiente como para separarse por ramas o si se distribuyen uniformemente en el diagrama de redes.

Se ha reducido el estudio a aquellos marcadores con tasa de mutación inferior a 0,0025 (DYS19, DYS389I, DYS390, DYS391, DYS392, DYS393, DYS385, DYS438, DYS437 y DYS448).

Para poder representar mejor todos los haplogrupos presentes en la población GMA, se ha reducido el número de individuos empleados en el estudio de Network. Se han estudiado 79 individuos y todos los haplogrupos presentes en la población se encuentran representados en función de su porcentaje.

En la figura 5.9 se puede ver la distribución de los individuos en función de su haplogrupo. Se ve claramente que los haplogrupos que forman el torso de la representación (líneas negras) son individuos de los haplogrupos G2a (GMA43 y GMA127), J2a (GMA63 y GMA87) y R1b1b2. Primero se forma el grupo del haplogrupo R1b1b2, con los dos grupos más grandes (9 y 8 individuos respectivamente) y posteriormente se desarrollan los otros dos grupos.

Por lo tanto, los individuos de los que divergen el resto son aquellos con los haplotipos GMA62 (n=8) y GMA114 (n=9) para los 10 marcadores Y-STRs estudiados.

Resulta muy interesante la distribución que siguen los haplogrupos dentro del diagrama de red, ya que como se puede observar, todos los individuos con un mismo haplogrupo se distribuyen dentro de las mismas ramas y muy cercanos entre ellos.

Se distinguen claramente dos grupos, uno formado por los haplogrupos P, Q, R y T; todos dentro de la misma rama evolutiva; y otro que contiene el resto de haplogrupos, mucho más antiguos que los anteriores.

Dentro de esta segunda rama, se pueden distinguir dos brazos, uno que incluye los haplogrupos G, H, I y J2, todos ellos dentro de la misma rama evolutiva. El otro brazo contiene a los individuos con haplogrupos J1 y E. Sin embargo, los varones J1 se encuentran muy cercanos a los J2. Los individuos con haplogrupo E, aun siendo un haplogrupo mucho más antiguo que el resto y anterior en el árbol evolutivo del cromosoma Y (Figura 5.10) [39]; se localizan muy alejados del nodo central, en la periferia de la imagen, ya que los estudios de Network tienden a agrupar el mayor número de individuos con mismo haplotipo en el torso de la red.

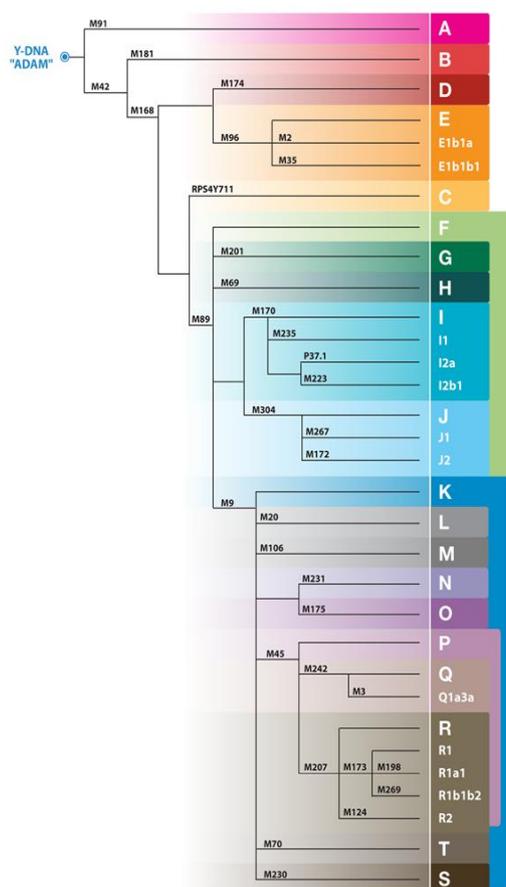


Figura 5.10: Árbol evolutivo de los haplogrupos de cromosoma Y. Imagen tomada de *Human Origins by Godfrey Oswald*.

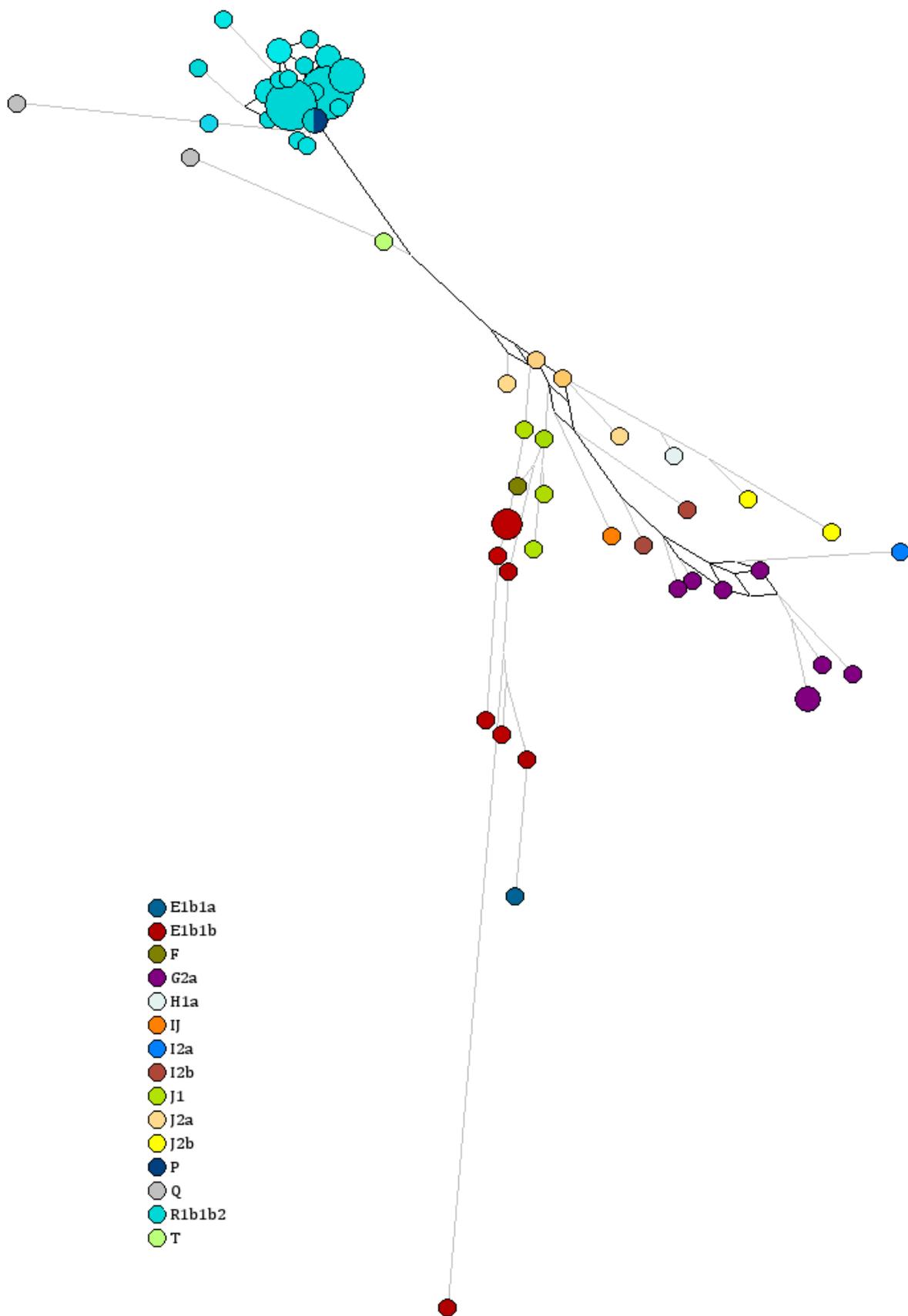


Figura 5.9: Red Network de todos los haplogrupos presentes en la población GMA

3.3.2. Estudio de los haplogrupos mayoritarios en la población GMA

En general, se puede ver como los individuos de las tres poblaciones se distribuyen de forma equitativa en a lo largo del árbol, sin observar la agrupación de los individuos en función de la provincia de procedencia. Además, se dan casos en que individuos de las tres provincias estudiadas comparten un mismo haplotipo para los 10 marcadores STRs estudiados.

- **R1b1b2:** El haplotipo más antiguo del haplogrupo R1b1b2 en el GMA107 que está compuesto por 4 individuos (GMA57, GMA62, GMA72 y GMA107). (Figura 5.11). De este grupo se generan el resto de haplotipos encontrados. Los más alejados del grupo central son los haplotipos GMA11 y GMA47.

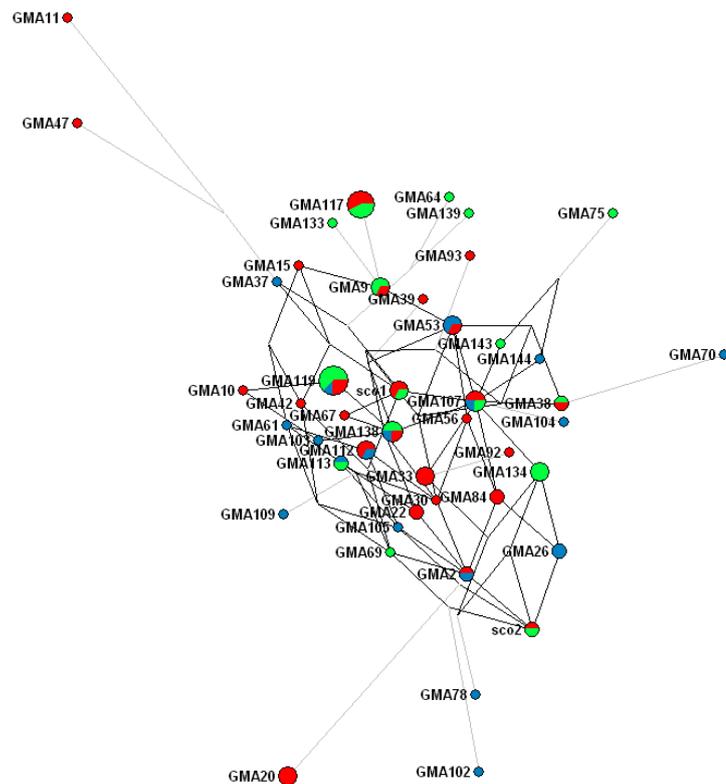


Figura 5.11: Red Network de todos los individuos con haplogrupo R1b1b2.

- **J1:** El torso de la red está formado por los haplotipos GMA1 (n=2) y GMA121; del haplogrupo J1c3. El individuo más alejado del centro es el individuo GMA122 que tiene haplogrupo IJ-M429 (de este derivan todos los individuos I y J).
- **E1b1b:** En el centro de la red se encuentran los haplotipos GMA129, GMA132 y GMA60. El individuo que se encuentra más alejado es GMA141 que tiene un haplogrupo E1b1b1*-M35(xM78,M81,M123).

- **G2a:** El torso de la red está compuesto por dos haplotipos (GMA43 y GMA127). A partir de estos dos individuos van apareciendo el resto de los individuos a medida que se acumulan mutaciones.
- **J2:** El inicio de la red lo forman tres haplotipos (GMA35, GMA63, GMA87), todos con haplogrupo J2a4. Los individuos más alejados son GMA23 y GMA41. Ambos son los únicos individuos dentro de los que tienen haplogrupo J2 que no pertenecen al haplogrupo J2a4, sino que pertenecen a J2b1 y J2b2f, respectivamente.

3.4 Estudio de Apellidos

Los apellidos son ampliamente empleados para detectar señales de estructura genética en poblaciones humanas [64,148]. En las sociedades occidentales, los apellidos se transmiten a lo largo de la vía paterna y se heredan como alelos genéticos [51,63]. Por ello, su transmisión debería coincidir con la de la región no recombinante del cromosoma Y.

El uso generalizado de los apellidos en Europa empezó durante la Edad Media. El corto tiempo evolutivo de éstos, los nacimientos ilegítimos, las adopciones y los cambios de los apellidos a partir de su forma original han hecho que la relación entre los apellidos y NRY se vea debilitada. En la gran mayoría de los países, el número de apellidos conocidos supera en varios órdenes de magnitud el número de linajes de cromosoma Y presentes, incluso el número de haplotipos presentes [149]. Esta gran disparidad implica que apellidos diferentes compartan el mismo linaje de cromosoma Y. Parece más probable que exista una relación más fuerte entre apellidos muy raros y linajes de cromosoma Y [148]. Sin embargo, este tipo de apellidos se da en un número muy reducido de individuos de la población y por ello son de poco interés en estudios de genética de poblaciones, epidemiología y genética forense [149].

Dentro de los individuos estudiados para cromosoma Y, se ha realizado el estudio del primer apellido de ellos para establecer una relación entre la distribución de los haplotipos y haplogrupos de cromosoma Y y los apellidos observados en la población masculina.

3.4.1. Haplogrupos y apellidos

Se han detectado 20 apellidos repetidos entre los 108 apellidos diferentes encontrados entre los 146 individuos estudiados (58 individuos), la mayoría

asociados al haplogrupo R1b1b2-M269 (28/58). Otros haplogrupos frecuentes entre los apellidos más observados son varios linajes del haplogrupo E (E1b1b1c-M123, E1b1b1a1b-V13, E1b1b1b1b-M81, E1b1b1*-M35, E1b1b1a1d-V65, E1b1a1g-U175 y E1b1b1a1c-V22), linajes del haplogrupo I (I2a-P37.2 y I2b1-M233), linajes del haplogrupo J (J2a4-L26, J2b2f-L283 y J1c3-P58) y el linaje G2a-P15. La figura 5.12 muestra la distribución de los haplogrupos entre los 4 apellidos más frecuentes encontrados en la población así como para aquellos que se encuentran en 3 o en 2 individuos. Se observa como de los 5 individuos que se apellidan García, 4 pertenecen al haplogrupo R1b1b2-M269. Sin embargo, no es un hecho destacable ya que tanto este apellido como el haplogrupo son muy comunes en la población española. Solamente 3 apellidos engloban individuos que presentan el mismo haplogrupo, Muñoz (2 ind), Martín (3 ind) y Simón (2 ind), todos ellos R1b1b2-M269.

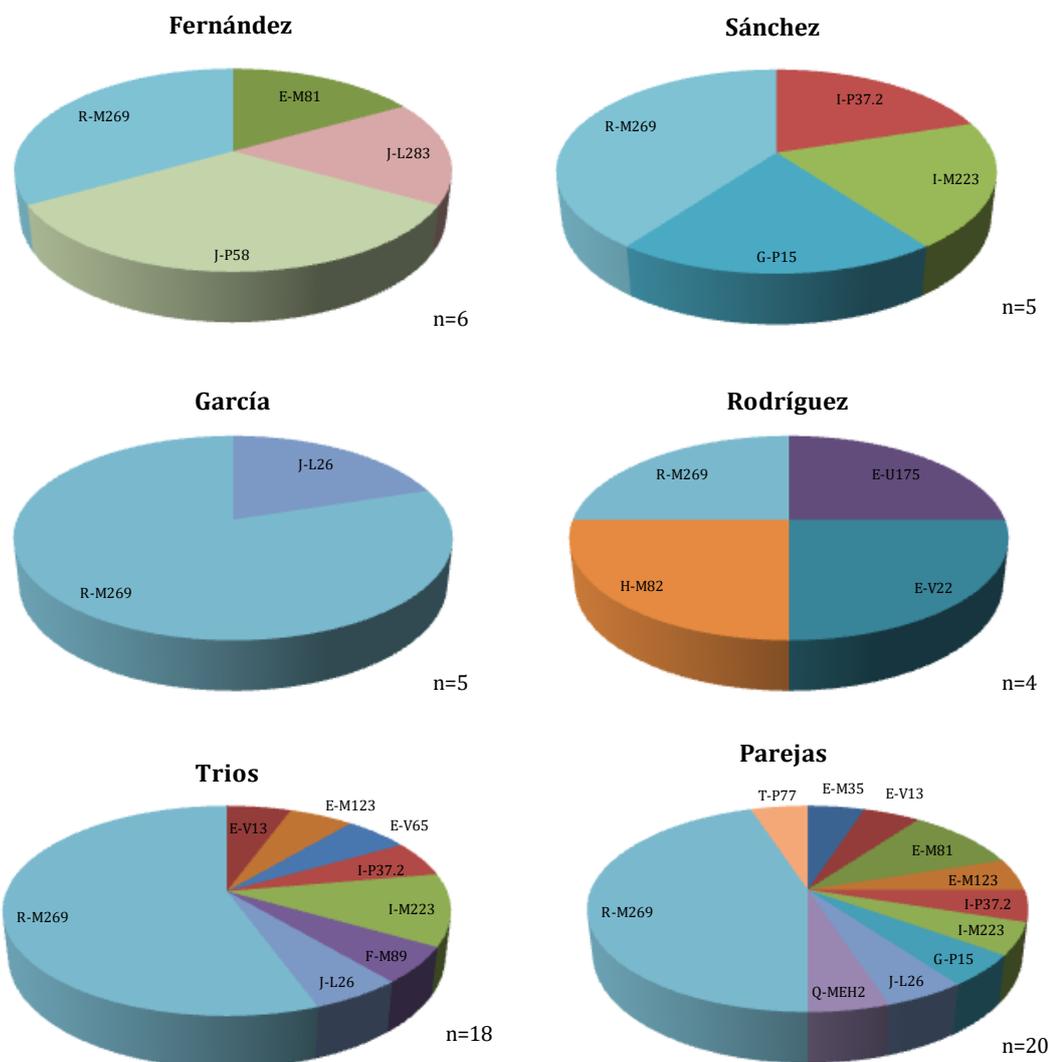


Figura 5.12: Distribución de los haplogrupos de cromosoma Y entre los apellidos observados en más de una ocasión en la población GMA.

El haplogrupo R1b1b2 es uno de los más frecuentes en el continente europeo con un gradiente de frecuencias que incrementa de este a oeste. Este gradiente indica que el haplogrupo se expandió por Europa a partir de una única fuente proveniente del Oriente Próximo durante el Neolítico [137]. Las frecuencias más altas del haplogrupo E se observan en el norte de África siendo de más del 80% en población berebere marroquí [138]. Sin embargo, este haplogrupo es muy raro en Europa, excepto en la Península Ibérica donde se encuentra en una frecuencia media de 5% [138,139]. Tanto los linajes del haplogrupo J como los del I se distribuyen principalmente en poblaciones del sud-este Europeo. La presencia de linajes que no se presentan a altas frecuencias en la población del sud-oeste Europeo se puede interpretar como consecuencia de la historia genética de la Península Ibérica, en la que se han recibido grandes influencias tanto del norte de Europa como las poblaciones de la cuenca Mediterránea.

Por último, 6 de los 108 apellidos presentes en la muestra tienen origen etimológico árabe. Cinco de ellos están ligados al haplogrupo R1b1b2-M269 mientras que sólo uno está ligado al haplogrupo J1c3-P58, típico de las poblaciones de la península arábiga.

3.4.2. Haplotipos y apellidos

A diferencia de los haplogrupos, los haplotipos basados en múltiples STRs de cromosoma Y son indicadores de coancestralidad reciente mucho más sensibles [148].

Se han observado 139 haplotipos únicos en toda la población, donde 133 haplotipos sólo se han descrito en una ocasión, 5 haplotipos en dos ocasiones y un haplotipo 3 veces. Se ha señalado que varones no relacionados entre sí que comparten el mismo apellido son significativamente más propensos a compartir haplotipos que varones con apellidos diferentes [150,151]. Sin embargo, en la población GMA ninguno de los individuos que presentan haplotipos compartidos tienen el mismo apellido. Además, a excepción de Martín y Gómez, el resto de individuos con haplotipo compartido presentan apellidos únicos.

3.5 Estudio de Frecuencias Alélicas

Se realizaron estudios de análisis de correspondencias a través de las frecuencias alélicas de las poblaciones españolas y del norte de África con Statistica v9.1. Para

facilitar la interpretación de los resultados se omitieron los marcadores en las figuras.

En la figura 5.13, se observan dos agrupaciones principales. Una se encuentra formada por las poblaciones de la Península Ibérica y otras poblaciones europeas, tanto mediterráneas como del norte de Europa, enmarcadas en color rojo. Debido a las frecuencias genéticas en los marcadores DYS438 (alelos 12 y 13) y DYS392 (alelos 13), la población del País Vasco (BasSPA) se aleja del grupo de poblaciones españolas.

Por otro lado, enmarcadas en color azul, se encuentran las poblaciones del norte de África. De ese grupo se aleja la población árabe argelina (ArgARG).

Debido a la gran diversidad presente en la población turca contemporánea, así como por el contacto que existe entre ésta y las poblaciones vecinas de Oriente Medio, los Balcanes y caucásicas; cabría esperar que la población turca se localizara entre las poblaciones mediterráneas y africanas. Sin embargo, la población escogida se localiza en el centro de Anatolia y está formada por individuos de cuatro asentamientos rurales con una fuerte estructura patrilineal. Este hecho ha influido fuertemente en la diversidad genética de esta región [93].

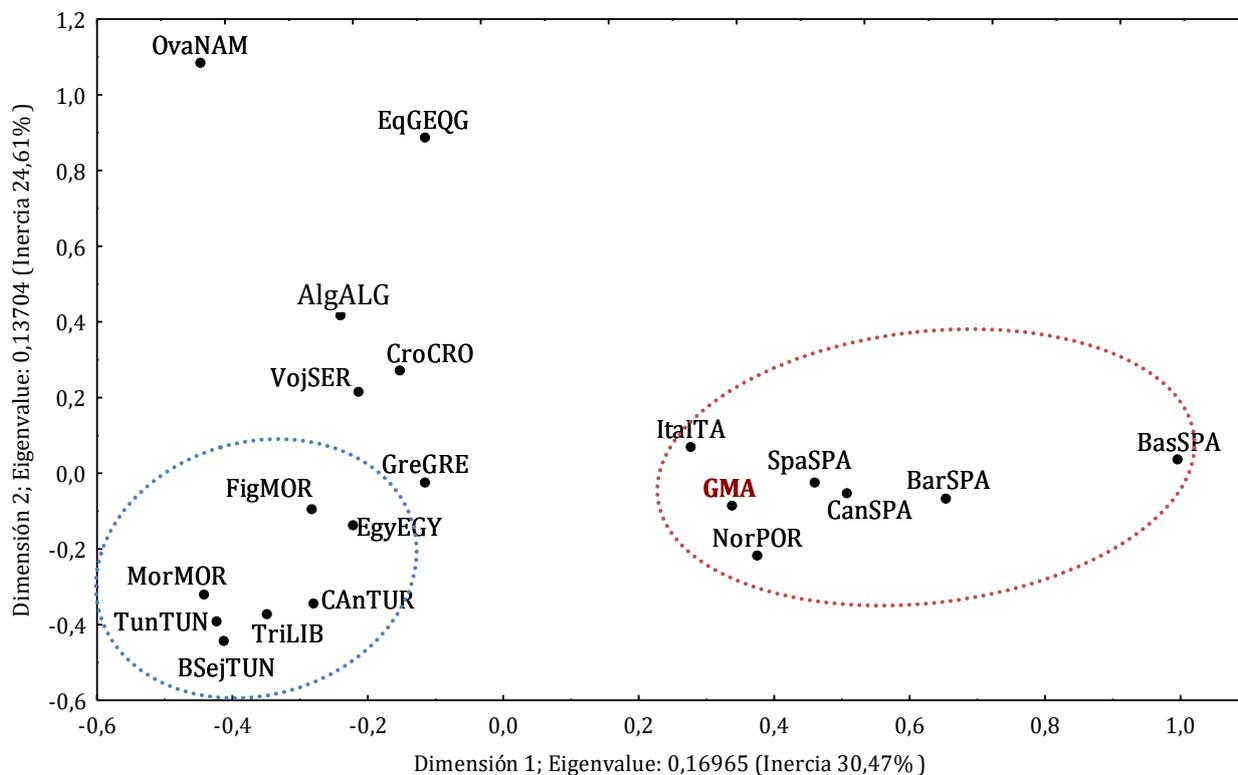


Figura 5.13: Análisis de correspondencia en 2 dimensiones que establece la relación entre las frecuencias alélicas y las poblaciones estudiadas.

Cabe destacar las poblaciones de Guinea Ecuatorial (EqGEQG) y Ovambo de Namibia (OvaNAM), que se sitúan muy alejadas del núcleo central de poblaciones, ambas poblaciones del África subsahariana. La población Ovambo es un grupo etnolingüístico de las sabanas del norte de Namibia y el sur de Angola formado por ocho grupos tribales cuyo idioma es una lengua bantú. Estudios de AMOVA han demostrado que esta población no presenta diferencias significativas con la población Bantú [103].

3.6 Estudio de Distancias Genéticas

Las distancias genéticas *Rst* se calcularon mediante la herramienta de cálculo de AMOVA y MDS de YRHD; para descifrar las relaciones genéticas entre la población GMA y 16 poblaciones incluidas dentro de la base de datos YHRD (Tabla 5.4).

Para una mejor visualización de los resultados, las distancias genéticas calculadas se representaron mediante escalamiento multidimensional no métrico con *IBM SPSS Statistics 20* (Figura 5.14).

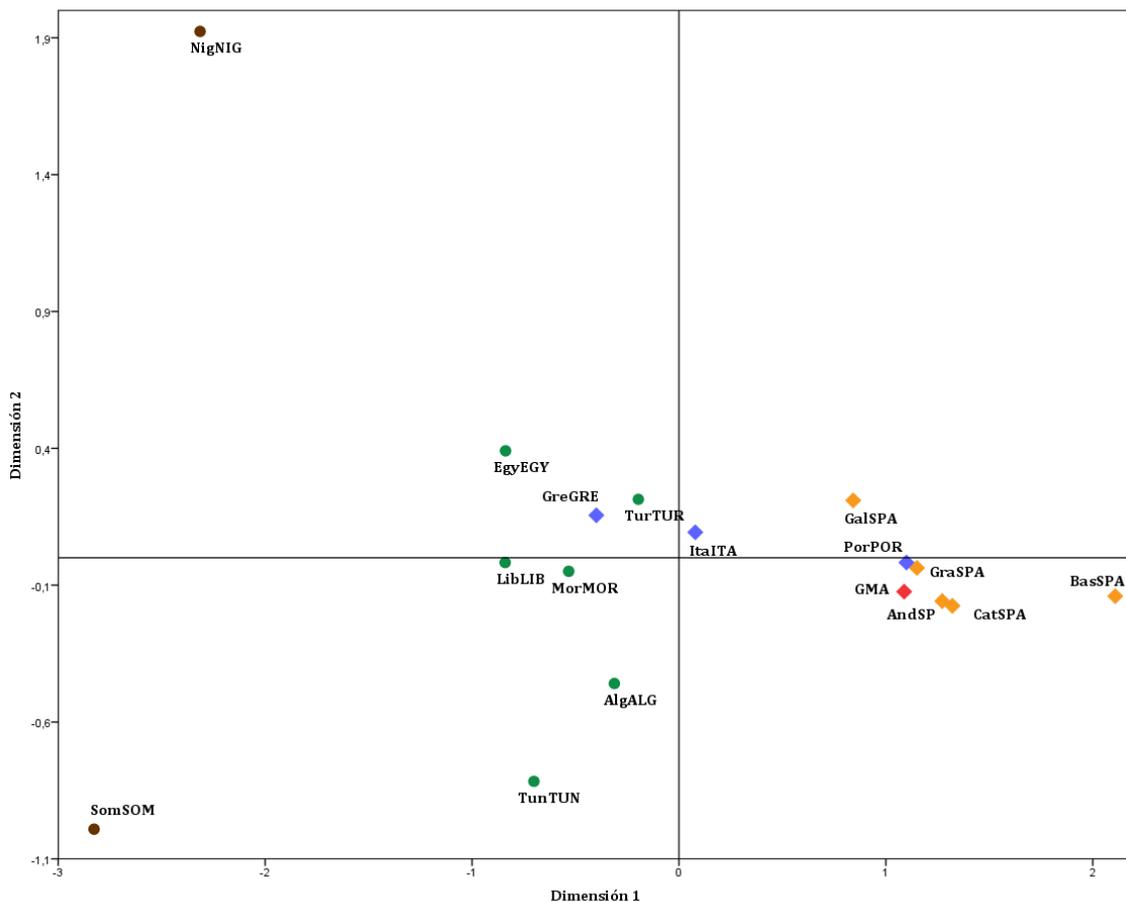


Figura 5.14: MDS calculado a partir de la distancia *Rst*; valor de stress 0,09543.

La dimensión 1 separa claramente las poblaciones africanas (Argelia, Libia, Marruecos, Túnez, Egipto, Nigeria y Somalia), localizadas en el área negativa, de las poblaciones españolas (País Vasco, Cataluña, Galicia, Andalucía, Granada y GMA) localizadas en el área positiva.

Se pueden distinguir claramente tres distribuciones que coinciden con la distribución geográfica de los individuos así como con el origen étnico de éstos. En la izquierda de la figura distinguimos las poblaciones del África subsahariana (Nigeria y Somalia), en el centro podemos encontrar todas las poblaciones del norte de África y muy próximas a ellas las poblaciones de Italia, Grecia y Turquía, y en la derecha de la imagen encontramos todas las poblaciones españolas y entre ellas la población Portuguesa.

3.7 Estudio de los Componentes Ancestrales

Para completar los análisis de frecuencias alélicas y distancias genéticas, se estudió la estructura de la población GMA comparándola con otras poblaciones tomadas de la bibliografía a través del programa STRUCTURE. Las características de las poblaciones empleadas quedan resumidas en la tabla 5.4.

El modelo que mejor explica las diferencias existentes entre las poblaciones estudiadas es el que asume 7 componentes ancestrales ($K=7 \ln P[D]= -10392,46$).

En este análisis, en vez de observar la distribución de los clústeres genéticos en función de la localización geográfica de los individuos como se detecta en el análisis de los STRs autosómicos (Figura 3.7); las diferentes componentes genéticas se identifican con los diferentes haplogrupos encontrados en las poblaciones (Figura 5.15).

El haplogrupo R se encuentra representado por la componente de color amarillo, predominante en las poblaciones europeas. El haplogrupo E se distribuye entre las componentes verde, azul y celeste. Las componentes verde y azul corresponden a subhaplogrupos ampliamente distribuidos entre las poblaciones del norte de África mientras que la componente celeste corresponde a los subhaplogrupos más ancestrales encontrados en poblaciones aisladas del África Subsahariana, como la población Ovambo de Namibia. Las componentes roja y naranja corresponden a subhaplogrupos del haplogrupo J, más concretamente al J1 (naranja) y J2 (roja). La región conocida como la media luna fértil se considera el origen del haplogrupo J y sus principales linajes, J1 y J2 [152]. Estudios sobre el haplogrupo J1 demuestran que se encuentra en frecuencias muy altas en las poblaciones árabe parlantes del norte de África [152,153]. El linaje J2 está ampliamente distribuido en Europa,

sobretudo en la cuenca mediterránea [152]; de hecho la región con mayor componente J2 es Italia (22%). Estos resultados encajan con los encontrados por *Onofri et al.* donde un 19% de la población italiana estudiada pertenece al linaje J2 [154]. Por último, la componente de color rosa corresponde a linajes del haplogrupo G, éste se encuentra en Europa Occidental y el norte de África en frecuencias muy bajas (1-10%) excepto en el Cáucaso e Italia (15-30%) [141].

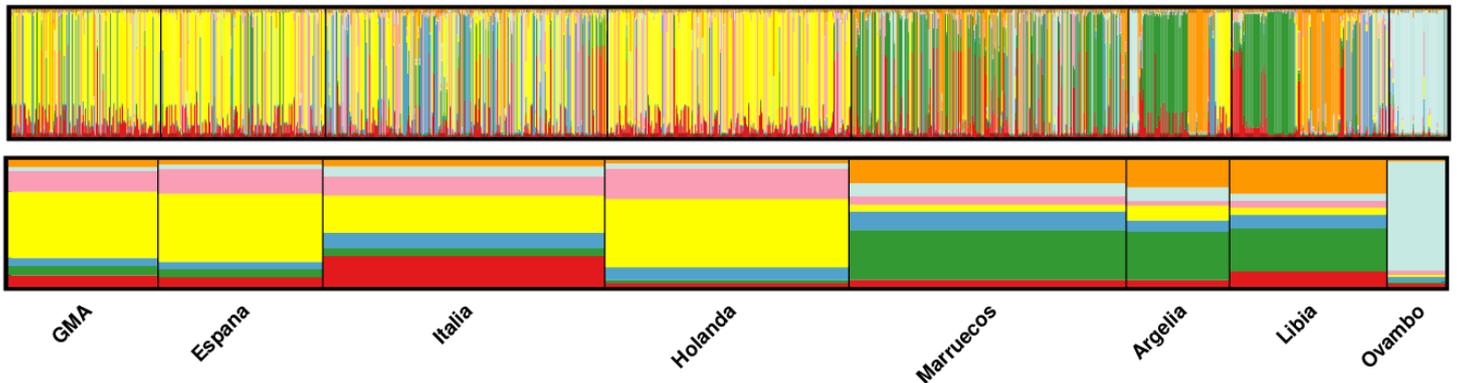


Figura 5.15: Análisis estructural de $K= 7$ para 8 poblaciones. Asignación media de cada individuo a los clústeres resultantes del análisis de estructura. Cada individuo se representa con una fina línea vertical, que está dividida en los colores asignados a cada K representando la pertenencia estimada de cada individuo a cada clúster K (*arriba*). Asignación media de cada población a los clústeres resultantes del análisis de estructura (*abajo*).

En conclusión, los datos genéticos obtenidos a través del estudio de STRs del cromosoma Y revelan que ningún componente africano ha permanecido en la población del Sur de la Península Ibérica, a pesar de haber estado ocupada por éstos durante 800 años. La presencia de haplogrupos típicamente africanos en la población GMA no es significativa cuando se compara con las frecuencias de éstos en poblaciones europeas, tanto mediterráneas como del norte de Europa. La expulsión de los pueblos del norte de África y la repoblación de la zona con gente del resto de la Península fue tan efectiva que es difícil identificar algún trazo del legado genético dejado por los antiguos pobladores. Las similitudes entre la población GMA y la población europea, o más concretamente otras poblaciones españolas, hace que las posibles diferencias existentes sean muy difíciles de identificar.

4. BIBLIOGRAFÍA

- [1] M. Jobling, M. Hurles, C. Tyler-Smith, *Human Evolutionary Genetics: Origins, People and Disease*, Garland Science/Taylor & Francis Group, New York, 2003.
- [2] J.M. Butler, *Forensic DNA typing: biology, technology, and genetics of STR markers*, Academic Press, 2005.
- [3] M.A. Mensah, M.S. Hestand, M.H.D. Larmuseau, M. Isrie, N. Vanderheyden, M. Declercq, et al., Pseudoautosomal region 1 length polymorphism in the human population., *PLoS Genet.* 10 (2014) e1004578. doi:10.1371/journal.pgen.1004578.
- [4] A.G. Hinch, N. Altemose, N. Noor, P. Donnelly, S.R. Myers, Recombination in the human Pseudoautosomal region PAR1., *PLoS Genet.* 10 (2014) e1004503. doi:10.1371/journal.pgen.1004503.
- [5] A. Flaquer, C. Fischer, T.F. Wienker, A new sex-specific genetic map of the human pseudoautosomal regions (PAR1 and PAR2)., *Hum. Hered.* 68 (2009) 192–200. doi:10.1159/000224639.
- [6] R.J. Blaschke, G. Rappold, The pseudoautosomal regions, SHOX and disease, *Curr. Opin. Genet. Dev.* 16 (2006) 233–239. doi:10.1016/j.gde.2006.04.004.
- [7] D. Bachtrog, Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration., *Nat. Rev. Genet.* 14 (2013) 113–24. doi:10.1038/nrg3366.
- [8] B.T. Lahn, Functional Coherence of the Human Y Chromosome, *Science* (80-.). 278 (1997) 675–680. doi:10.1126/science.278.5338.675.
- [9] C. Yao, Z. Wang, Y. Zhou, W. Xu, Q. Li, D. Ma, et al., A study of Y chromosome gene mRNA in human ejaculated spermatozoa., *Mol. Reprod. Dev.* 77 (2010) 158–66. doi:10.1002/mrd.21116.
- [10] B. Charlesworth, The evolution of sex chromosomes, *Science* (80-.). 251 (1991) 1030–1033. doi:10.1126/science.1998119.
- [11] R. Bergero, D. Charlesworth, The evolution of restricted recombination in sex chromosomes, *Trends Ecol. Evol.* 24 (2009) 94–102. doi:10.1016/j.tree.2008.09.010.
- [12] B. Trombetta, F. Cruciani, P.A. Underhill, D. Sellitto, R. Scozzari, Footprints of X-to-Y Gene Conversion in Recent Human Evolution, *Mol. Biol. Evol.* 27 (2010) 714–725. doi:10.1093/molbev/msp231.
- [13] B. Charlesworth, The evolution of chromosomal sex determination and dosage compensation., *Curr. Biol.* 6 (1996) 149–162. doi:10.1016/S0960-9822(02)00448-7.
- [14] M. Kayser, K.N. Ballantyne, Y chromosome in forensic science, *Forensic DNA Appl. An Interdiscip. Perspect.* (n.d.).
- [15] M. Brión, J.J. Sanchez, K. Balogh, C. Thacker, A. Blanco-Verea, C. Børsting, et al., Introduction of a single nucleotide polymorphism-based “Major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages, *Electrophoresis.* 26 (2005) 4411–4420. doi:10.1002/elps.200500293.
- [16] C. Ebensperger, R. Studer, J.T. Epplen, Specific amplification of the ZFY gene to screen sex in man., *Hum. Genet.* 82 (1989) 289–290. doi:10.1007/BF00291174.
- [17] P. Fattorinil, S. Cacció, S. Gustincich, J. Wolfe, B. Maria Altamura, G. Graziosil, Sex determination and species exclusion in forensic samples with probe cY97, *Int. J. Legal Med.* 104 (1991) 247–250. doi:10.1007/BF01369578.
- [18] L. Roewer, J. Arnemann, N.K. Spurr, K.H. Grzeschik, J.T. Epplen, Simple repeat

- sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts., *Hum. Genet.* 89 (1992) 389–394. doi:10.1007/BF00194309.
- [19] L. Roewer, J.T. Epplen, Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work, *Forensic Sci. Int.* 53 (1992) 163–171. doi:10.1016/0379-0738(92)90193-Z.
- [20] P. Vasquez, M. Garde, B. Martínez, Amplificación y tipado de marcadores genéticos moleculares de tipo microsatélites (STR's) autosómicos y del cromosoma Y, *Curso Online Genética Forense. Univ. Zaragoza.* (n.d.).
- [21] S. Rozen, H. Skaletsky, J.D. Marszalek, P.J. Minx, H.S. Cordum, R.H. Waterston, et al., Abundant gene conversion between arms of palindromes in human and ape Y chromosomes, *Nature.* 423 (2003) 873–876.
- [22] H. Skaletsky, T. Kuroda-Kawaguchi, P.J. Minx, H.S. Cordum, L. Hillier, L.G. Brown, et al., The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes, *Nature.* 423 (2003) 825–837.
- [23] E.K. Hanson, J. Ballantyne, Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications, *Leg. Med.* 8 (2006) 110–120. doi:10.1016/j.legalmed.2005.10.001.
- [24] P. Balaesque, T.E. King, E.J. Parkin, E. Heyer, D. Carvalho-Silva, T. Kraaijenbrink, et al., Gene conversion violates the stepwise mutation model for microsatellites in y-chromosomal palindromic repeats., *Hum. Mutat.* 35 (2014) 609–17. doi:10.1002/humu.22542.
- [25] R. Kittler, A. Erler, S. Brauer, M. Stoneking, M. Kayser, Apparent intrachromosomal exchange on the human Y chromosome explained by population history, *Eur. J. Hum. Genet.* 11 (2003) 304–314.
- [26] J.M. Butler, M.C. Kline, A.E. Decker, Addressing Y-chromosome short tandem repeat allele nomenclature, *J. Genet. Geneal.* 4 (2008) 125–148.
- [27] M. Kayser, R. Kittler, A. Erler, M. Hedman, A.C. Lee, A. Mohyuddin, et al., A comprehensive survey of human Y-chromosomal microsatellites, *Am. J. Hum. Genet.* 74 (2004) 1183–1197.
- [28] K.N. Ballantyne, M. Kayser, Additional Y-STRs in Forensics: Why, which and when, in: J. Shewale (Ed.), *Forensic DNA Appl. An Interdiscip. Perspect.*, 1st ed., Taylor & Francis Group, Boca Raton, 2014: pp. 221–245.
- [29] K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, Q. Ayub, et al., Toward male individualization with rapidly mutating y-chromosomal short tandem repeats., *Hum. Mutat.* 35 (2014) 1021–32. doi:10.1002/humu.22599.
- [30] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, et al., A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages., *Forensic Sci. Int. Genet.* 6 (2012) 208–18. doi:10.1016/j.fsigen.2011.04.017.
- [31] K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, et al., Mutability of Y-chromosomal microsatellites: Rates, characteristics, molecular bases, and forensic implications, *Am. J. Hum. Genet.* 87 (2010) 341–353. doi:10.1016/j.ajhg.2010.08.006.
- [32] S. Willuweit, L. Roewer, Y Chromosome Haplotype Reference Database, *Forensic Sci. Int. Genet.* 15 (2013) 43–48. doi:10.1016/j.fsigen.2014.11.024.
- [33] C. Bouakaze, C. Keyser, S. Amory, E. Crubézy, B. Ludes, First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA., *Int. J. Legal Med.* 121 (2007)

- 493–9. doi:10.1007/s00414-007-0177-3.
- [34] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: An evolutionary marker comes of age, *Nat. Rev. Genet.* 4 (2003) 598–612.
- [35] a. Van Geystelen, R. Decorte, M.H.D. Larmuseau, Updating the Y-chromosomal phylogenetic tree for forensic applications based on whole genome SNPs, *Forensic Sci. Int. Genet.* 7 (2013) 573–580. doi:10.1016/j.fsigen.2013.03.010.
- [36] M.A. Jobling, Y-chromosomal SNP haplotype diversity in forensic analysis, *Forensic Sci. Int.* 118 (2001) 158–162.
- [37] P.A. Underhill, T. Kivisild, Use of y chromosome and mitochondrial DNA population structure in tracing human migrations, *Annu. Rev. Genet.* 41 (2007) 539–564.
- [38] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, *Genome Res.* 18 (2008) 830–838.
- [39] M. Van Oven, A. Van Geystelen, M. Kayser, R. Decorte, M.H. Larmuseau, Seeing the wood for the trees: A minimal reference phylogeny for the human Y chromosome, *Hum. Mutat.* 35 (2014) 187–191. doi:10.1002/humu.22468.
- [40] F. Cruciani, B. Trombetta, A. Massaia, G. Destro-Bisol, D. Sellitto, R. Scozzari, A revised root for the human y chromosomal phylogenetic tree: The origin of patrilineal diversity in Africa, *Am. J. Hum. Genet.* 88 (2011) 814–818. doi:10.1016/j.ajhg.2011.05.002.
- [41] M. Mezzavilla, M. Geppert, C. Tyler-Smith, L. Roewer, Y. Xue, Insights into the origin of rare haplogroup C3* Y chromosomes in South America from high-density autosomal SNP genotyping, *Forensic Sci. Int. Genet.* 15 (2015) 115–20. doi:10.1016/j.fsigen.2014.11.005.
- [42] P.A. Underhill, G.D. Poznik, S. Rootsi, M. Järve, A.A. Lin, J. Wang, et al., The phylogenetic and geographic structure of Y-chromosome haplogroup R1a, *Eur. J. Hum. Genet.* 23 (2015) 124–31. doi:10.1038/ejhg.2014.50.
- [43] M. Kayser, A. Caglià, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, et al., Evaluation of Y-chromosomal STRs: A multicenter study, *Int. J. Legal Med.* 110 (1997) 125–133.
- [44] M. Prinz, K. Boll, H. Baum, B. Shaler, Multiplexing of Y chromosome specific STRs and performance for mixed samples, *Forensic Sci. Int.* 85 (1997) 209–218.
- [45] G.B. Barra, T.H. Santa Rita, C.F. Chianca, L.F.R. Velasco, C.F. De Sousa, L.F.A. Nery, et al., Fetal male lineage determination by analysis of Y-chromosome STR haplotype in maternal plasma, *Forensic Sci. Int. Genet.* 15 (2015) 105–110. doi:10.1016/j.fsigen.2014.11.006.
- [46] S.M.B. Maiquilla, J.M. Salvador, G.C. Calacal, M.S. Sagum, M.R.M. Dalet, F.C. Delfin, et al., Y-STR DNA analysis of 154 female child sexual assault cases in the Philippines, *Int. J. Legal Med.* 125 (2011) 817–824. doi:10.1007/s00414-010-0535-4.
- [47] S. Singh, G. Kaur, R.K. Gorea, Y-Chromosome DNA identification of aspermic male offender from vaginal swab of victim, (2015).
- [48] E. Hanson, J. Ballantyne, A Y-short tandem repeat specific DNA enhancement strategy to aid the analysis of late reported (>=6 days) sexual assault cases., *Med. Sci. Law.* 54 (2014) 209–218. doi:10.1177/0025802413519761.
- [49] L. Gusmao, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, et al., DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, *Int. J. Legal Med.* 157 (2005) 187–197. doi:10.1007/s00414-005-0026-1.

- [50] L. Roewer, Y chromosome STR typing in crime casework, *Forensic Sci. Med. Pathol.* 5 (2009) 77–84. doi:10.1007/s12024-009-9089-5.
- [51] M.A. Jobling, C. Tyler-Smith, Fathers and sons: The Y chromosome and human evolution, *Trends Genet.* 6 (1995) 799–803.
- [52] M. Karmin, L. Saag, M. Vicente, M.A.W. Sayres, M. Järve, U.G. Talas, et al., A recent bottleneck of Y chromosome diversity coincides with a global change in culture., *Genome Res.* 25 (2015) 459–466. doi:10.1101/gr.186684.114.
- [53] S.-Z. Gao, Y. Zhang, D. Wei, H.-J. Li, Y.-B. Zhao, Y.-Q. Cui, et al., Ancient DNA reveals a migration of the ancient Di-qiang populations into Xinjiang as early as the early Bronze Age., *Am. J. Phys. Anthropol.* 157 (2015) 71–80. doi:10.1002/ajpa.22690.
- [54] L. Roewer, A. Carracedo, Second Forensic Y chromosome User Workshop, Institute of Legal Medicine (Humboldt-University) in Berlin, Germany 16-17 June, 2000: Foreword, *Forensic Sci. Int.* 118 (2001) 105.
- [55] L. Roewer, M. Krawczak, S. Willuweit, M. Nagy, C. Alves, A. Amorim, et al., Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes, *Forensic Sci. Int.* 118 (2001) 106–113.
- [56] M. Kayser, S. Brauer, S. Willuweit, H. Schädlich, M.A. Batzer, J. Zawacki, et al., Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations, *J. Forensic Sci.* 47 (2002) 513–519.
- [57] R. Lessig, S. Willuweit, M. Krawczak, F.-C. Wu, C.-E. Pu, W. Kim, et al., Asian online Y-STR haplotype reference database, *Leg. Med.* 5 (2003) S160–S163.
- [58] R.I. Richards, G.R. Sutherland, Heritable unstable DNA sequences, *Nat. Genet.* 1 (1992) 7–9.
- [59] A. Pérez-Lezaun, F. Calafell, M. Seielstad, E. Mateu, D. Comas, E. Bosch, et al., Population genetics of Y-chromosome short tandem repeats in humans, *J. Mol. Evol.* 45 (1997) 265–270.
- [60] B.M. Ciminelli, F. Pompei, P. Malaspina, M. Hammer, F. Persichetti, P.F. Pignatti, et al., Recurrent simple tandem repeat mutations during human Y-chromosome radiation in Caucasian subpopulations, *J. Mol. Evol.* 41 (1995) 966–973.
- [61] N. V. Trofimova, S.S. Litvinov, R.I. Khusainova, L.N. Penkin, V.L. Akhmetova, F.S. Akhatova, et al., Genetic characterization of populations of the Volga-Ural region according to the variability of the Y-chromosome, *Russ. J. Genet.* 51 (2015) 108–115. doi:10.1134/S1022795414120138.
- [62] C. Ottoni, M.H.D. Larmuseau, N. Vanderheyden, C. Martínez-Labarga, G. Primativo, G. Biondi, et al., Deep into the roots of the Libyan Tuareg: A genetic survey of their paternal heritage, *Am. J. Phys. Anthropol.* 145 (2011) 118–124. doi:10.1002/ajpa.21473.
- [63] M.A. Jobling, In the name of the father: Surnames and genetics, *Trends Genet.* 17 (2001) 353–357.
- [64] T.E. King, M.A. Jobling, What's in a name? Y chromosomes, surnames and the genetic genealogy revolution, *Trends Genet.* 25 (2009) 351–360.
- [65] M.H.D. Larmuseau, J. Vanoverbeke, G. Gielis, N. Vanderheyden, H.F.M. Larmuseau, R. Decorte, In the name of the migrant father—Analysis of surname origins identifies genetic admixture events undetectable from genealogical records, *Heredity (Edinb)*. 109 (2012) 90–95. doi:10.1038/hdy.2012.17.
- [66] T.R. Rebbeck, C.B. Ambrosone, D.A. Bell, S.J. Chanock, R.B. Hayes, F.F. Kadlubar, et al., SNPs, Haplotypes, and cancer: Applications in molecular epidemiology, *Cancer*

- Epidemiol. Biomarkers Prev. 13 (2004) 681–687.
- [67] W. Kim, T.-K. Yoo, S.-J. Kim, D.-J. Shin, C. Tyler-Smith, H.-J. Jin, et al., Lack of association between Y-chromosomal haplogroups and prostate cancer in the Korean population, *PLoS One*. 2 (2007) e172.
- [68] Y. Sato, T. Shinka, S. Nozawa, M. Yoshiike, E. Koh, J. Kanaya, et al., Y chromosome haplogroup D2a1 is significantly associated with high levels of luteinizing hormone in Japanese men., *Andrology*. (2015). doi:10.1111/andr.12026.
- [69] F.J. Charchar, L.D.S. Bloomer, T. a Barnes, M.J. Cowley, C.P. Nelson, Y. Wang, et al., Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome., *Lancet*. 379 (2012) 915–22. doi:10.1016/S0140-6736(11)61453-0.
- [70] S. Shrestha, M.W. Smith, T.H. Beaty, S.A. Strathdee, Theory and methodology for utilizing genes as biomarkers to determine potential biological mixtures, *Ann. Epidemiol.* 15 (2005) 29–38.
- [71] C. Krausz, L. Quintana-Murci, G. Forti, Y chromosome polymorphisms in medicine, *Ann. Med.* 36 (2004) 573–583.
- [72] S. Lindström, H.O. Adami, J. Adolfsson, F. Wiklund, Y chromosome haplotypes and prostate cancer in Sweden, *Clin. Cancer Res.* 14 (2008) 6712–6716.
- [73] K. Mullis, F. Faloona, S. Scharf, Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction, *Cold Spring Harb. Symp. Quant. Biol.* 51 (1986) 263–273.
- [74] W. Bär, B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, P. Lincoln, et al., DNA recommendations. Further report of the DNA Commission of the ISFG regarding the use of short tandem repeat systems, *Forensic Sci Int.* 87 (1997) 179–184.
- [75] S. Willuweit, L. Roewer, Y chromosome haplotype reference database (YHRD): Update, *Forensic Sci. Int. Genet.* 1 (2007) 83–87.
- [76] L. Excoffier, H. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows , *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [77] A. Tereba, Tools for Analysis of Population Statistics, Profiles DNA. (2001).
- [78] T.W. Athey, Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach, *J. Genet. Geneal.* 2 (2006) 34–39.
- [79] T.W. Athey, Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach, *J. Genet. Geneal.* 1 (2005) 1–7.
- [80] M. Muzzio, V. Ramallo, J.M.B. Motti, M.R. Santos, J.S. López Camelo, G. Bailliet, Software for Y-haplogroup predictions: a word of caution, *Int. J. Legal Med.* (2010) 1–5.
- [81] W. Athey, Comments on the article, “Software for Y Haplogroup Predictions, a Word of Caution,” *Int. J. Legal Med.* (2010) 1–3.
- [82] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies, *Genetics*. 164 (2003) 1567–1587.
- [83] M.J. Hubisz, D. Falush, M. Stephens, J.K. Pritchard, Inferring weak population structure with the assistance of sample group information, *Mol. Ecol. Resour.* 9 (2009) 1322–1332.
- [84] J.B. Kruskal, Nonmetric multidimensional scaling: A numerical method, *Psychometrika*. 29 (1964) 115–129.
- [85] M. Jakobsson, N.A. Rosenberg, CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population

- structure, *Bioinformatics*. 23 (2007) 1801–1806.
- [86] N.A. Rosenberg, DISTRUCT: a program for the graphical display of population structure, *Mol. Ecol. Notes*. 4 (2004) 137–138.
- [87] R. Aboukhalid, M. Bouabdellah, M. Abbassi, K. Bentayebi, M. Elmzibri, D. Squalli, et al., Haplotype frequencies for 17 Y-STR loci (AmpFLSTR®Y-filer™) in a Moroccan population sample, *Forensic Sci. Int. Genet.* 4 (2010) e73–e74.
- [88] L. Palet, C. Coudray, C. Galey, C. Keyser, M. Melhaoui, C. Gagnor, et al., Y-STR genetic diversity in Moroccans from the Figuig oasis, *Forensic Sci. Int. Genet.* 4 (2010).
- [89] C. Robino, F. Crobu, C. Di Gaetano, A. Bekada, S. Benhamamouch, N. Cerutti, et al., Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample, *Int. J. Legal Med.* 122 (2008) 251–255.
- [90] S. Frigi, F. Pereira, L. Pereira, B. Yacoubi, L. Gusmão, C. Alves, et al., Data for Y-chromosome haplotypes defined by 17 STRs (AmpFLSTR® Yfiler™) in two Tunisian Berber communities, *Forensic Sci. Int.* 160 (2006) 80–83.
- [91] B. Arredi, E.S. Poloni, S. Paracchini, T. Zerjal, D.M. Fathallah, M. Makrelouf, et al., A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa, *Am. J. Hum. Genet.* 75 (2004) 338–345. doi:10.1086/423147.
- [92] S. Triki-Fendri, P. Sánchez-Diz, D. Rey-González, I. Ayadi, S. Alfadhli, A. Rebai, et al., Population genetics of 17 Y-STR markers in West Libya (Tripoli region), *Forensic Sci. Int. Genet.* 7 (2013) e59–61. doi:10.1016/j.fsigen.2013.02.002.
- [93] Y.D. Alakoc, O. Gokcumen, A. Tug, T. Gultekin, E. Gulec, T.G. Schurr, Y-chromosome and autosomal STR diversity in four proximate settlements in Central Anatolia, *Forensic Sci. Int. Genet.* 4 (2010).
- [94] C. Nuñez, M. Baeta, M. Fernández, M. Zarrabeitia, B. Martinez-Jarreta, M.M. de Pancorbo, Highly discriminatory capacity of the PowerPlex(®) Y23 System for the study of isolated populations., *Forensic Sci. Int. Genet.* 17 (2015) 104–7. doi:10.1016/j.fsigen.2015.04.005.
- [95] C. Sánchez, C. Barrot, A. Xifró, M. Ortega, I.G. de Aranda, E. Huguet, et al., Haplotype frequencies of 16 Y-chromosome STR loci in the Barcelona metropolitan area population using Y-Filer™ kit, *Forensic Sci. Int.* 172 (2007) 211–217.
- [96] M.L. Pontes, L. Cainé, D. Abrantes, G. Lima, M.F. Pinheiro, Allele frequencies and population data for 17 Y-STR loci (AmpF ϕ STR® Y-filer™) in a Northern Portuguese population sample, *Forensic Sci. Int.* 170 (2007) 62–67. doi:DOI: 10.1016/j.forsciint.2006.04.008.
- [97] S. Turrina, R. Atzei, D. De Leo, Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay, *Int. J. Legal Med.* 120 (2006) 56–59.
- [98] P. Katsaloulis, K. Tsekoura, M. Vouropoulou, P. Miniati, Genetic population study of 11 Y chromosome STR loci in Greece., *Forensic Sci. Int. Genet.* 7 (2013) e56–8. doi:10.1016/j.fsigen.2013.02.001.
- [99] G. Mršić, B. Gršković, A. Vrdoljak, M. Popović, I. Valpotić, Š. Anđelinović, et al., Croatian national reference Y-STR haplotype database, *Mol. Biol. Rep.* 39 (2012) 7727–41. doi:10.1007/s11033-012-1610-3.
- [100] I.S. Veselinovic, D.M. Zgonjanin, M.P. Maletin, O. Stojkovic, M. Djurendic-Brenesel, R.M. Vukovic, et al., Allele frequencies and population data for 17 Y-chromosome STR loci in a Serbian population sample from Vojvodina province, *Forensic Sci. Int.* 176 (2008) e23–e28. doi:DOI: 10.1016/j.forsciint.2007.04.003.
- [101] A. a. Westen, T. Kraaijenbrink, L. Clarisse, L.J.W. Grol, P. Willemse, S.B. Zuniga, et al.,

- Analysis of 36 Y-STR marker units including a concordance study among 2085 Dutch males, *Forensic Sci. Int. Genet.* 14 (2015) 174–181. doi:10.1016/j.fsigen.2014.10.012.
- [102] E. Arroyo-Pardo, L. Gusmão, A.M. López-Parra, C. Baeza, M.S. Mesa, A. Amorim, Genetic variability of 16 Y-chromosome STRs in a sample from Equatorial Guinea (Central Africa), *Forensic Sci. Int.* 149 (2005) 109–113. doi:DOI: 10.1016/j.forsciint.2004.06.006.
- [103] J. Fujihara, I. Yuasa, T. Muro, R. Iida, E. Tsubota, H. Nakamura, et al., Allele frequencies and haplotypes for 28 Y-STRs in Ovambo population, *Leg. Med.* 11 (2009) 205–208. doi:DOI: 10.1016/j.legalmed.2009.03.009.
- [104] S. Elmrgghni, Y.M. Coulson-Thomas, M. Kaddura, R.A. Dixon, D.R. Williams, Population genetic data for 17 Y STR markers from Benghazi (East Libya), *Forensic Sci. Int. Genet.* 6 (2012) 224–7. doi:10.1016/j.fsigen.2011.05.001.
- [105] A. Laouina, B. El Houate, H. Yahia, H. Azeddoug, R. Boulouiz, F. Chbel, Allele frequencies and population data for 17 Y-STR loci (The AmpFI STR® Y-filer™) in Casablanca resident population., *Forensic Sci. Int. Genet.* 5 (2011) e1–3. doi:10.1016/j.fsigen.2010.10.016.
- [106] L. Cherni, L. Pereira, A. Goios, B.Y. Loueslati, H. Khodjet el Khil, I. Gomes, et al., Y-chromosomal STR haplotypes in three ethnic groups and one cosmopolitan population from Tunisia., *Forensic Sci. Int.* 152 (2005) 95–9. doi:10.1016/j.forsciint.2005.02.007.
- [107] I. Ayadi, L. Ammar-Keskes, A. Rebai, Haplotypes for 13 Y-chromosomal STR loci in South Tunisian population (Sfax region), *Forensic Sci. Int.* 164 (2006) 249–53. doi:10.1016/j.forsciint.2005.10.006.
- [108] F. Manni, P. Leonardi, A. Barakat, H. Rouba, E. Heyer, M. Klintschar, et al., Y-chromosome analysis in Egypt suggests a genetic regional continuity in Northeastern Africa., *Hum. Biol.* 74 (2002) 645–58.
- [109] J. Henke, L. Henke, P. Chatthopadhyay, M. Kayser, M. Dülmer, S. Cleef, et al., Application of Y-chromosomal STR haplotypes to forensic genetics., *Croat. Med. J.* 42 (2001) 292–7.
- [110] B. Ambrosio, A. Novelletto, C. Hernandez, J.M. Dugoujon, C. Fortes-Lima, J.N. Rodriguez, et al., Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain), *Forensic Sci. Int. Genet.* 6 (2012) e66–e71. doi:10.1016/j.fsigen.2011.05.007.
- [111] O. García, I. Yurrebaso, I.D. Mancisidor, S. López, S. Alonso, L. Gusmão, Data for 27 Y-chromosome STR loci in the Basque Country autochthonous population., *Forensic Sci. Int. Genet.* (2015). doi:10.1016/j.fsigen.2015.09.010.
- [112] C. Robino, A. Ralf, S. Pasino, M.R. De Marchi, K.N. Ballantyne, A. Barbaro, et al., Development of an Italian RM Y-STR haplotype database: Results of the 2013 GEFI collaborative exercise., *Forensic Sci. Int. Genet.* 15 (2015) 56–63. doi:10.1016/j.fsigen.2014.10.008.
- [113] P. Grignani, G. Peloso, P. Fattorini, C. Previderè, Highly informative Y-chromosomal haplotypes by the addition of three new STRs DYS437, DYS438 and DYS439., *Int. J. Legal Med.* 114 (2000) 125–9.
- [114] S. Presciuttini, A. Caglià, M. Alù, A. Asmundo, L. Buscemi, L. Caenazzo, et al., Y-chromosome haplotypes in Italy: the GEFI collaborative database., *Forensic Sci. Int.* 122 (2001) 184–8.
- [115] C. Robino, S. Inturri, S. Gino, C. Torre, C. Di Gaetano, F. Crobu, et al., Y-chromosomal

- STR haplotypes in Sicily., *Forensic Sci. Int.* 159 (2006) 235–40. doi:10.1016/j.forsciint.2005.05.015.
- [116] G. Ferri, S. Ceccardi, F. Lugaresi, C. Bini, F. Ingravallo, A. Cicognani, et al., Male haplotypes and haplogroups differences between urban (Rimini) and rural area (Valmarecchia) in Romagna region (North Italy)., *Forensic Sci. Int.* 175 (2008) 250–5. doi:10.1016/j.forsciint.2007.06.007.
- [117] N. Cerri, A. Verzeletti, B. Bandera, F. De Ferrari, Population data for 12 Y-chromosome STRs in a sample from Brescia (northern Italy)., *Forensic Sci. Int.* 152 (2005) 83–7. doi:10.1016/j.forsciint.2005.02.006.
- [118] M.E. Ghiani, G. Vona, Y-chromosome-specific microsatellite variation in a population sample from Sardinia (Italy)., *Coll. Antropol.* 26 (2002) 387–401.
- [119] G. Ferri, M. Alù, B. Corradini, E. Radheshi, G. Beduschi, Slow and fast evolving markers typing in Modena males (North Italy)., *Forensic Sci. Int. Genet.* 3 (2009) e31–3. doi:10.1016/j.fsigen.2008.05.004.
- [120] V. Onofri, F. Alessandrini, C. Turchi, B. Fraternali, L. Buscemi, M. Pesaresi, et al., Y-chromosome genetic structure in sub-Apennine populations of Central Italy by SNP and STR analysis., *Int. J. Legal Med.* 121 (2007) 234–7. doi:10.1007/s00414-007-0153-y.
- [121] A. Verzeletti, N. Cerri, F. Gasparini, A. Poglio, E. Mazzeo, F. De Ferrari, Population data for 15 autosomal STRs loci and 12 Y chromosome STRs loci in a population sample from the Sardinia island (Italy)., *Leg. Med. (Tokyo)*. 11 (2009) 37–40. doi:10.1016/j.legalmed.2008.06.003.
- [122] V. Rodríguez, C. Tomàs, J.J. Sánchez, J.A. Castro, M.M. Ramon, A. Barbaro, et al., Genetic sub-structure in western Mediterranean populations revealed by 12 Y-chromosome STR loci., *Int. J. Legal Med.* 123 (2009) 137–41. doi:10.1007/s00414-008-0302-y.
- [123] F. Brisighelli, A. Blanco-Verea, I. Boschi, P. Garagnani, V.L. Pascali, A. Carracedo, et al., Patterns of Y-STR variation in Italy, *Forensic Sci. Int. Genet.* 6 (2012) 834–839. doi:10.1016/j.fsigen.2012.03.003.
- [124] M. Piglionica, S.L. Baldassarra, E. Giardina, A. Stella, F.D. D'Ovidio, P. Frati, et al., Population data for 17 Y-chromosome STRs in a sample from Apulia (Southern Italy)., *Forensic Sci. Int. Genet.* 7 (2013) e3–4. doi:10.1016/j.fsigen.2012.08.003.
- [125] A.T. Fernandes, A. Brehm, L. Gusmão, A. Amorim, Y-chromosome STR haplotypes in the Madeira archipelago population., *Forensic Sci. Int.* 122 (2001) 178–80.
- [126] A. González-Neira, L. Gusmão, M. Brión, M. V Lareu, A. Amorim, A. Carracedo, Distribution of Y-chromosome STR defined haplotypes in Iberia., *Forensic Sci. Int.* 110 (2000) 117–26.
- [127] M. Carvalho, M.-J. Anjos, L. Andrade, V. Lopes, M. V Santos, J.-J. Gamero, et al., Y-chromosome STR haplotypes in two population samples: Azores Islands and Central Portugal, *Forensic Sci. Int.* 134 (2003) 29–35. doi:DOI: 10.1016/S0379-0738(03)00097-5.
- [128] C. Alves, V. Gomes, M.J. Prata, A. Amorim, L. Gusmão, Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFISTR Yfiler) in Portugal, *Forensic Sci. Int.* 171 (2007) 250–255. doi:DOI: 10.1016/j.forsciint.2006.10.022.
- [129] K.S. Parreira, M.V. Lareu, P. Sánchez-Diz, I. Skitsa, A. Carracedo, DNA typing of short tandem repeat loci on Y-chromosome of Greek population., *Forensic Sci. Int.* 126 (2002) 261–4.

- [130] C. Robino, S. Varacalli, S. Gino, A. Chatzikiyriakidou, A. Kouvatsi, C. Triantaphyllidis, et al., Y-chromosomal STR haplotypes in a population sample from continental Greece, and the islands of Crete and Chios., *Forensic Sci. Int.* 145 (2004) 61–4. doi:10.1016/j.forsciint.2004.02.026.
- [131] E. Bosch, F. Calafell, A. González-Neira, C. Flaiz, E. Mateu, H.-G. Scheil, et al., Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns., *Ann. Hum. Genet.* 70 (2006) 459–87. doi:10.1111/j.1469-1809.2005.00251.x.
- [132] L. Kovatsi, J.L. Saunier, J.A. Irwin, Population genetics of Y-chromosome STRs in a population of Northern Greeks., *Forensic Sci. Int. Genet.* 4 (2009) e21–2. doi:10.1016/j.fsigen.2009.01.001.
- [133] C. Hallenberg, K. Nielsen, B. Simonsen, J. Sanchez, N. Morling, Y-chromosome STR haplotypes in Danes., *Forensic Sci. Int.* 155 (2005) 205–10. doi:10.1016/j.forsciint.2004.12.019.
- [134] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, *Genetics*. 164 (2003) 1567–1587. doi:Article.
- [135] L.J. Martínez-González, M. Saiz, M.J. Álvarez-Cubero, A. Gómez-Martín, J.C. Álvarez, C. Martínez-Labarga, et al., Distribution of Y chromosomal STRs loci in Mayan and Mestizo populations from Guatemala, *Forensic Sci. Int. Genet.* 6 (2011) 136–142. doi:DOI: 10.1016/j.fsigen.2011.04.003.
- [136] C. Vullo, V. Gomes, C. Romanini, A.M. Oliveira, O. Rocabado, J. Aquino, et al., Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations, *Int. J. Legal Med.* 129 (2015) 673–680.
- [137] P. Balaesque, G.R. Bowden, S.M. Adams, H.Y. Leung, T.E. King, Z.H. Rosser, et al., A predominantly neolithic origin for European paternal lineages, *PLoS Biol.* 8 (2010). doi:10.1371/journal.pbio.1000285.
- [138] L. Alvarez, C. Santos, R. Montiel, B. Caeiro, A. Baali, J.-M. Dugoujona, et al., Y-chromosome variation in South Iberia: insights into the North African contribution., *Am. J. Hum. Biol.* 21 (2009) 407–9. doi:10.1002/ajhb.20888.
- [139] S.M. Adams, E. Bosch, P.L. Balaesque, S.J. Ballereau, A.C. Lee, E. Arroyo, et al., The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula., *Am. J. Hum. Genet.* 83 (2008) 725–36. doi:10.1016/j.ajhg.2008.11.007.
- [140] A. Kitchen, C. Ehret, S. Assefa, C.J. Mulligan, Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East., *Proc. Biol. Sci.* 276 (2009) 2703–10. doi:10.1098/rspb.2009.0408.
- [141] S. Rootsi, N.M. Myres, A. a Lin, M. Järve, R.J. King, I. Kutuev, et al., Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus, *Eur. J. Hum. Genet.* (2012) 1275–1282. doi:10.1038/ejhg.2012.86.
- [142] K.L. Young, G. Sun, R. Deka, M.H. Crawford, Paternal genetic history of the basque population of Spain., *Hum. Biol. an Int. Rec. Res.* 83 (2011) 455–475. doi:10.3378/027.083.0402.
- [143] N. Rai, G. Chaubey, R. Tamang, A.K. Pathak, V.K. Singh, M. Karmin, et al., The phylogeography of Y-chromosome haplogroup h1a1a-m82 reveals the likely Indian origin of the European Romani populations., *PLoS One.* 7 (2012) e48477. doi:10.1371/journal.pone.0048477.

- [144] F.L. Mendez, T.M. Karafet, T. Krahn, H. Ostrer, H. Soodyall, M.F. Hammer, Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa., *Hum. Biol.* 83 (2011) 39–53. doi:10.3378/027.083.0103.
- [145] A. Sala, D. Corach, Analysis of admixture and genetic structure of two Native American groups of Southern Argentinean Patagonia., *Mol. Biol. Rep.* 41 (2014) 1533–43. doi:10.1007/s11033-013-2999-z.
- [146] V. Battaglia, V. Grugni, U.A. Perego, N. Angerhofer, J.E. Gomez-Palmieri, S.R. Woodward, et al., The first peopling of South America: new evidence from Y-chromosome haplogroup Q., *PLoS One.* 8 (2013) e71390. doi:10.1371/journal.pone.0071390.
- [147] A. Bíró, T. Fehér, G. Bárány, H. Pamjav, Testing Central and Inner Asian admixture among contemporary Hungarians., *Forensic Sci. Int. Genet.* 15 (2015) 121–6. doi:10.1016/j.fsigen.2014.11.007.
- [148] T.E. King, S.J. Ballereau, K.E. Schürer, M.A. Jobling, Genetic signatures of coancestry within surnames, *Curr. Biol.* 16 (2006) 384–388. doi:10.1016/j.cub.2005.12.048.
- [149] R. Calderón, C.L. Hernández, P. Cuesta, J.M. Dugoujon, Surnames and Y-chromosomal markers reveal low relationships in southern Spain, *PLoS One.* 10 (2015). doi:10.1371/journal.pone.0123098.
- [150] T.E. King, M.A. Jobling, Founders, drift, and infidelity: The relationship between y chromosome diversity and patrilineal surnames, *Mol. Biol. Evol.* 26 (2009) 1093–1102. doi:10.1093/molbev/msp022.
- [151] G.R. Bowden, P. Balaesque, T.E. King, Z. Hansen, A.C. Lee, G. Pergl-Wilson, et al., Excavating past population structures by surname-based sampling: The genetic legacy of the vikings in Northwest England, *Mol. Biol. Evol.* 25 (2008) 301–309. doi:10.1093/molbev/msm255.
- [152] M. Regueiro, R. Garcia-Bertrand, K. Fadhlou-Zid, J. Álvarez, R.J. Herrera, From Arabia to Iberia: A Y chromosome prospective, *Gene.* 564 (2015) 141–152. doi:10.1016/j.gene.2015.02.042.
- [153] J. Chiaroni, R.J. King, N.M. Myres, B.M. Henn, A. Ducourneau, M.J. Mitchell, et al., The emergence of Y-chromosome haplogroup J1e among Arabic-speaking populations., *Eur. J. Hum. Genet.* 18 (2010) 348–53. doi:10.1038/ejhg.2009.166.
- [154] V. Onofri, A. Tagliabracci, I. Boschi, F. Brisighelli, F. Scarnicci, V.L. Pascali, et al., Y chromosome J2 subtyping in an Italian sample: Population and forensic implications, *Forensic Sci. Int. Genet. Suppl. Ser.* 1 (2008) 233–234. doi:10.1016/j.fsigs.2007.10.170.
- [155] M. Hammer, A.J. Redd, *Forensic Applications of Y Chromosome STRs and SNPs*, 2006.

ABSTRACT

The genetic data of mitochondrial control region of 191 unrelated donor residents of the provinces of Granada, Málaga and Almería (GMA) were analyzed in order to determine the genetic legacy left by the female inhabitants of the former Kingdom of Granada. A total of 169 haplotypes have been described, 159 of them unique haplotypes. Observed haplotype frequencies and haplogroup frequencies were also analyzed. AMOVA and STRUCTURE analysis show that the populations of the three provinces can be treated genetically as a single population.

Ancestral components are distributed over the populations according to haplogroup designation instead of geographical location. The main haplogroups described in GMA population are European haplogroups (H, 36.13%; U, 17.28% and R, 13.61%). The presence of African haplogroups in GMA population is not relevant when their frequency is compared to those in other European populations.

For better characterization, the population GMA was compared with North African, Iberian Peninsula and South European populations. Analysis of haplogroup frequencies and genetic distances based on mtDNA haplogroup frequencies show that GMA population is located within the Spanish populations group. The analytical results affirm that, owing to the studied genetic markers, the GMA population does not show any characteristics that reflect female genetic influence of North African people.

1. INTRODUCCIÓN

La gran mayoría del genoma humano se localiza dentro del núcleo de las células. Sin embargo, dentro de las mitocondrias se encuentra un pequeño genoma llamado ADN mitocondrial.

El ADN mitocondrial (ADNmt) es el genoma eucariótico mejor estudiado, se conoce su estructura y secuencia completa en casi todos los organismos. En el caso de la especie humana es doble, circular cerrado y con una longitud de 16 569 pares de bases (pb) aproximadamente. Sus dos cadenas son complementarias y antiparalelas, asemejándose más a un ADN bacteriano que al ADN nuclear que está estrechamente asociado con proteínas y repartido en cromosomas (Figura 6.1). La distribución asimétrica de los nucleótidos entre ambas cadenas, da lugar a una cadena “ligera” y una “pesada”. La cadena pesada o H contiene un mayor número de guaninas, que tienen la mayor masa molecular de los cuatro nucleótidos, que la ligera o L [1].

La **región codificante** contiene 37 genes, de los cuales 13 codifican para proteínas de la cadena de transporte de electrones implicadas en la fosforilación oxidativa (OXPHOS) y en la producción de ATP [2], 22 son para ARN de transferencia y 2 para ARN ribosómicos que asisten en la traducción del ARN mitocondrial. De ellos, 28 están codificados en la cadena pesada mientras que la cadena ligera codifica para 8 ARN de transferencia y una enzima llamada ND6 [3].

La regulación de la molécula de ADNmt se sitúa en la **región control** (D-loop) con aproximadamente 1121 pares de bases de longitud, contiene los promotores de transcripción de ambas cadenas y el origen de replicación de la cadena pesada. Ésta se encuentra entre la posición 16 024 y la 16 569 continuando desde la posición 1 hasta la 576. Esta región no codifica para ningún gen por lo que se denomina frecuentemente **región no codificante** (Figura 6.1).

Debido a que no codifica para proteínas, la región control tiene una alta variabilidad entre individuos por lo que se ha convertido en objeto de análisis en estudios antropológicos e investigación en genética forense. Se divide en otras regiones en función de su grado de variabilidad. Las más estudiadas y conocidas por su alto grado de variabilidad son la **región hipervariable 1, HV1**, (16 024-16 365), la **región hipervariable 2, HV2**, (73 a la 340), y una tercera región hipervariable, menos utilizada, **HV3**, (440-560) [4]. Existe una mayor variabilidad en la HV1 que en la HV2, por lo tanto ha sido la más usada en estudios de poblaciones humanas y de la que se tiene una mayor información [5].

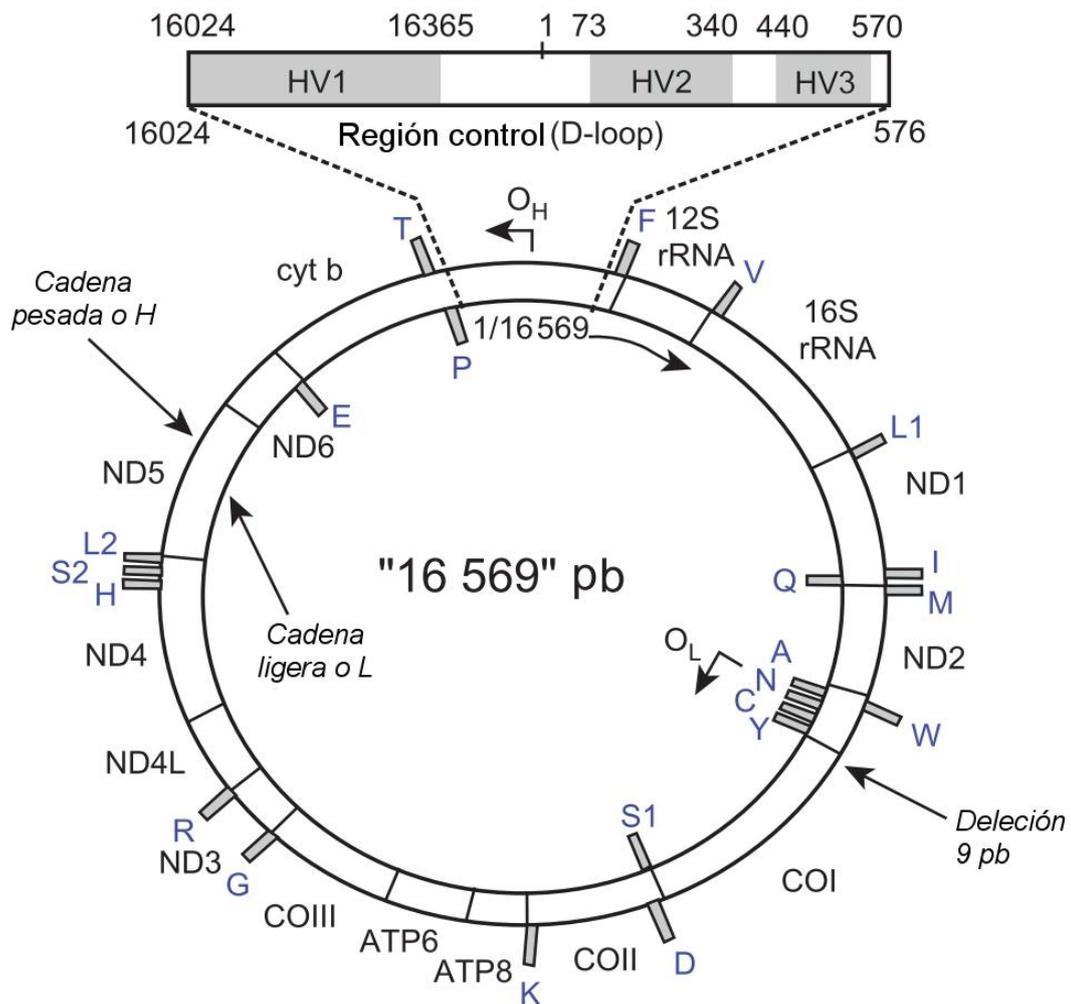


Figura 6.1: Estructura del ADN mitocondrial humano. Se muestran los 37 genes presentes en el genoma mitocondrial. Imagen tomada y modificada de Butler. 2012 [3].

Características del ADN mitocondrial

Este tipo de ADN presenta ciertas características que lo hacen especialmente útil para estudios de identificación y evolución molecular:

- **Alto número de copias en cada individuo:** Hay entre 10 y 100 mitocondrias por célula y dentro de cada mitocondria entre 10 y 100 copias de ADNmt [6,7]. El resultado de este **alto número de copias** es que al intentar obtener ADNmt de una muestra siempre es mucho más sencillo que obtener ADN nuclear, lo que le confiere una gran utilidad en muestras de restos mínimos y/o antiguos.
- **Transmisión exclusiva por vía materna:** el ADNmt de hermanos, de hijos y madre, y en general de aquellos individuos relacionados por vía materna; es idéntico siempre y cuando no se den mutaciones [8-10]. Se ha comprobado que las mitocondrias espermáticas entran dentro del ovocito durante la

fecundación, pero éstas llevan y adquieren una señal de ubiquitina para su degradación selectiva entre el estadio de 4 a 8 células [11].

- **Caracterización completa** de su genoma en 1981 por el grupo de Frederick Sanger en Cambridge [2].
- **Presencia de numerosos polimorfismos.**
- **Rápida tasa de evolución.** El ADNmt tiene una **alta tasa de mutación**, de 5-10 veces mayor que el ADN nuclear, frecuencia estimada de $0,017 \cdot 10^{-6}$ sustituciones por sitio y año. Esta característica hace del ADNmt una herramienta muy útil para el análisis con fines de identificación y también para el estudio del proceso evolutivo [12]. La fácil aparición de mutaciones y la fijación de éstas produce una alta hipervariabilidad entre la población humana.

Tabla 6.1: Comparación entre el ADN nuclear y el ADN mitocondrial humanos [3].

Características	ADN nuclear	ADN mitocondrial
Tamaño del genoma	≈ 3,2 billones pb	≈16 569 pb
Copias por célula	2 (1 alelo de cada progenitor)	>1000
% del total de ADN/célula	99,75%	0,25%
Estructura	Lineal, empaquetado en cromosomas	Circular
Herencia	Ambos progenitores	Materna
Ploidía	Diploide	Haploide
Recombinación generacional	Sí	No
Reparación en la replicación	Sí	No
Exclusividad	Único entre individuos (excepto gemelos monozigóticos)	Idéntico en toda la línea materna
Tasa de mutación	Relativamente baja	5-10 mayor al ADN nuclear
Secuencia de referencia	Descrita en 2001 por el Human Genome Project [13]	Descrita en 1981 por Anderson et al. [2]

1.1. Origen y Evolución del ADN Mitocondrial

La mitocondria es un orgánulo celular donde se genera la mayor parte del ATP mediante el sistema OXPHOS. Hace aproximadamente 2000 millones de años, la tensión de oxígeno ambiental de la atmósfera en la Tierra aumentó bruscamente, lo que sugiere el origen de la mitocondria como orgánulo en los eucariotas primitivos [14].

Según la hipótesis endosimbionte, las mitocondrias provienen de una bacteria “independiente” que se estableció en una célula huésped amitocondriada. Las

evidencias bioquímicas y moleculares indican que las mitocondrias se originaron en un único suceso evolutivo a partir de un ancestro α -proteobacteriano [15,16]. En cualquier genoma mitocondrial, los genes funcionales son un subconjunto que han sido identificados en el genoma mitocondrial menos divergente; y los codones codificantes para proteínas mitocondriales mantienen el orden de los genes homólogos bacterianos [17].

La caracterización de un gran número de genomas mitocondriales ha permitido identificar dos tipos diferentes, *ancestrales* y *derivados* [18]. Los ancestrales han retenido vestigios de un ancestro eubacteriano mientras que los derivados se alejan del patrón ancestral prácticamente sin características primitivas y con una divergencia estructural acompañada por una reducción considerable del tamaño total por una transferencia de material genético del genoma mitocondrial al nuclear, incluyendo genes que codifican para proteínas de la cadena respiratoria y proteínas ribosomales [19].

1.2. Polimorfismos de ADN Mitocondrial

Las sustituciones puntuales son la mayor parte de los polimorfismos del ADNmt, aunque también existen deleciones e inserciones de una o varias pares de bases [20,21].

Para el estudio de la variabilidad mitocondrial humana, el análisis de “RFLPs” en todo el genoma mitocondrial y la secuenciación de ciertas regiones del ADNmt, (HV1 y/o HV2) son la forma más habitual de detectar sustituciones puntuales.

Con la secuenciación de la totalidad o de las regiones con más variabilidad del genoma mitocondrial se encuentran cambios nucleotídicos o mutaciones en la secuencia de bases, lo que define los haplotipos mitocondriales.

Wesley Brown y Douglas Wallace a finales de los años 70, aplican por primera vez el ADNmt como marcador molecular para la reconstrucción del pasado de las poblaciones humanas [12,22]. En los primeros estudios se analizaron numerosas muestras mediante una única enzima de restricción [23] o un número pequeño de éstas mediante varias enzimas [24]. La tendencia en estudios posteriores fue utilizar numerosas enzimas de restricción sobre un mayor número de muestras [25–28].

El desarrollo y generalización de la PCR y el avance de las técnicas de secuenciación posibilitaron una nueva aproximación al estudio molecular de las poblaciones humanas, consistente en la secuenciación de porciones de la molécula del ADNmt. La resolución proporcionada por los nuevos análisis de secuenciación era mucho mayor que la de los análisis de restricción más sofisticados y permitió realizar

una disección más detallada de los tipos mitocondriales individuales. Inicialmente, los análisis se centraron en la región control o D-loop mitocondrial, ampliándose posteriormente a otras regiones no codificantes del genoma mitocondrial.

La región control del ADNmt engloba la gran mayoría de posiciones polimórficas del genoma mitocondrial relevantes para la discriminación de haplotipos en un contexto forense. Hasta hace poco, se han tomado las dos regiones hipervariables como fuente de información en casos forenses. Sin embargo, el uso único de estas dos regiones disminuye el poder de discriminación del ADN mitocondrial frente al análisis de la región control completa. Además, se ha demostrado que hay posiciones fuera de las HVs que aportan gran información filogenética [29].

1.2.1. Nomenclatura

La comisión de ADN de la Sociedad Internacional de Genética Forense ha establecido unas recomendaciones para todos los laboratorios forenses que trabajan con ADNmt para facilitar el intercambio de información entre laboratorios. Además, estas mismas recomendaciones se han seguido en los datos de referencia de ADNmt de las poblaciones.

La Unión Internacional de Química Pura y Aplicada (IUPAC) establece una norma de nomenclatura para la designación de los nucleótidos; G, guanina; C, citosina, A, adenina; y T, timina. También establece la nomenclatura para heteroplasmas. Como el código IUPAC emplea letras mayúsculas, las letras minúsculas se emplean para designar mezclas, posiciones donde hay deleciones y no deleciones de una base así como inserciones y no inserciones. Por ello, siempre se indicará la posición del nucleótido que cambia precedido por el nucleótido de referencia y seguido por el nucleótido mutado. Así, las sustituciones puntuales se nombrarían *A263G*; las deleciones como *C524DEL*, *C524del* o *C524-*; y las inserciones como *-315.1C* [29].

Debido a que se reportan las diferencias entre una muestra problema y una secuencia de referencia, se pueden dar múltiples alineamientos. En el caso de las inserciones en regiones homopoliméricas no se puede generar un alineamiento correcto para establecer entre que posiciones ha tenido lugar la inserción. Por ello, se ha consensado nombrarlos en el extremo 3' del tracto homopolimérico [29].

1.2.2. Heteroplasmia

Una heteroplasmia es la presencia de más de un tipo de ADNmt en un individuo por mutaciones somáticas o heredadas. Se cree que son un estado intermedio en la

fijación de mutaciones en el ADNmt de un individuo y tienden a ocurrir en posiciones con una alta tasa de mutación [30]. La proximidad a la cadena de transporte de electrones, mayor fuente de radicales libres; y la baja protección y los escasos mecanismos de reparación que presenta el ADNmt, son las principales causas de la generación de heteroplasmias.

La heteroplasmia se puede dar a diferentes niveles: individuos con más de un tipo de ADNmt en un mismo tejido; individuos con tipos de ADNmt distintos entre diferentes tejidos y/o; individuos heteroplásmicos en un tejido y homoplásmico en otros tejidos [31]. Incluso, se han observado heteroplasmias dentro de una misma mitocondria [32] (Figura 6.2).

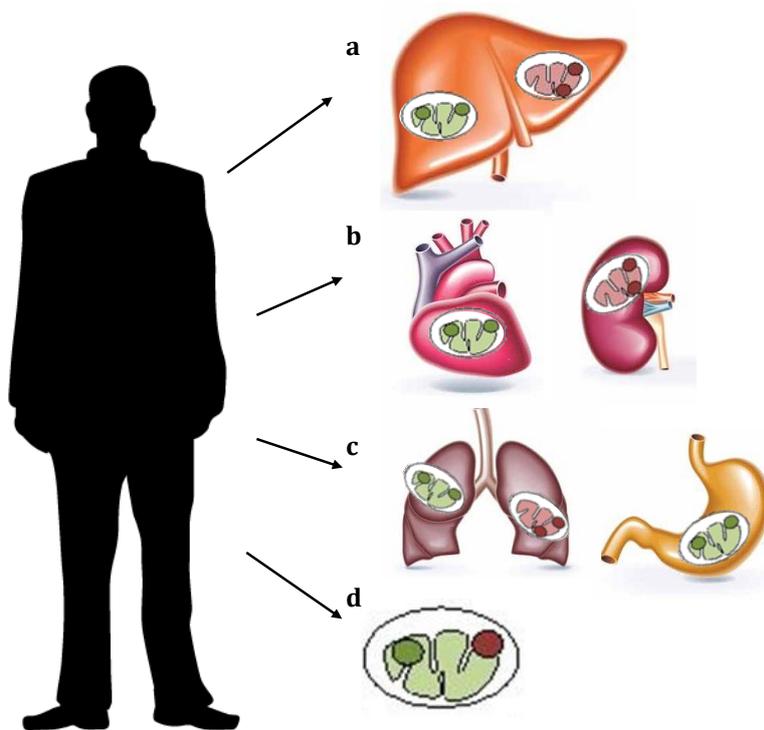


Figura 6.2: Diferentes niveles a los que se puede observar heteroplasmia. *a.* Heteroplasmia dentro de mismo tejido; *b.* heteroplasmia entre tejidos diferentes; *c.* heteroplasmia en unos tejidos y homoplasmia en otros tejidos; *d.* dentro de una misma mitocondria.

Diferentes estudios demuestran que las heteroplasmias son más frecuentes en tejidos con una alta actividad metabólica, hígado [33]; y cuando las moléculas de ADNmt pasa a través de cuellos de botella estrechos durante el desarrollo, cabello [34].

Existen dos tipos de heteroplasmias:

- *Heteroplasmia de secuencia:* Se define por la presencia de dos nucleótidos en una posición del genoma mitocondrial, que se muestran como picos superpuestos en un electroferograma.
- *Heteroplasmia de longitud:* Se manifiesta como la variación en el número de bases existentes en un fragmento. Se da generalmente en los fragmentos

homopoliméricos de policitosinas; en HV1, de las posiciones 16 193 a 19 193 y en HV2, de las posiciones 303 a 310.

Las heteroplasmas no se distribuyen aleatoriamente en el genoma mitocondrial pero se observan en mayor frecuencia en determinadas posiciones, principalmente en la región control [33]. Las heteroplasmas en la región codificante del ADN mitocondrial están implicadas en muchas enfermedades mitocondriales, otras enfermedades, cáncer y envejecimiento. Se ha observado que existe una gran correlación entre la presencia de heteroplasmas en tejidos y la edad del individuo [33].

1.3. Haplogrupos de ADN Mitocondrial

Un haplogrupo es una agrupación de haplotipos que comparten ciertas sustituciones diagnósticas (definidas por enzimas de restricción o por su secuencia) y que presentan un origen común. Los polimorfismos que definen cada haplogrupo se produjeron exclusivamente en las líneas antecesoras de todos los haplotipos que lo integran [35]. Con la aplicación de este tipo de estudios en muestras de diferentes continentes se reveló la existencia de localizaciones geográficas de los haplogrupos, de manera que podían distinguirse haplogrupos exclusivamente africanos, europeos y asiáticos; lo que ha resultado ser de gran utilidad para la inferencia de patrones migratorios dentro y fuera de cada continente (Figura 6.3).

En 2008, se construyó un árbol filogenético actualizado de la variación descrita en el ADNmt basado en 55 publicaciones, *Phylotree* [36]. Desde entonces, este árbol se va actualizando cada seis meses y sirve de referencia para los estudios filogenéticos del ADNmt así como de predictores de haplogrupos mitocondriales. *Phylotree* incluye todos los polimorfismos genéticos presentes en el genoma mitocondrial completo. De esta manera se han descrito 3925 haplogrupos mitocondriales [37].

Los primeros haplogrupos mitocondriales descritos se nombraron como A, B, C y D y fueron en Nativos Americanos [38]. A partir de entonces, los siguientes haplogrupos descubiertos se fueron designando con otras letras del alfabeto. Los clados se dividen en subclados que se designan por su letra correspondiente seguida de un número. Las posteriores subdivisiones se nombran alternando letras minúsculas y números.

Se asume que todos los linajes de ADNmt derivan de un ancestro común materno, *Eva mitocondrial*; que vivió hace aproximadamente 200 000 años en África [39,40]. Este linaje es el denominado haplogrupo L0, que se originó en el este de África y posteriormente se desplazó al sudeste africano. El haplogrupo L1 se originó

directamente del haplogrupo ancestral L0 y dio lugar a los clades L2-L6. Del haplogrupo L3 se originaron todos los haplogrupos no africanos actuales por ser el grupo que caracterizó la salida del hombre moderno de África.

Los haplogrupos N y M son los dos grandes linajes de los que descienden todos los haplogrupos no africanos [41,42].

Los haplogrupos A, B, C y D se encuentran principalmente en poblaciones Nativas Americanas, así como en el norte y este de Asia [43]. Por otro lado, los haplogrupos E y F se encuentran principalmente distribuidos en el sur y el este de Asia.

El haplogrupo R es un haplogrupo antiguo y complejo y prácticamente la mitad del pool genético mitocondrial del oeste europeo se compone de linajes mitocondriales derivados de éste. Una de las líneas descendentes del haplogrupo R, el haplogrupo H, cubre un 40-50% de la variación total en Europa, observándose un descenso de las frecuencias hacia el este y el sur del continente [44,45]. El haplogrupo I se extiende por toda Europa con frecuencias relativamente bajas.

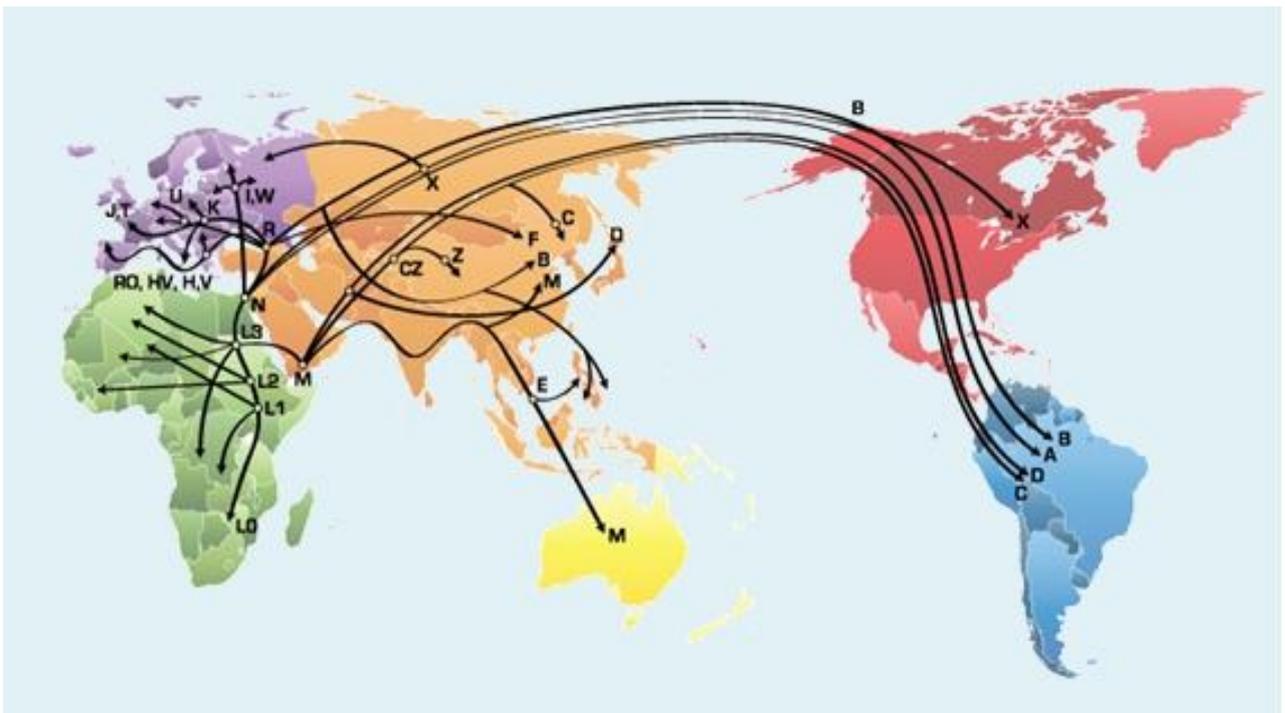


Figura 6.3. Expansión de las poblaciones humanas siguiendo la distribución de haplogrupos de ADNmt. Imagen tomada y modificada de *AncestrybyDNA*.

El haplogrupo J es uno de los cuatro principales haplogrupos específicos Europeos, junto con el H, el U y el V. Se cree que se introdujo en el continente durante el Neolítico por las primeras sociedades agricultoras. El haplogrupo K se originó en el oeste asiático y se extendió por Europa. Es muy característico encontrarlo en poblaciones específicas como los vascos [46].

Por último, el haplogrupo T está ampliamente distribuido en bajas frecuencias. El haplogrupo W se encuentra principalmente en el oeste europeo, aunque se observa en frecuencias leves en toda Eurasia. El haplogrupo X se originó en el Sudoeste asiático y de allí se distribuyó hacia Europa, el norte de África y Oriente Próximo [47]. Los haplogrupos Y y Z son característicos de poblaciones asiáticas, estando sobretodo en poblaciones siberianas.

1.4. Aplicaciones

1.4.1. ADNmt como herramienta en estudios de filiación

El estudio del ADNmt complementa al análisis de polimorfismos del ADN nuclear. Es de gran importancia en el estudio de parentesco entre familiares relacionados por vía materna, cuando no se cuenta con familiares de primer grado y se ha de recurrir al estudio de familiares lejanos [48–50] o bien cuando se ha de realizar la identificación de restos óseos contando con parientes maternos, donde habitualmente la cantidad de ADN nuclear esperada es baja o muy baja (Figura 6.4) [51].

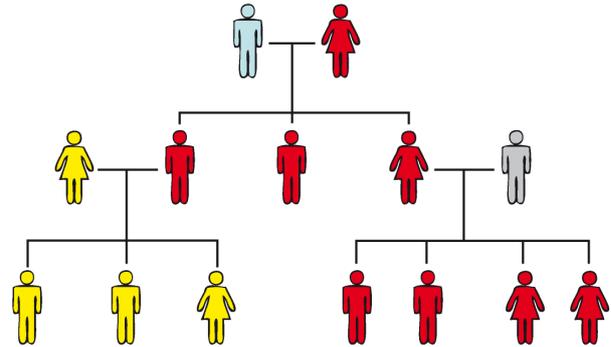


Figura 6.4. Herencia del ADNmt.

1.4.2. ADNmt como herramienta en estudios de criminalística

En ocasiones, el ADNmt es la única información que se puede conseguir, bien sea por la escasa cantidad de la muestra (pelos sin bulbo) o por su altísima degradación. De este modo, cuando un sospechoso posee un haplotipo de ADNmt diferente al de un indicio biológico, hay una buena razón para su exclusión, al igual que ocurre con otros sistemas basados en el tipado de ADN.

La caracterización de un grandísimo número de poblaciones humanas mediante el ADNmt y la inclusión de estos datos en bases de datos poblacionales, facilita en gran medida el uso del mismo en estudios de criminalística. La determinación de los polimorfismos genéticos presentes en el ADNmt de una muestra indubitada y su posterior comparación con las bases de datos disponibles, da la frecuencia de

encontrar ese haplotipo en una determinada población y, por lo tanto, orienta sobre el posible origen geográfico de la muestra.

1.4.3. Estudio de la antropología evolutiva y la historia de las poblaciones

La tasa de mutación observada en el ADNmt hace de éste una muy buena herramienta para la reconstrucción filogenética profunda de las poblaciones a través de las migraciones ancestrales, en vez del estudio de generaciones recientes.

Cuando se realizó la secuenciación del ADNmt por primera vez [2] hubo una gran repercusión en el estudio de las relaciones filogenéticas entre los diferentes grupos humanos actuales. Su herencia y su ausencia de recombinación le permiten preservar la información de los sucesos mutacionales que han surgido en un linaje mitocondrial determinado en las generaciones pasadas.

La caracterización del ADNmt en poblaciones humanas, tanto actuales como antiguas, en otros primates, y en fósiles del género *Homo*, ha contribuido, en las últimas décadas, a aclarar cuestiones centrales de la evolución humana [52]. Asimismo, el ADNmt destaca como uno de los sistemas más utilizados en la caracterización genética de las poblaciones humanas, habiendo sido aplicado como herramienta filogenética [53–57]. La variación en el ADNmt puede ser usada para construir árboles filogenéticos, o varios árboles alternativos dispuestos en una red [58–61], que describen la relación entre secuencias individuales. También es una herramienta muy utilizada en la caracterización genética e inferencia del origen e historia demográfica de numerosas poblaciones antiguas y modernas de los diferentes continentes.

Gracias al estudio del ADNmt de poblaciones de todo el mundo se ha podido determinar el origen de los humanos anatómicamente modernos. Se ha establecido un árbol mitocondrial enraizado en África con unas ramas relativamente cortas implicando a un antecesor común reciente (Eva mitocondrial) [62]. Los datos derivados del ADNmt apoyan un modelo de evolución humana en la que los humanos anatómicamente modernos se desarrollaron en África hace unos 150 000 años y se expandieron fuera de África hace 100 000 años reemplazando totalmente a las otras poblaciones humanas de forma que todas las poblaciones actuales son descendientes de esta población africana. Esta hipótesis queda confirmada ya que se han secuenciado 6 genomas mitocondriales del hombre de Neandertal que han permitido sugerir que no ha habido contribución Neandertal en la diversidad genética mitocondrial de los humanos modernos [63,64].

Además, el análisis del ADNmt permite la reconstrucción de la evolución de los humanos modernos. Las técnicas de secuenciación de nueva generación han revolucionado la investigación en el campo del estudio de las poblaciones humanas a través de ADNmt y ha permitido el análisis de un total de 25 genomas mitocondriales completos de homínidos. Entre ellos se encuentra el genoma mitocondrial del homínido de Denisova que representa la especie humana antigua más inusual. El análisis de su genoma mitocondrial sugiere que el homínido de Denisova representa un grupo hermano al hombre de Neandertal y a los humanos modernos [65,66].

El genoma mitocondrial humano más antiguo analizado hasta la fecha se trata de un fósil del Pleistoceno de Kostenki (Rusia), de data estimada de entre 33 000 y 30 000 años de antigüedad. Los datos derivados sugieren una continuidad en la población Europea durante los últimos 30 000 años [65].

1.4.4. Investigación en biomedicina

Las mutaciones presentes en el genoma mitocondrial están ampliamente relacionadas con múltiples enfermedades, no sólo con enfermedades mitocondriales [67] sino que también con otro tipo de enfermedades. Mutaciones en la región codificante del ADNmt se han asociado a enfermedades como aterosclerosis coronaria [68] o sordera [69]. Además, también se ha establecido una relación entre los haplogrupos mitocondriales y diversas enfermedades. Se ha asociado el haplogrupo N con un mayor riesgo de presentar ciertas características de lupus eritematoso en población Han china [70]. Por otro lado, se ha encontrado una asociación significativa entre el haplogrupo T y la presencia de retinopatía diabética en individuos caucásicos de Europa central [71].

2. MATERIAL Y MÉTODOS

2.1. Población de Estudio

Se han estudiado un total de 191 individuos adultos no relacionados entre sí de la región del sureste de la Península Ibérica [Almería (61), Granada (66), Málaga (64)], ver capítulo 2 páginas 39-40. Todos los donantes fueron informados del propósito de este estudio, aceptaron su participación y firmaron un consentimiento informado según la guía ética de la declaración de Helsinki.

2.2. Análisis Genético Molecular de la Muestra

La toma de muestras, así como su posterior análisis están estrictamente registrados por rigurosos controles, tanto intralaboratorio como interlaboratorio.

2.2.1 Extracción

Se tomaron muestras del epitelio bucal de todos los individuos mediante hisopos estériles de algodón. El ADN genómico se obtuvo mediante un protocolo de extracción orgánica con fenol:cloroformo:alcohol isoamílico (25:24:1) (Apéndice I).

2.2.2 Cuantificación

La estimación de la cantidad y calidad de ADN nuclear de las muestras se ha realizado a través de electroforesis en geles de agarosa al 0,8% (Apéndice I).

2.2.3 Amplificación

Las muestras se amplificaron para la región control completa del ADN mitocondrial (Tabla 6.2 y Figura 6.5).

Para ello se ha amplificado la región control completa, los 1121 pb, en una sola reacción empleando los primers adecuados. Se ha empleado una pareja de primers que han sido proporcionados por Eurofins MWG Operon:

L15997-H619 (1191 pb)

Cuyas secuencias son:

- L15997: 5' CACCATTAGCACCCAAAGCT 3'
- H619: 5' GGTGATGTGAGCCCGTCTAA 3'

Aquellas muestras en las que no se consiguió amplificar región control entera, se estudiaron mediante dos fragmentos de amplificación empleando los primers adecuados:

L15997-H17 (589 pb)

- L15997: 5' CACCATTAGCACCCAAAGCT 3'
- H17: 5' CCCGTGAGTGGTTAATAGGGT 3'

L16555-H619 (633 pb)

- L16555: 5' CCCACACGTTCCCCTTAAAT 3'
- H619: 5' GGTGATGTGAGCCCGTCTAA 3'

Tabla 6.2: Componentes, composición y volumen de la reacción de amplificación.

Reactivo	Componentes	Volumen
PCR Master	25 U Taq DNA Polimerasa en 20 mM Tris-HCl, 100 mM KCl, 3 mM MgCl ₂ , 0,01% Brij 35 (v/v), mix de dNTPs, (dATP, dCTP, dGTP, dTTP) a 0,4 mM	7,5 µl
Primer F	Primer Forward 10 µM	0,3 µl
Primer R	Primer Reverse 10 µM	0,3 µl

La cantidad de ADN por muestra osciló entre 1-2ng/µl y el volumen final de reacción fue de 15µl.

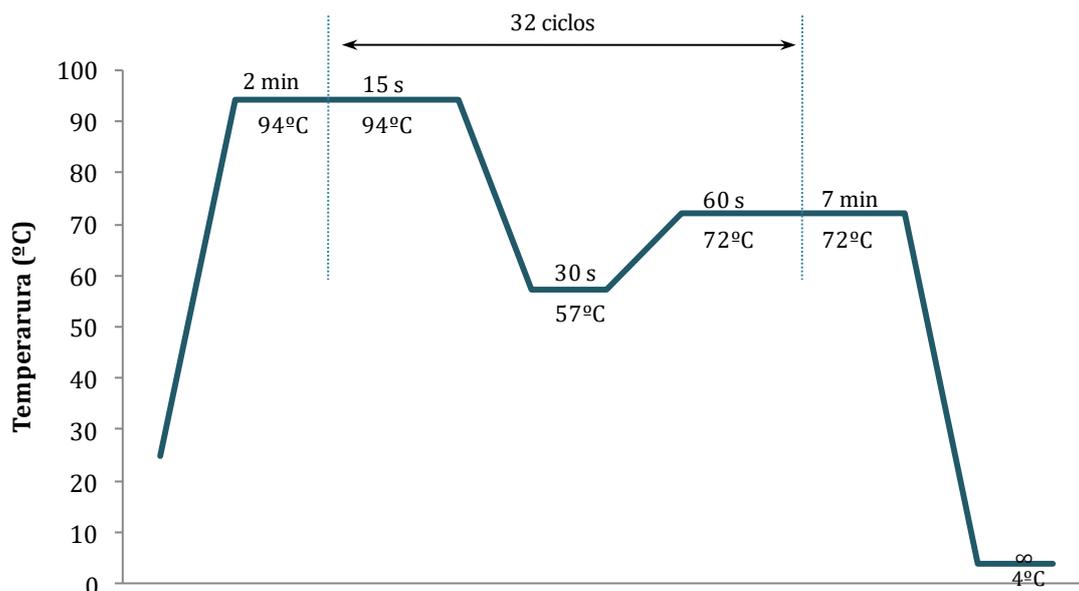


Figura 6.5: Gráfico de las condiciones de amplificación del ADNmt.

En todas las reacciones se han amplificado sistemáticamente dos controles:

- control positivo: verificar que la reacción de amplificación ha sido correcta.
- control negativo: descartar una posible contaminación por ADN extraño.

Una vez finalizada la reacción de amplificación comprobamos la reacción mediante la migración electroforética de los productos amplificados en geles de agarosa al 2%. De esta manera es posible saber si la amplificación ha sido positiva y nos indica si debemos aumentar o disminuir la cantidad de producto amplificado que se va a emplear en posteriores etapas.

2.2.4 Purificación del producto amplificado

Para eliminar todos aquellos nucleótidos (dNTP's) y primers que hayan podido quedar libres en la reacción de amplificación y puedan interferir negativamente en la reacción de secuenciación, se ha de purificar el producto resultante de la amplificación.

Para ello, se emplea un kit comercial, MicroElute® Cycle-Pure Kit (Omega Bio-Tek); diseñado para purificar productos de amplificación en un pequeño volumen de elución (10-15µl). Este protocolo se basa en la unión del ADN a una membrana gracias a un buffer de unión, una serie de lavado con un buffer de lavado y la elución del ADN de la membrana mediante la incubación con un buffer de elución.

2.2.5 Secuenciación

La secuenciación permite determinar el orden de nucleótidos que componen el fragmento de ADN mitocondrial a estudiar. El método empleado es el método de Sanger o de los didesoxinucleótidos. Este método se basa en la incorporación a la cadena de ADN de didesoxidonucleotidos marcados con fluorocromos de diferente color como terminadores. Estos nucleótidos carecen de grupo hidroxilo en su extremo 3' por lo que la polimerasa no puede seguir incorporando nucleótidos a la cadena una vez ha incorporado uno de ellos. En la reacción de secuenciación se añaden una mezcla de nucleótidos no marcados y didesoxinucleótidos que hace que obtengamos fragmentos de distinta longitud de la misma cadena de forma lineal.

El kit empleado para la reacción de secuenciación es BigDye® Terminator v3.1 Cycle Sequencing (Tabla 6.3 y Figura 6.6).

Tabla 6.3: Componentes, composición y volumen de la reacción de secuenciación.

Reactivo	Componentes	Volumen
BigDye® Terminator v3.1	dNTPs/ddNTPs, AmpliTaq DNA Polymerase	0,5 µl
Ready Reaction Mix	FS, MgCl ₂	
5X Sequencing Buffer	Tris-HCL, pH 9.0 y MgCl ₂	2,5 µl
Primer	Primer 1 µM	3 µl
DMSO	Dimetilsulfóxido	0,7 µl
H₂O	Agua estéril	5,5 µl

La cantidad de ADN por muestra fue de 3µl y el volumen final de reacción fue de 12µl.

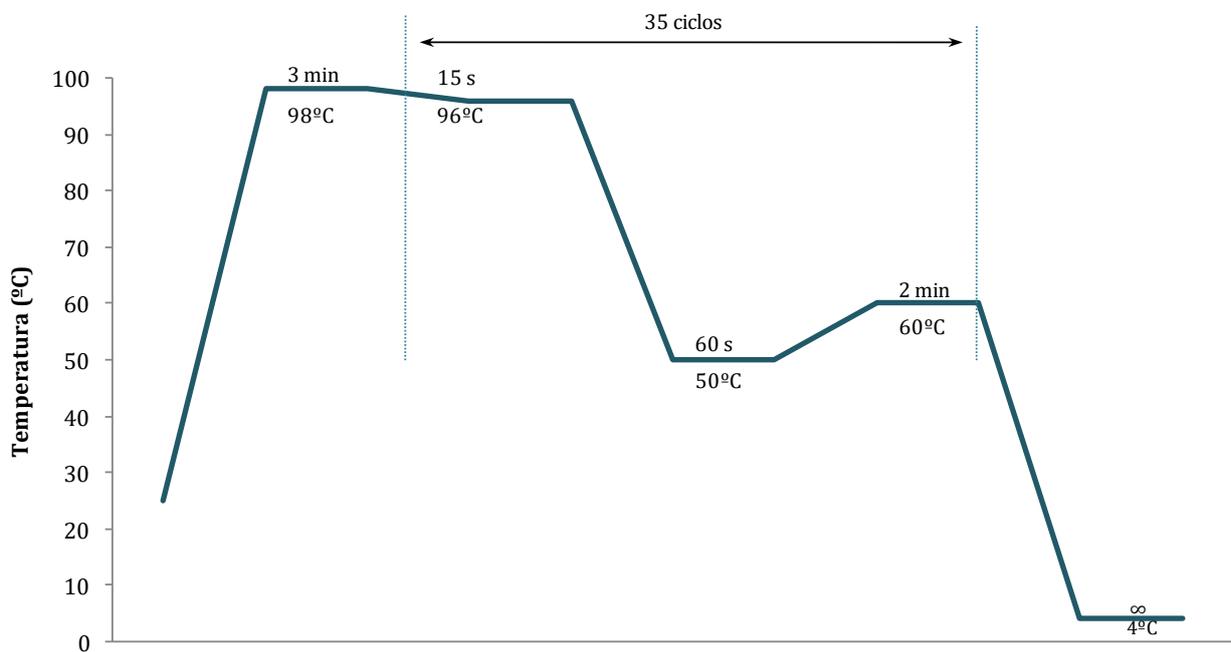


Figura 6.6: Gráfico de las condiciones de secuenciación del ADNmt.

2.2.6 Purificación del producto secuenciado

Para eliminar todos aquellos nucleótidos y terminadores que hayan podido quedar libres en la reacción de secuenciación y puedan interferir en la correcta detección y visualización de la secuencia de ADN, se ha de purificar el producto secuenciado. Existen varios métodos, pero el empleado en este trabajo ha sido la purificación mediante columnas DTR (Dye Terminator Removal). Estas columnas tienen una resina que elimina el 99% de sales, buffers, dNTPs y otros materiales que se unen a la resina [72].

2.2.7 Visualización de los resultados

Las muestras amplificadas fueron visualizadas por electroforesis capilar en un analizador genético ABI Prism 3130 de Applied Biosystems®.

Una vez purificadas las muestras, éstas pueden ser analizadas directamente mediante electroforesis capilar (Apéndice II), sin necesidad de un procesamiento previo.

La electroforesis se llevó a cabo con Performance Optimized Polymer 7 (POP-7™; Life Technologies) y módulo SeqSanger_POP7_50cm_KB.

Las condiciones de electroforesis fueron las siguientes:

- Tiempo de inyección.....30s
- Voltaje de inyección..... 1kV
- Voltaje de electroforesis..... 15kV
- Temperatura de electroforesis..... 50°C
- Tiempo de electroforesis20min

Una vez recogido el dato bruto o “raw data”, el software *Sequencing Analysis* 5.3.1 [73] permite analizar, representar y editar este dato bruto. Mediante este software se asignaron los nucleótidos a cada posición en las secuencias obtenidas. Una vez los datos fueron analizados, el software *SeqScape*® *Software for Mutation Profiling* v2.6 [74] permitió la comparación de las secuencias a través de alineamiento simple con secuencias conocidas. Los resultados se compararon con la región control completa de la secuencia de Anderson [75] para determinar los polimorfismos que presenta cada individuo. Las diferencias respecto a la secuencia de referencia se nombraron de acuerdo con las recomendaciones de la comisión de ADN de la ISFG [29].

Los datos se encuentran en la base de datos EMPOP (*EDNAP Mitochondrial DNA Population Database*) [76] bajo el número de acceso EMP00678. Además, las secuencias completas del ADNmt se encuentran en GenBank con los números de acceso KU686486 - KU686676.

2.3. Análisis Estadístico de la Muestra

Los datos de los perfiles genéticos de los individuos que se prestaron para realizar el estudio poblacional fueron introducidos en tablas mediante un código y su perfil genético.

El cálculo de las frecuencias de los haplotipos encontrados en cada una de las poblaciones se realizó mediante una hoja de cálculo. Del mismo modo se calculó el número de polimorfismos presentes. El número de posiciones polimórficas, transiciones, transversiones e indels; la diversidad nucleotídica y el número de diferencias de emparejamiento (*number of pairwise differences*) se calcularon mediante *Arlequin 3.5.1.3* [77].

Para determinar la subestructura genética de la población del Sudeste de la Península Ibérica, se testó la hipótesis de una distribución aleatoria de los individuos entre parejas de poblaciones mediante un test exacto de diferenciación de la población y el análisis de la varianza molecular (AMOVA) con 10 000 repeticiones experimentales mediante *Arlequin 3.5.1.3*. Además, se utilizó *STRUCTURE 2.3.4* [78,79] para inferir grupos de individuos a partir de los datos genéticos de loci ligados. Para todas las simulaciones y cálculos, se asumieron los modelos de mestizaje y no mestizaje, incluyendo información poblacional y asumiendo un modelo de frecuencias alélicas correlacionadas. Se estableció un periodo de *burn-in* de 50 000 iteraciones seguido de 100 000 iteraciones adicionales, estimando de 1 a 10 contribuciones (*K*). Cada carrera se repitió 10 veces y se calcularon las probabilidades para cada *K* en cada grupo de carreras. Las 10 réplicas para cada *K* se evaluaron con *CLUMPP 1.1.2* (<http://www.stanford.edu/group/rosenberglab/clumpp.html>) [80]. Los grupos combinados resultantes se visualizaron mediante *DISTRUCT 1.1* (<http://www.stanford.edu/group/rosenberglab/distruct.html>) [81].

Los haplogrupos de ADN mitocondrial se calcularon mediante *HaploGrep* [82] basado en *PhyloTree.org* [36]. La aproximación que emplea Haplogrep para determinar el haplogrupo más probable incluye el cálculo del peso filogenético de cada mutación encontrada. Tras contribuir con los haplotipos mitocondriales a la base de datos EMPOP, los administradores de la base de datos recomendaron la reasignación de los haplogrupos de un determinado número de muestras basándose en la estimación de los haplogrupos mitocondriales a través del software EMMA [37]. La determinación de los haplogrupos se realiza a través de una aproximación de máxima probabilidad basada en la estimación del ratio de fluctuación de las mutaciones. En función de este ratio, se determinan unos valores coste de las diferencias entre los perfiles del ADNmt. Los valores más bajos corresponden con los haplogrupos asignados.

Mediante *Network 4.6.1.3* se determinó el ancestro común dentro de la población desde donde se originaron el resto de los haplotipos de la población.

Se llevó a cabo un análisis de correspondencias con el programa *Statistica v9.1* (Statsoft Inc. Tulsa USA) a fin de entender la asociación entre las frecuencias de los haplogrupos y las poblaciones. Las poblaciones empleadas quedan resumidas en la tabla 6.4.

El estudio de distancias genéticas con respecto a otras poblaciones se realizó con distintas poblaciones tomadas de la bibliografía a través del estudio de las frecuencias de los haplogrupos presentes en la población. Las poblaciones escogidas para realizar estos estudios de comparación se muestran en la tabla 6.4. Se determinaron las distancias *Fst*, Reynolds y Slatkin mediante *Arlequin 3.5.1.3*. Para una mejor visualización de los resultados, las distancias genéticas calculadas se representaron mediante escalamiento multidimensional no métrico [83] con *IBM SPSS Statistics 20*.

Tabla 6.4: Poblaciones, abreviatura, situación geográfica, acceso EMPOP y autor del estudio publicado.

POBLACIÓN	Código	País	N	Acceso EMPOP	Bibliografía
ANÁLISIS DE POBLACIONES					
Granada_Málaga_Almería	GMA	España	191	EMP00678	Presente estudio
Marruecos	MorMOR	Marruecos	509	EMP00280-283	Aboukhalid <i>et al.</i> , 2013 [84]
Argelia	AlgALG	Argelia	336	-	Bekada <i>et al.</i> , 2015 [85]
Egipto	EgyEGY	Egipto	277	EMP00022	Saunier <i>et al.</i> , 2009 [86]
Túnez	TunTUN	Túnez	64	EMP00074	Turchi <i>et al.</i> , 2009 [87]
Tuareg	TuaLIB	Libia	129	-	Otoni <i>et al.</i> , 2009 [88]
España	SpaSPA	España	316	-	Álvarez <i>et al.</i> 2007 [89]
Huelva	HueSPA	España	158	-	Hernández <i>et al.</i> , 2014 [90]
País Vasco	BasSPA	España	106	EMP00365	Cardoso <i>et al.</i> , 2012 [91]
Portugal	PorPOR	Portugal	292	EMP00092, EMP00552-554	Marques <i>et al.</i> , 2015 [92]
Italia	ItaITA	Italia	583	-	Brisighelli <i>et al.</i> , 2012 [93]
Holanda	HolHOL	Holanda	680	EMP00666	Chaitanya <i>et al.</i> , 2016 [94]
Estonia	EstEST	Estonia	114	-	Stoljarova <i>et al.</i> 2015 [95]
Somalia	SomSOM	Somalia	192	EMP00397	Mikkelsen <i>et al.</i> 2012 [96]

3. RESULTADOS Y DISCUSIÓN

3.1. Estudio Poblacional

Se caracterizaron 191 individuos de las provincias de Granada, Málaga y Almería, obteniendo los haplotipos para las regiones HV1 y HV2 de 50 de ellos y la región control completa de 141 individuos (Tabla 6.5). Se encontraron un total de 169 haplotipos, de los cuales 159 fueron haplotipos únicos. De los 50 haplotipos estudiados para las regiones HV1 y HV2, se han observado 3 haplotipos en dos ocasiones. De los 141 individuos de los que se tiene la secuencia de la región control completa, se han observado cinco haplotipos en dos ocasiones, un haplotipo 4 veces, otro haplotipo en 5 individuos y otro haplotipo en 7 individuos. La diversidad haplotípica de la región control del ADN mitocondrial en la población es de 0,9973 y la capacidad de discriminación del 94,09% [(Nº Haplotipos/n)·100].

Tabla 6.5: Haplotipos en la población GMA.

Muestra	Hg	N Polimorfismos
16024-16365; 73-340		
GMA09, 41, 45	R0	3 A263G -315.1C
GMA16, 22, 28	R0	3 A263G -309.1C -315.1C
GMA01	J2b1a5	1 C16069T T16126C C16179A C16193T C16278T T16311C A73G C150T T152C A263G C295T -309.1C -315.1C
GMA02	H1c3a	1 C16176T A16219G T146C A257G A263G -315.1C
GMA03	L1b1a+189	1 T16126C C16187T T16189C C16223T C16264T C16270T C16278T T16311C A73G T152C C182T G185T A189G T195C G247A A263G -315.1C
GMA04	HV0	1 T16298C A263G -309.1C -309.2C -315.1C
GMA05	K2a5	1 T16224C T16311C A73G T146C T152C A263G -309.1C -315.1C C324T
GMA06	U5b1b1+152	1 T16189C C16192T C16270T A73G C150T T152C A263G -309.1C -315.1C
GMA07	H7h	1 G16213A A263G -309.1C -315.1C
GMA08	H1aj1	1 C16192T A263G -309.1C -315.1C
GMA10	J1b1a1	1 C16069T T16126C G16145A T16172C C16222T C16261T A73G C242T A263G C295T -309.1C -315.1C
GMA11	U5a1f1	1 C16192T C16256T C16270T T16311C A73G A263G -309.1C -315.1C
GMA12	HV0	1 T16075C T16189C T16298C A200G A263G -309.1C -315.1C
GMA13	L3h1b1a	1 C16179T C16223T T16243C C16256A A16284G T16311C C16320T A73G T152C A189C C194T T195C A263G -309.1C -315.1C
GMA14	R0a	1 T16126C T16209C T16362C A263G -309.1C -315.1C
GMA15	U5a1a1+152	1 A16227G C16256T C16270T A16316G C16320T A73G T152C A263G -309.1C -315.1C

GMA17	H1ag1a	1 C16218T A263G -309.1C -315.1C
GMA18	U5b1f1a	1 C16192T C16270T G16319A A73G C150T A263G -315.1C
GMA19	U5a1b1	1 C16192T C16256T C16270T C16291T A73G A263G -315.1C
GMA20	U3	1 C16260T A16343G A73G C150T A263G -315.1C
GMA21	H1aj1	1 C16192T A263G -309.1C -309.2C -315.1C
GMA23	H1e+16129	1 G16129A T195C A263G -309.1C -315.1C
GMA24	H1bf1	1 C16239T C16242T A263G -309.1C -315.1C
GMA25	H1a1	1 A16162G T16209C C16239T A73G T146C A263G -315.1C
GMA26	H5	1 T16304C A263G -309.1C -315.1C
GMA27	H3x1	1 G16213A A162A93G T152Y A263G
GMA29	HV0	1 A16180G T16189C T16298C A263G -309.1C -309.2C -315.1C
GMA30	H6a1a7	1 A16170G T16362C C150T T239C A263G -309.1C -309.2C -315.1C
GMA31	U6a7a	1 C16169Y T16172C A16219G C16278T A73G T152C A263G -315.1C
GMA32	H1ao	1 C16278T A93G T146C A263G -315.1C
GMA33	H2a1	1 C16354T A263G -309.1C -315.1C
GMA34	H39a	1 C16169T -315.1C
GMA35	H1br	1 A16335G A237G A263G -315.1C
GMA36	U5a1c	1 A16180G C16192T C16256T C16270T C16320T A73G A263G -309.1C -315.1C
GMA37	D1f2	1 C16142T C16223T T16325C T16362C A73G T152C A263G -315.1C
GMA38	M5a1	1 G16129A C16223T C16291T T16298C A73G A263G -309.1C -315.1C
GMA39	J2b1a	1 C16069T T16126C C16193T C16278T A73G C150T T152C A263G C295T -309.1C -315.1C
GMA40	J1c+C16261 +189	1 C16069T T16126C A73G G185A A189G G228A A263G C295T -309.1C -315.1C
GMA42	M65a+@T16311	1 A16289G A263G -309.1C -309.2C -315.1C
GMA43	M18a	1 C16223T A16318T A73G A93G C194T T246C A263G -309.1C -315.1C
GMA44	H1e1a1	1 C16260T T16311C A93G T152C A263G -309.1C -315.1C
GMA46	J2b1a5	1 C16069T T16126C C16193T C16278T T16311C A73G C150T T152C A263G C295T -309.1C -315.1C
GMA47	I3a	1 T16086C G16129A C16223T A73G T152C T199C G207A T239C T250C A263G -309.1C -315.1C
GMA48	U5b	1 C16192T C16270T A73G C150T A189G T195C A263G -315.1C
GMA49	V7a	1 G16153A T16189C T16298C A263G -315.1C
GMA50	H5a3a1	1 T16304C A93G A263G -309.1C -315.1C
16024-576		
GMA58, 122, 131, 174, 176, 201, 220	R0	7 T16519C A263G -309.1C -315.1C
GMA88, 102, 145, 199, 233	R0	5 T16519C A263G -315.1C
GMA65, 66, 74, 78	V+@72	4 C16291T T16298C A263G -309.1C -309.2C -315.1C
GMA114, 138	U3	2 A16343G A73G C150T A263G -315.1C
GMA155, 156	H1a1	2 A16162G T16209C T16519C A73G A263G -315.1C

GMA60, 64	H1b1+16362	2	T16189C T16356C T16362C T16519C A263G -315.1C A523del C524del
GMA77, 165	J2b1a	2	C16069T T16126C C16193T C16278T A73G C150T T152C A263G C295T -309.1C -315.1C T489C
GMA91, 146	J1c2e2	2	C16069T T16126C C16278T C16366T T16519C A73G G185A A188G G228A A263G C295T -309.1C -315.1C C462T T489C A523del C524del
GMA51	H5'36	1	T16189C A263G -309.1C -309.2C -315.1C C456T
GMA53	H7a2	1	C16176T T16519C T152C A263G -309.1C -309.2C -315.1C
GMA54	T2a1b	1	T16126C C16294T C16296T T16324C T16519C A93G A263G -309.1C -309.2C -315.1C
GMA55	H27	1	G16129A A16316G T16519C T63G C64T G66A C150T T195C A263G -315.1C
GMA56	L2a1	1	T16189C C16223T C16278T C16294T A16309G T16362C G16390A A73G T146C C150T T152C T195C A263G -309.1C -315.1C
GMA59	T	1	T16126C C16266T C16294T T16304C T16519C A73G C198T A263G -309.1C -315.1C
GMA61	U4a2	1	T16356C T16519C A73G T195C A263G T310C C315del G499A -524.1A -524.2C -524.3A -524.4C
GMA62	H3x	1	T16368C T16519C C150T T195C A263G -309.1C -309.2C -315.1C
GMA63	U5a2a1+152	1	C16114A C16192T C16256T C16270T C16294T T16311C G16526A A73G T152C A263G -309.1C -315.1C
GMA68	HV0	1	C16291T T16298C T16311C T72C A263G -309.1C -309.2C -315.1C
GMA69	HV0+195	1	A16293T T16298C T72C T195C A263G -309.1C -315.1C
GMA70	R8a1+16093	1	T16093C A16183C T16189C T16249C A16265C T16311C T16519C A73G T195C A263G -309.1C -315.1C
GMA71	U4a2	1	T16356C T16519C A73G T195C A263G T310C G499A -524.1A -524.2C -524.3A -524.4C
GMA72	H20a	1	C16218T A16328A T16362C A249del A263G T292C -315.1C
GMA73	U5a2	1	C16192T C16256T C16270T G16526A A73G A263G -309.1C -315.1C
GMA75	K1a	1	C16278T T16311C T16519C A73G A263G -315.1C C497T
GMA76	H10+(16093)	1	T16093C G16129A T16519C G121C A263G -315.1C
GMA79	H1+T16189	1	A16183C T16189C A16293C T16519C A263G -315.1C
GMA80	U4a2	1	T16356C T16519C A73G T195C A200G A263G T310C -315.1C G499A -524.1A -524.2C -524.3A -524.4C
GMA81	H2a5a1	1	C16291T T16519C A263G -309.1C -315.1C
GMA82	I2a	1	G16129A G16145A C16223T A16312G G16391A T16519C A73G T152C T199C A202G T204C G207A T250C A263G -309.1C -315.1C
GMA83	H5	1	T16304C A263G -309.1C -315.1C C456T
GMA84	H3z	1	T16136C T16519C G207A A263G T293C -309.1C -315.1C
GMA85	V+@72	1	T16298C A263G -309.1C -309.2C -315.1C
GMA86	H3v+16093	1	T16093C T16519C T152C C182T A263G -309.1C -315.1C T408A
GMA87	T2b	1	T16126C C16294T C16296T T16304C T16519C A73G

			A263G -309.1C -315.1C
GMA89	T2c1+146	1	T16126C C16292T C16294T C16296T T16422C T16519C A73G T146C A263G -315.1C C518T A523del C524del
GMA90	HV4b	1	C16069T T16519C T152C A263G -309.1C -315.1C A523del C524del
GMA92	J2a1a1a	1	C16069T T16126C G16145A T16231C C16261T T16519C A73G C150T T152C T195C A215G A263G C295T 310.1T -315.1C T319C T489C G513A
GMA93	H32	1	G16145A T16519C A73G T152C A263G -315.1C
GMA94	H1e2c	1	G16274A T16298C A73G A263G -309.1C -309.2C -315.1C
GMA95	H57	1	T16519C C64T T152C A263G -309.1C -315.1C
GMA97	H1ba	1	C16270T T16519C A263G -315.1C
GMA98	J1c3e	1	C16069T T16126C G16390A A73G G185A G228A A263G C295T -309.1C -315.1C C462T T489C
GMA99	T1	1	T16126C A16163G T16189C C16294T T16519C A73G A263G -315.1C
GMA100	H3z	1	T16519C A263G T293C -315.1C
GMA101	HV0	1	G16129A T16298C T72C A263G -309.1C -309.2C -315.1C -524.1A -524.2C
GMA103	H1e1a1	1	T16189C T16311C T16519C A93G A263G -309.1C -309.2C -315.1C
GMA104	U6a3b	1	T16172C A16182C A16183C T16189C C16223T C16278T T16519C A73G T146C T152C G185A A188G A263G -309.1C -315.1C
GMA106	I	1	G16129A C16223T T16311C G16391A T16519C A73G T199C T204C T250C A263G -309.1C -315.1C -573.1C -573.2C -573.3C -573.4C
GMA107	A2+(64)+16129	1	C16111T G16129A C16223T C16290T G16319A T16362C C64T A73G T146C A153G A235G A263G -309.1C -315.1C A523del C524del
GMA108	H7a2	1	C16176T T16519C T152C A263G -309.1C -315.1C
GMA109	H2a2a	1	A16181G T16519C A263G -315.1C
GMA112	U3b1a	1	T16086C A16335G A16343G A73G C150T A249del A263G -309.1C -315.1C A73G A523del C524del
GMA117	H1e1a1	1	C16169T T16519C A93G A200G A263G T279C -315.1C
GMA119	H1e+16129	1	G16129A T16519C T146C A263G -315.1C
GMA121	H1+T16189	1	A16183C T16189C T16519C A263G -315.1C
GMA124	R0	1	T16519C A263G -309.1C -309.2C -315.1C
GMA128	H3c1	1	C16260T C16262T T16519C A189G A263G -315.1C
GMA129	H6	1	T16093C T16362C A16482G T239C A263G -309.1C -315.1C
GMA130	W	1	C16223T C16292T G16346A T16519C A73G A189G T195C C198T T204C G207A A263G -309.1C -315.1C
GMA132	U5b1b1 +@C16192	1	A16183C T16189C C16270T T16519C A73G C150T A263G -309.1C -309.2C -315.1C
GMA135	HV0	1	T16189C T16298C T16362C T72C A263G -315.1C
GMA139	J1c1	1	C16069T T16126C T16519C A73G G185A G228A A263G C295T -309.1C -315.1C C462T T482C T489C
GMA143	U4a2	1	T16356C T16519C A73G A153G T195C A263G T310C C315del G499A
GMA147	U6a7a	1	T16172C A16219G C16278T A73G T152C A263G -315.1C

GMA148	H11a	1	A16293G T16311C T195C A263G -315.1C -315.2C
GMA149	U5b1e	1	T16172C A16183C T16189C C16270T G16274A T16311C T16325C C16465T T16519C A73G C150T T195C T196C A263G -309.1C -315.1C
GMA150	H1e+16129	1	G16129A T16325C T16519C A263G -315.1C
GMA153	H2a2b	1	A16235G C16291T A263G -309.1C -309.2C -315.1C
GMA154	T1a7	1	T16126C A16163G C16186T T16189C C16193del G16274A C16294T T16519C A73G A263G -309.1C - 309.2C -315.1C A512G
GMA159	HV0	1	A16220C T16298C T16519C T72C A200G A263G -309.1C -315.1C
GMA161	H1e1a4	1	T16311C C16327T T16519C A263G -309.1C -309.2C -315.1C
GMA162	U5b2a1a2	1	G16390A T16519C A73G C150T A263G -315.1C
GMA163	U5a1+@C16192	1	G16129A C16256T C16270T A16399G A73G A263G -315.1C
GMA166	U5a1b1	1	C16192T C16256T C16270T C16291T A16399G A73G A263G -315.1C
GMA167	U4a3a	1	A16265G C16294T T16356C T16362C T16519C A73G T195C G247A A263G -309.1C -315.1C G499A -524.1A -524.2C -524.3A -524.4C -524.5A -524.6C
GMA168	T1a	1	T16126C A16163G C16186T T16189C C16294T T16519C A73G A263G -315.1C
GMA169	R9	1	T16304C A263G -309.1C -315.1C
GMA170	U3	1	C16260T A16343G A73G C150T A263G -315.1C
GMA171	M5a1	1	G16129A C16223T C16291T T16298C T16362C T16519C A73G A263G -309.1C -309.2C -315.1C T489C -524.1A -524.2C
GMA172	X3a	1	A16182C A16183C T16189C C16223T C16278T T16519C A73G T146C A153G C256T A263G -309.1C - 309.2C -315.1C
GMA173	T2b13a	1	A16051G T16126C A16146G C16294T T16304C T16519C A73G T195C A263G -315.1C
GMA175	H1aj1	1	C16192T T16519C A263G -309.1C -309.2C -315.1C
GMA177	H1e+16129	1	G16129A T16519C T195C A263G -309.1C -309.2C -315.1C
GMA178	I	1	G16129A C16223T T16519C A73G T199C T204C T250C A263G -315.1C -573.1C -573.2C
GMA180	H1c3a	1	C16176T A16219G T16519C T146C A257G A263G -309.1C -315.1C
GMA181	L2a1+143	1	C16223T C16278T C16294T A16309G G16390A A73G G143A T152C T195C A263G -309.1C -315.1C A523del C524del
GMA182	H1e+16129	1	G16129A A16183C T16189C T16519C A263G -309.1C -309.2C -315.1C
GMA183	R0+T16189	1	T16189C T16356C T16519C A73G A263G -309.1C -315.1C
GMA184	H1+T16189	1	A16183C T16189C T16519C A263G -309.1C -309.2C -315.1C
GMA185	H1e1a6	1	T16519C C150T A263G -315.1C
GMA186	K1a5a	1	C16278T T16362C T16519C A73G T195C A263G -315.1C C497T
GMA187	R1	1	T16311C A73G A263G -315.1C
GMA188	H10+(16093)	1	T16093C T16362C T239C A263G -309.1C -315.1C
GMA189	H3c1	1	C16260T T16519C -44.1C A263G -315.1C

GMA191	H1e+16129	1 G16129A T16519C A263G -309.1C -315.1C
GMA192	H1e1a1	1 T16304C A93G A263G -309.1C -315.1C
GMA194	U5b1d1a	1 C16270T A73G C150T T195C A263G -315.1C
GMA196	K1a	1 G16129A T16224C T16311C T16519C A73G A263G -315.1C C497T -524.1A -524.2C
GMA197	H1ba	1 A16051G C16270T T16311C T16519C A263G -315.1C
GMA198	U6a	1 T16172C A16219G C16278T C16290T A73G A263G -315.1C
GMA200	V10a	1 C16261T T16298C T16311C T16362C T16519C T72C A263G -309.1C -315.1C
GMA202	HV17	1 C16111T T16519C T152C A263G -315.1C -524.1A -524.2C
GMA203	H3w	1 C16248T T16519C A263G -309.1C -315.1C
GMA204	T2	1 T16126C C16294T C16296T T16519C A73G A263G -309.1C -315.1C
GMA205	W1e1	1 C16223T C16292T C16295T T16519C A73G A189G C194T T195C T204C G207A A263G -309.1C -315.1C
GMA206	K1a5a	1 G16196A T16224C T16311C T16362C T16519C A73G T152C A263G -309.1C -315.1C C497T -524.1A -524.2C
GMA209	H2a5b1	1 A16328A T16362C A249del A263G T292C -309.1C -315.1C A390G
GMA210	R0	1 T16519C A108C G109T A263G -315.1C
GMA212	H1	1 T16209C T16519C A263G -309.1C -315.1C
GMA216	H6	1 T16519C T239C A263G -309.1C -315.1C
GMA217	T2b	1 T16126C C16294T C16296T T16304C T16519C A73G A263G -315.1C
GMA218	U6a'b'd+T16311	1 T16172C A16219G T16311C A73G T152C G207A A263G -309.1C -315.1C
GMA221	K	1 T16224C T16311C T16519C A73G A263G -309.1C -315.1C
GMA225	U5b2b3	1 T16224C C16270T T16362C T16519C A73G C150T A263G -309.1C -315.1C A517T
GMA226	K1+16362	1 T16224C T16311C T16362C T16519C A73G A263G -309.1C -315.1C
GMA227	R0	1 A263G -309.1C -315.1C
GMA228	H3w	1 C16248T C16294T T16519C A263G -315.1C
GMA229	U5b1d1a	1 C16270T A73G C150T A263G -309.1C -315.1C -524.1A -524.2C
GMA231	H1+T16189	1 T16189C T16519C A263G -309.1C -309.2C -315.1C
GMA232	H2q	1 C16188G T16519C A263G -315.1C
GMA234	H1z	1 T16178C A16183C T16189C T16519C A263G -309.1C -309.2C -315.1C C327T
GMA237	H1ao	1 C16278T T16519C A93G T146C A263G -315.1C
GMA238	U4a1	1 C16134T C16266T T16325C T16356C T16519C A73G T152C T195C A263G -315.1C G499A -524.1A -524.2C

Se han identificado un total de 174 diferencias frente a la rCRS, de las cuales 152 se definen como posiciones polimórficas. Se han observado 136 sustituciones, 121 transiciones y 15 transversiones; y 20 indels. La diversidad nucleotídica es de $0,053 \pm 0,027$ y el número de diferencias de emparejamiento es de $9,269 \pm 4,285$.

3.2. Estudio de la Estructura de la Población

Para determinar si el estudio de los polimorfismos de la región control completa del ADNmt permite establecer una subestructura de la población de estudio en las tres provincias de las que proceden las muestras, se realizaron estudios de AMOVA.

El análisis AMOVA no muestra una subestructura genética significativa entre los individuos de las tres provincias de estudio, por lo que podemos agruparlos dentro de la misma población ($p > 0,05$). Cabe destacar que toda la variación determinada a través del estudio de los polimorfismos de la región control completa del ADNmt es la que se observa entre los individuos de una misma población (Tabla 6.6).

Tabla 6.6: Diseño AMOVA y resultados de las tres provincias del Sudeste de la Península Ibérica.

Diseño	Fuente de Variación	d.f	Suma de cuadrados	Componentes de Varianza	Porcentaje de variación
3 Grupos: Grupo 1= Almería Grupo 2= Granada Grupo 3= Málaga	Entre poblaciones	2	8,134	0,01244 Va	0,38
	Dentro de las poblaciones	188	615,772	3,27538 Vb	99,62
	Total	190	623,906	3,28782	
					Índices de fijación
					FST: 0,00378

Como se mencionó en el capítulo anterior, se ha descrito que el uso del software STRUCTURE no es relevante en el análisis de marcadores fuertemente ligados, como los encontrados en el ADN mitocondrial ya que los datos obtenidos por este software muestran resultados muy similares a los obtenidos por el estudio de haplogrupos [97]. Tanto en este estudio como en estudios previos [98] los análisis corroboran los resultados de AMOVA y nos ayudan a entender la estructura genética de la población.

El modelo con mayor valor de probabilidad es K=4 ($\ln P [D] = -2444,78$). Estas 4 componentes se distribuyen de forma similar entre las tres poblaciones. Esto indica que los datos genéticos favorecen un único nodo común para las tres poblaciones estudiadas. Estos resultados se ajustan a las expectativas históricas y

socioculturales ya que comparten el mismo origen además de una gran proximidad geográfica (Figura 6.7).

El estudio profundo del porcentaje de cada componente revela una asociación entre éstas y los haplogrupos encontrados en la población GMA. De esta forma, se puede ver como la componente verde claro (46,75%) caracteriza a los individuos con haplogrupos R, H, HV o V (58,64% de los individuos estudiados). La componente de color naranja (28,92%) caracteriza a los individuos con haplogrupos U o K (20,94%). Por otro lado, la componente azul claro (15,72%) diferencia a los individuos con haplogrupos J o T (11,52%). Por último, se observa la presencia de una cuarta componente (color rojo). Ésta no se distribuye dentro de un grupo de individuos concreto, si no que se detecta en el resto de haplogrupos descritos.

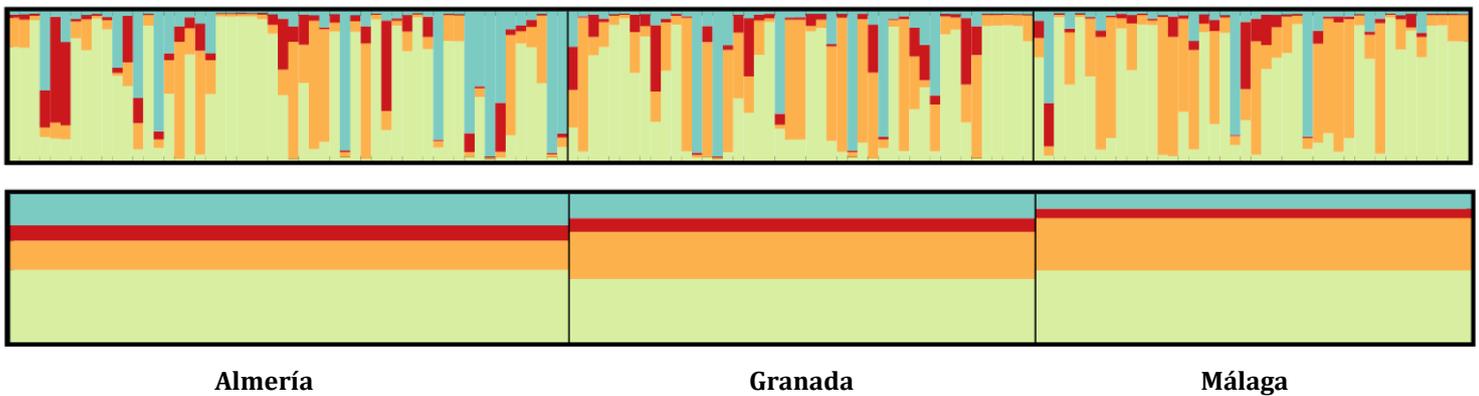


Figura 6.7: Análisis estructural de K= 4 para las 3 poblaciones. Asignación media de cada individuo a los clústeres resultantes del análisis de estructura. Cada individuo se representa con una fina línea vertical, que está dividida en los colores asignados a cada K representando la pertenencia estimada de cada individuo a cada clúster K (*arriba*). Asignación media de cada población a los clústeres resultantes del análisis de estructura (*abajo*).

Existen individuos que presentan una distribución similar de las 4 componentes detectadas que pertenecen a los haplogrupos L, M, W o X. El programa asigna las principales componentes a haplogrupos mucho más modernos, mientras que los haplogrupos de los que surgieron éstos se identifican como una mezcla de todas las componentes encontradas. Esto se debe al bajo número de individuos que pertenecen a los haplogrupos más antiguos (L o M). Si la distribución de los haplogrupos fuese más homogénea, posiblemente se detectaría otro patrón de distribución de componentes ancestrales en la población.

3.3. Estudio de Haplogrupos

Los haplogrupos predichos mediante el programa *HaploGrep* [82] basado en *Phylotree.org* [36] se presentan en la tabla 6.7 de forma más expandida y en la figura 6.8 de manera resumida. De acuerdo con las recomendaciones de los administradores de la base de datos EMPOP, se han reasignado los haplogrupos de 27 individuos que habían sido asignados al haplogrupo H2a2a por *Haplogrep*.

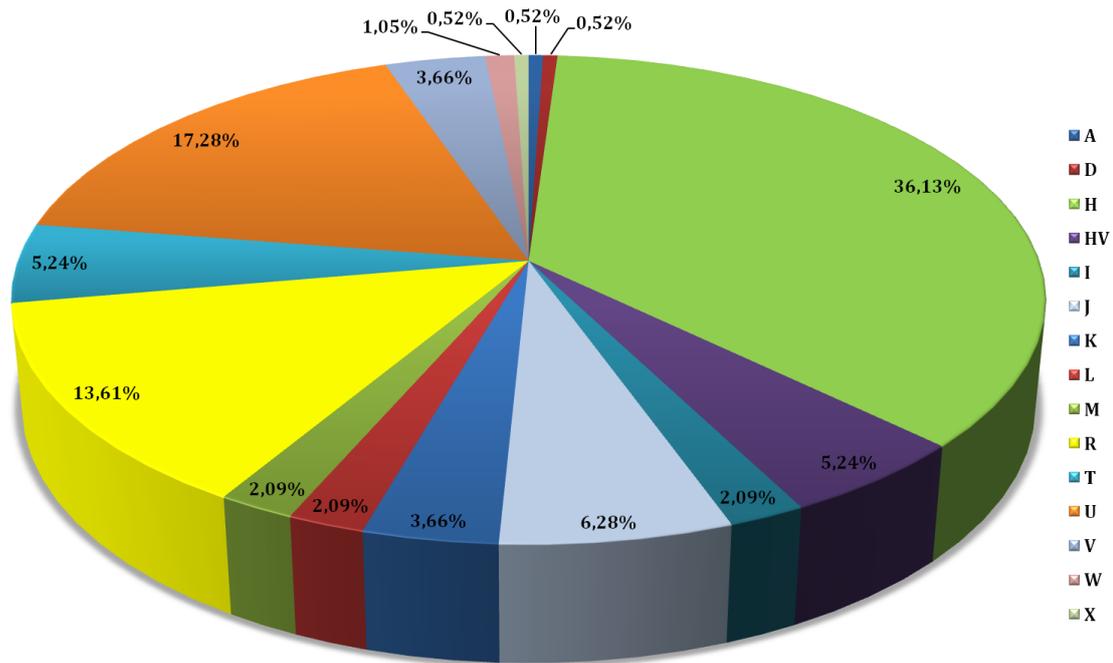


Figura 6.8: Gráfico de sectores de las frecuencias de los haplogrupos en la población GMA.

Tabla 6.7: Frecuencia de los haplogrupos predichos en la población GMA.

A2+(64) +16129	D1f2	H1	H1 +16189	H10 +(16093)	H11a	H1a1	H1ag1a	H1aj1	H1ao
0,0052	0,0052	0,0052	0,0209	0,0105	0,0052	0,0157	0,0052	0,0157	0,0105
H1b1 +16362	H1ba	H1bf1	H1br	H1c3a	H1e +16129	H1e1a1	H1e1a4	H1e1a6	H1e2c
0,0105	0,0105	0,0052	0,0052	0,0105	0,0314	0,0209	0,0052	0,0052	0,0052
H1q	H1z	H20a	H27	H2a1	H2a2b	H2a5a1	H2a5b1	H32	H39a
0,0052	0,0052	0,0052	0,0052	0,0052	0,0052	0,0052	0,0052	0,0052	0,0052
H3c1	H3e	H3v+16093	H3w	H3x	H3x1	H3z	H5	H5'36	H57
0,0105	0,0052	0,0052	0,0105	0,0052	0,0052	0,0105	0,0105	0,0052	0,0052

H5a3a1	H6	H6a1a7	H7a2	H7h	HV0	HV0 +195	HV17	HV4b	I
0,0052	0,0105	0,0052	0,0105	0,0052	0,0366	0,0052	0,0052	0,0052	0,0105
I2a	I3a	J1b1a1	J1c+16261 +189	J1c1	J1c2e2	J1c3e	J2a1a1a	J2b1a	J2b1a5
0,0052	0,0052	0,0052	0,0052	0,0052	0,0105	0,0052	0,0052	0,0157	0,0105
K	K1 +16362	K1a	K1a5a	K2a5	L1b1a +189	L2a1	L2a1+143	L3h1b1a	M18a
0,0052	0,0052	0,0105	0,0105	0,0052	0,0052	0,0052	0,0052	0,0052	0,0052
M5a1	M65a +@16311	R0	R0 +16189	R0a	R1	R8a1 +16093	R9	T	T1
0,0105	0,0052	0,1099	0,0052	0,0052	0,0052	0,0052	0,0052	0,0052	0,0052
T1a	T1a7	T2	T2a1b	T2b	T2b13a	T2c1 +146	U3	U3b1a	U4a1
0,0052	0,0052	0,0052	0,0052	0,0105	0,0052	0,0052	0,0209	0,0052	0,0052
U4a2	U4a3a	U5a1 +@16192	U5a1a1 +152	U5a1b1	U5a1c	U5a1f1	U5a2	U5a2a1 +152	U5b
0,0209	0,0052	0,0052	0,0052	0,0105	0,0052	0,0052	0,0052	0,0052	0,0052
U5b1b1 +@16192	U5b1b1 +152	U5b1d1a	U5b1e	U5b1f1a	U5b2a1a2	U5b2b3	U6a	U6a3b	U6a7a
0,0052	0,0052	0,0105	0,0052	0,0052	0,0052	0,0052	0,0052	0,0052	0,0105
U6a'b'd +16311	V+@72	V10a	V7a	W	W1e1	X3a			
0,0052	0,0262	0,0052	0,0052	0,0052	0,0052	0,0052			

El haplogrupo mayoritario presente en la población GMA es el haplogrupo H (36,13%). Frecuencias similares de este haplogrupo se observan en poblaciones de toda la Península Ibérica, que oscilan desde 59,5% en Galicia a 30% en Cataluña [99] y 32,9% en Huelva [100]. Dentro de este haplogrupo, se distinguen 43 subhaplogrupos (figura 6.9). El más abundante es el H1e+16129 (3,14%) seguido por los haplogrupos H1+16189 y H1e1a1 (2,09%), este último descrito por primera vez en 2012 en la población del País Vasco [101].

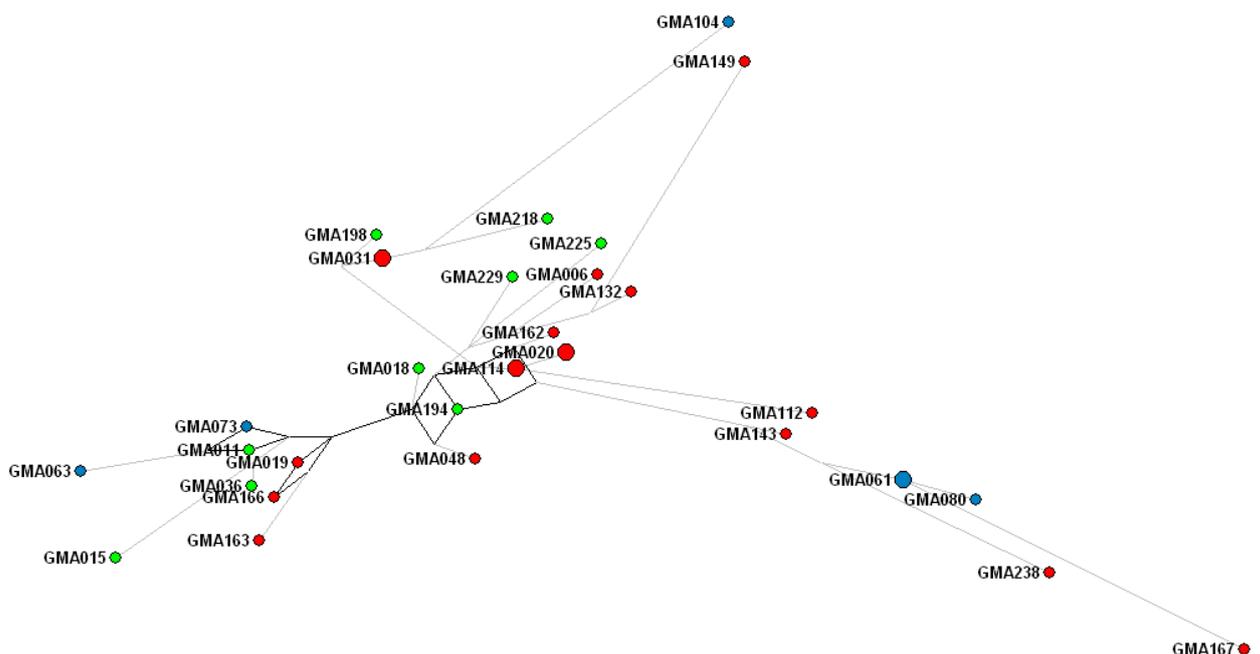
poblaciones del norte de África varían entre un 11% (Bereberes marroquíes) [105] y 6,5% (Argelia) [85]. Sin embargo, sólo está presente en un 2,62% de individuos de la población GMA. Estos datos son similares a los encontrados en otras poblaciones tanto españolas (Galicia 2,20% [106]; Córdoba 0,9% [107]; Zamora 0,50% [108]; País Vasco 0,94% [91]), como de la cuenca mediterránea (Italia 0,86% [93]).

Además, la frecuencia de los haplogrupos mayoritarios en África Subsahariana (L) dentro de la población GMA es del 2,09%. Estos datos son similares a los encontrados en otras poblaciones de la Península Ibérica. Por otro lado, las frecuencias de los linajes del haplogrupo L en las poblaciones del norte de África oscilan entre 47,62 % en la población de Túnez [87] y 20,58% en la población egipcia [86].

3.3.1. Estudio de los haplogrupos mayoritarios en la población GMA

Tras la determinación de los haplogrupos presentes en la población GMA, se realizaron estudios de redes Network para determinar aquellos haplotipos más antiguos. Este estudio también permite comprobar si las tres poblaciones difieren lo suficiente como para separarse por ramas o si se distribuyen uniformemente en el diagrama de redes.

- **U:** Los individuos más antiguos del haplogrupo U son GMA194 y GMA11. Además de estos dos individuos, el torso del haplogrupo se compone por los haplotipos GMA19, GMA73 y GMA166 (Figura 6.12). Los individuos que más mutaciones han acumulado respecto al individuo ancestral son GMA104, GMA149, GMA167 y GMA238.



Todos ellos presentan polimorfismos que han sido descritos en menor frecuencia como G185A, A188G, T196C y un diverso número de inserciones del dinucleótido AC en la posición 524, todas ellas con una frecuencia en MITOMAP [109] menor de 0,003%.

- **H:** El torso del haplogrupo H se caracteriza por la presencia de 4 haplotipos formado por 10 individuos (figura 6.11). Los haplotipos más alejados del nodo central son GMA55 y GMA72 con haplogrupos H27 y H20a respectivamente. Ambos presentan polimorfismos poco frecuentes como T63C, C64T y G66A en el caso de GMA55 y T292C en el caso de GMA72.

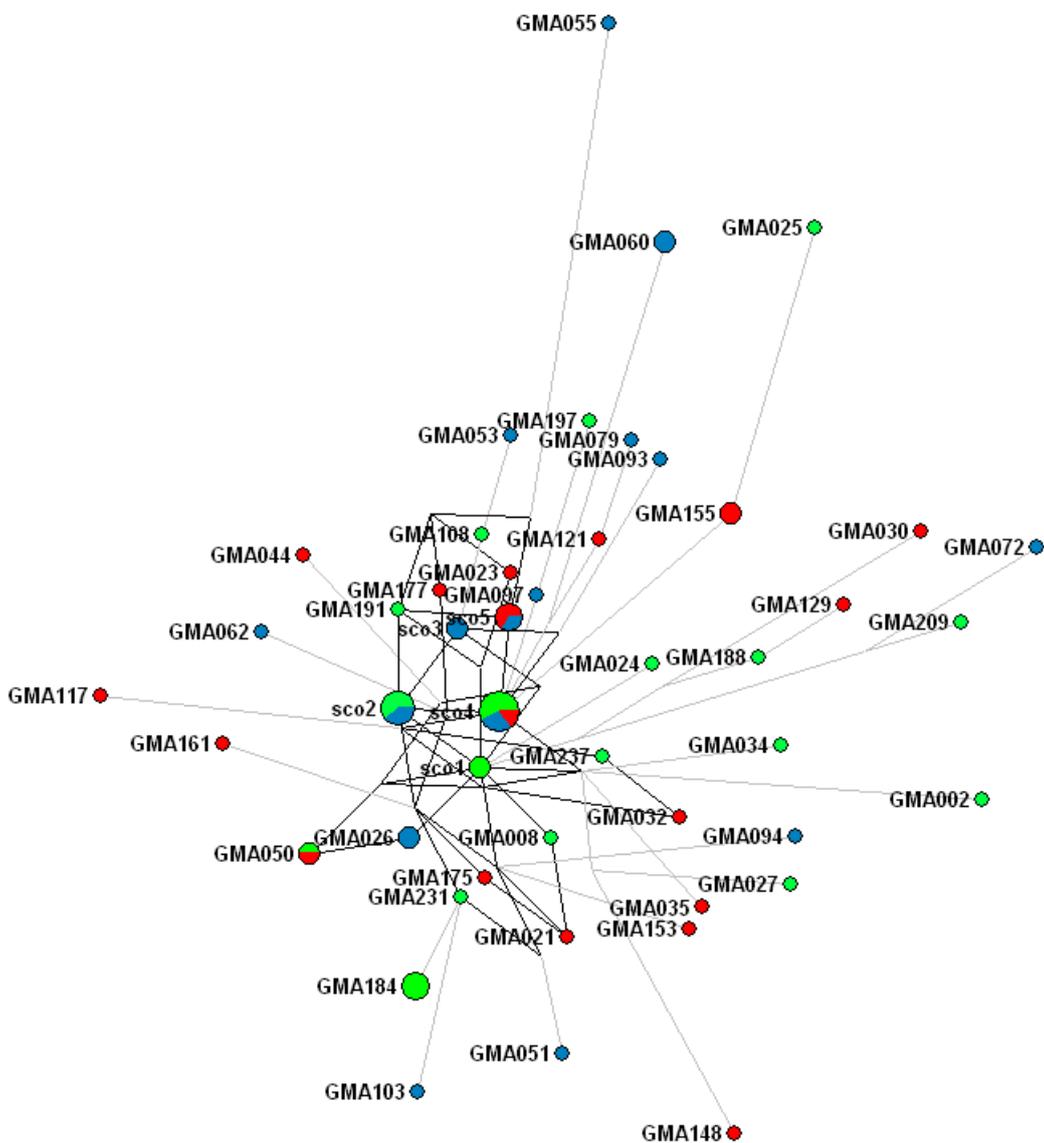


Figura 6.11: Red Network de todos los individuos con haplogrupo H.

3.3.2. Estudio de todos los haplogrupos en la población GMA

Para poder representar mejor todos los haplogrupos presentes en la población GMA, se ha reducido el número de individuos empleados en el estudio de Network. Primero se ha realizado un estudio de los dos haplogrupos mayoritarios en la población (H y U) y se han tomado únicamente aquellos haplotipos más antiguos para el estudio de todos los haplogrupos encontrados. Se han estudiado 83 individuos.

En la figura 6.13 se puede ver la distribución de los individuos en función de su haplogrupo. El torso de la figura está formado por individuos de los haplogrupos H (8 individuos), R0 (4 individuos), HV (4 individuos), K (1 individuo), T (3 individuos) y V (5 individuos). Primero se forma el grupo compuesto por 3 individuos R0 y 4 individuos H y a continuación aparece el individuo del haplogrupo K. Por otro lado, los individuos que más mutaciones han acumulado con respecto al nodo central pertenecen a los haplogrupos A2 (GMA107), L1b (GMA03) y J2a (GMA92). Estos resultados son esperables ya que el haplogrupo A es típico de poblaciones nativas americanas y se encuentra prácticamente restringido al continente Americano [43]. Por otro lado, como ya hemos comentado anteriormente; el haplogrupo L1 desciende directamente del haplogrupo L0, del que se originaron el resto de haplogrupos mitocondriales [39]. Además, aunque el haplogrupo J2a es exclusivo del continente Europeo, las frecuencias más altas de éste se observan en el norte de Europa [110].

Igual que se observa en el estudio de redes de los haplogrupos del cromosoma Y (figura 5.9), los individuos con un mismo haplogrupo mitocondrial también se agrupan dentro de las mismas ramas y cercanos entre ellos. Al contrario de lo que cabría esperar según la historia evolutiva de los haplogrupos, los más antiguos se encuentran muy alejados del nodo central, formado por individuos de haplogrupos H, R, HV y K, mucho más modernos. Esto se debe a que los estudios de Network analizan los haplotipos encontrados en la población intentando agrupar los que presentan más similitudes en el nodo central.

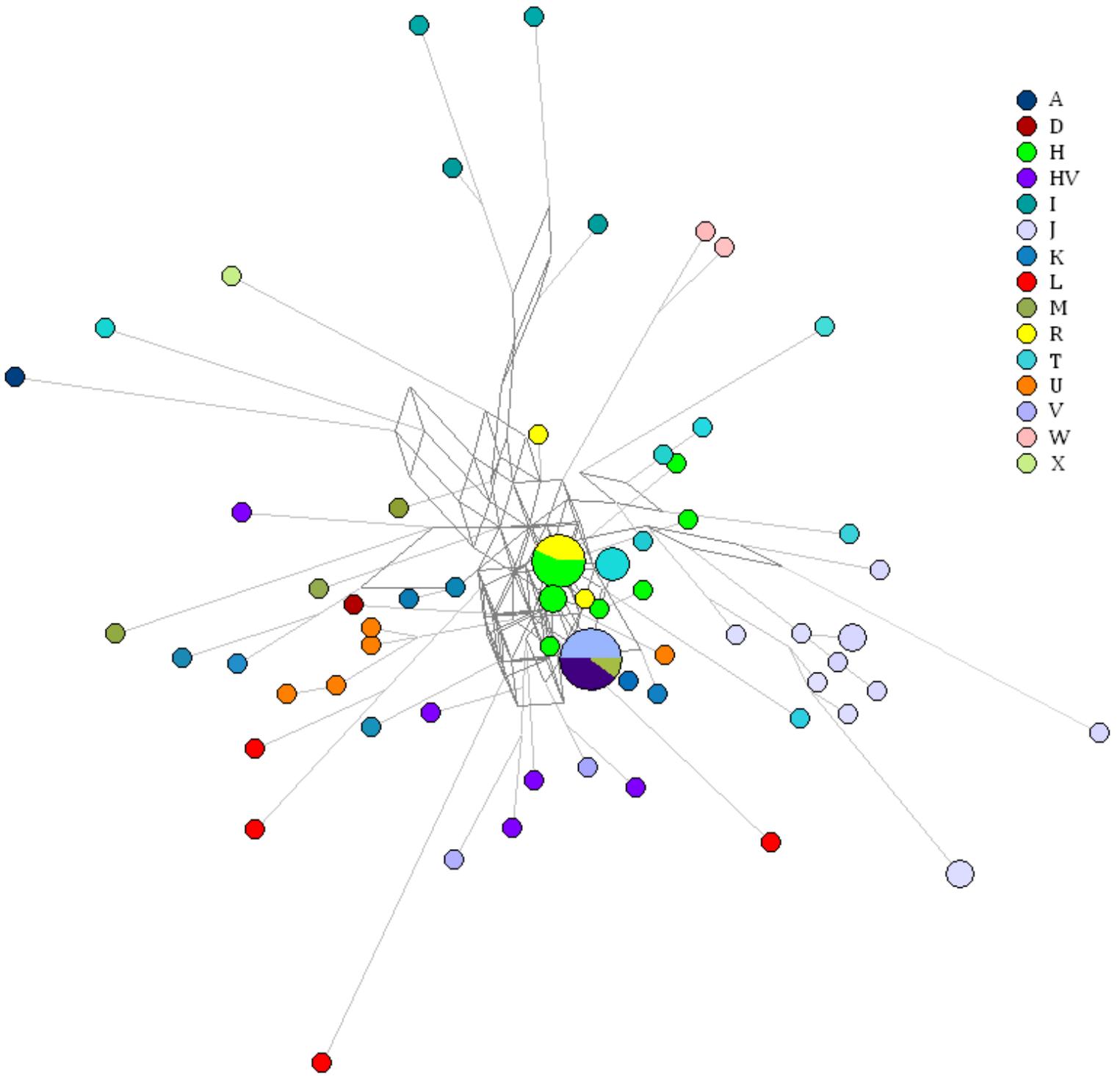


Figura 6.13: Red Network de todos los haplogrupos presentes en la población GMA.

3.4. Estudio de Frecuencias de Haplogrupos

Se realizaron estudios de análisis de correspondencias a través de las frecuencias de los haplogrupos de las poblaciones españolas y del norte de África con Statistica v9.1.

En la figura 6.14 se observan claramente como el eje de coordenadas X separa poblaciones según su origen. En el área positiva se agrupan las poblaciones europeas y en el área negativa se encuentran poblaciones africanas. Esto se debe a las diferencias encontradas entre las frecuencias de los haplogrupos. En las poblaciones europeas se observan altas frecuencias para los haplogrupos típicamente europeos (H, U y J), mientras que las poblaciones africanas presentan mayores frecuencias para haplogrupos africanos (L y M).

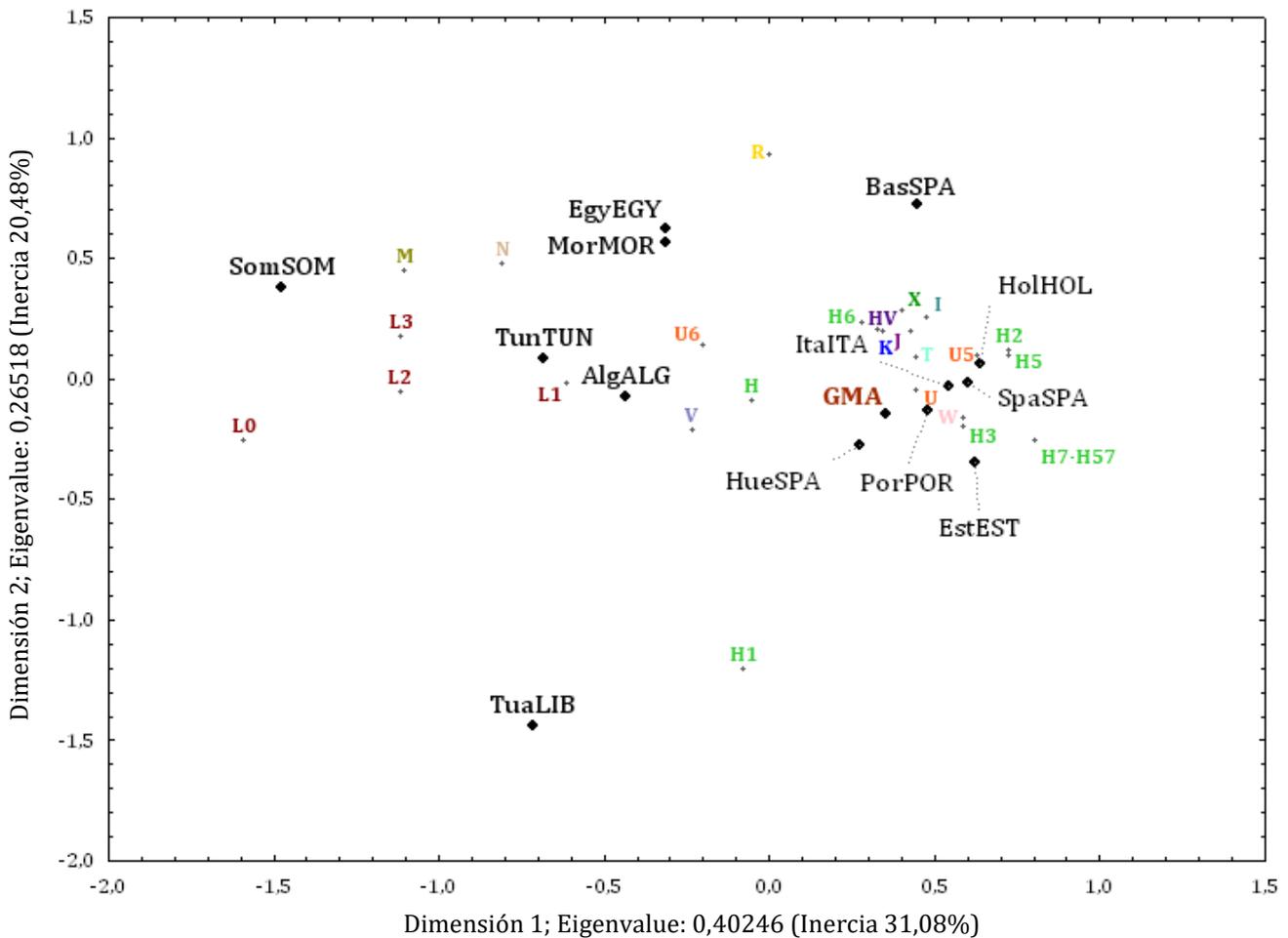


Figura 6.14: Análisis de correspondencia en 2 dimensiones que establece la relación entre las frecuencias de los haplogrupos y las poblaciones estudiadas.

Dentro del grupo de poblaciones europeas cabe destacar la ligera distancia que existe entre la población del País Vasco (BasSPA) y el resto de poblaciones. Esto se debe a la frecuencia del haplogrupo R, en concreto el haplogrupo R0 (37,74%) [91].

Por otro lado, se observa dentro de las poblaciones africanas un grupo central formado por las poblaciones del norte de África [(EgyEGY), (MorMOR), (TunTUN) y (AlgALG)]. Sin embargo, la población tuareg de Libia (TuaLIB) que, aún estando geográficamente situada en el norte de África, se aleja del grupo de poblaciones norafricanas. Esto se debe a que este grupo poblacional presenta dos componentes mitocondriales mayoritarios. Posee una fuerte influencia Europea determinada por las frecuencias de los haplogrupos H1 (61,24%) y V. Además, presenta un componente típicamente subsahariano presente por una mayor frecuencia del linaje L0 (6,20%), comparado con el resto de poblaciones del norte de África estudiadas [88].

Por último, es interesante comentar la posición que tiene la población somalí (SomSOM) en la figura. Somalia se localiza en el cuerno de África, donde de acuerdo con la teoría de la “Salida de África” se inició la dispersión de los humanos anatómicamente modernos hacia el resto del mundo. Como claramente indica la figura 6.13, los haplogrupos que mayores influencias ejercen sobre esta población son el L0, L2, L3 (58,69%) y M (15,796%) [96].

3.5. Estudio de Distancias Genéticas

Las distancias genéticas *de Reynolds*, *Slatkin* y *Fst* se calcularon mediante *Arlequin 3.5.1.3* partiendo de las frecuencias de los distintos haplogrupos mitocondriales; para descifrar las relaciones genéticas entre la población GMA y 13 poblaciones tomadas de la literatura (Tabla 6.4).

Para una mejor visualización de los resultados, las distancias genéticas calculadas se representaron mediante escalamiento multidimensional no métrico con *IBM SPSS Statistics 20* (Figura 6.15).

En este caso, la dimensión 2 separa claramente las poblaciones europeas, localizadas en el área positiva (GMA, Huelva, País Vasco, España, Portugal, Italia, Estonia y Holanda) de las poblaciones africanas localizadas en el área negativa (Egipto, Marruecos, Túnez, Argelia, Tuaregs de Libia y Somalia).

Al igual que ocurre en el análisis de correspondencias, la población del País Vasco se separa del resto de poblaciones españolas debido al altísimo porcentaje de haplotipos pertenecientes al haplogrupo R0 [91]. Por otro lado, la población

holandesa presenta un gran número de haplotipos pertenecientes al haplogrupo HV, J1c y K1a [94] comparada con el resto de poblaciones europeas estudiadas.

Como ya hemos comentado anteriormente, la población tuareg de Libia se aleja del grupo de poblaciones del norte de África por su alta frecuencia del haplogrupo H1. Éste se extendió desde la Península Ibérica tras la Última Gran Glaciación hacia el centro y norte de Europa [44], así como hacia el noroeste de África [105]. Por otro lado, la población somalí se aleja del grupo de poblaciones africanas debido a su origen subsahariano, con una gran proporción de individuos pertenecientes a haplogrupos como el L0, L2 y L3.

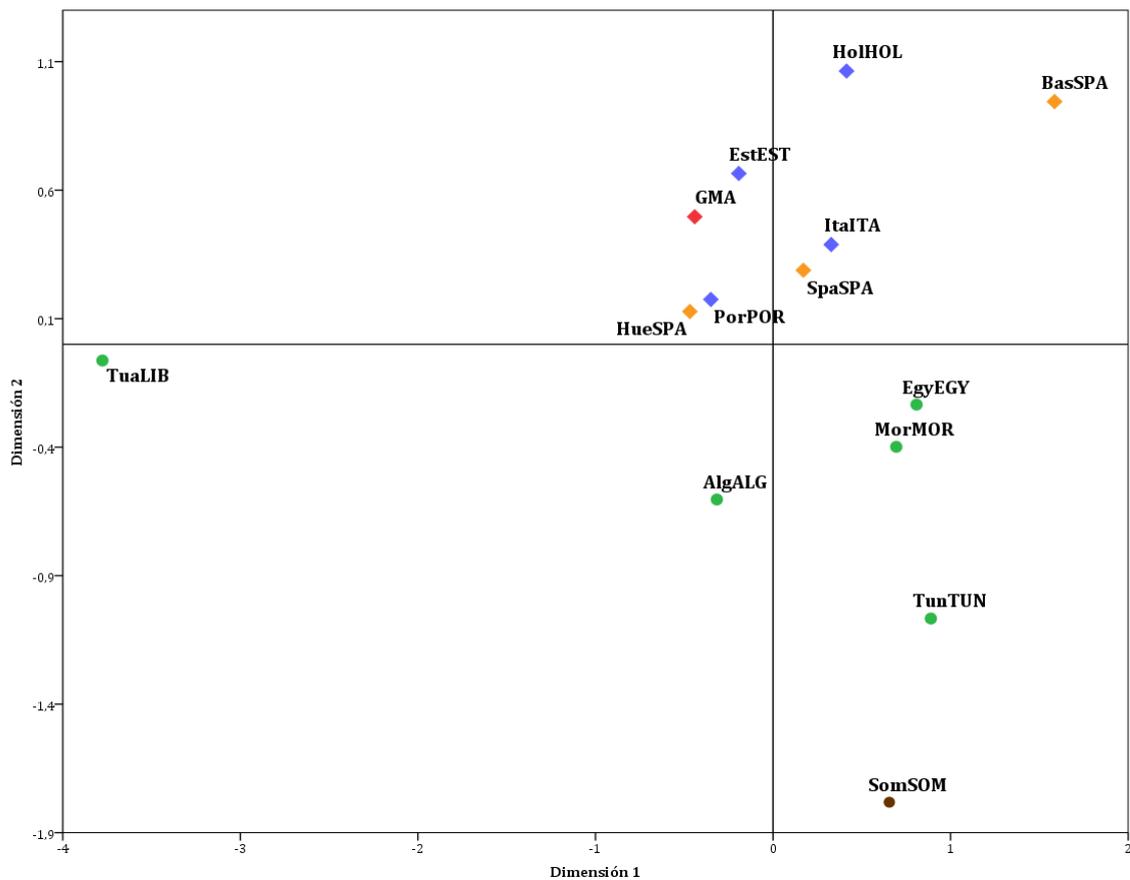


Figura 6.15: MDS calculado a partir de la distancia Fst; valor de stress 0,20281.

Todos los estudios realizados a través de la comparación de las frecuencias de los haplogrupos mitocondriales permiten afirmar que no se ha mantenido ninguna influencia materna del norte de África en la población GMA. Durante la repoblación del territorio que ocupaba en antiguo Reino de Granada se trasladaron familias enteras para habitar esta zona. Los datos obtenidos tras el análisis del ADN mitocondrial confirman que si se produjeron matrimonios entre cristianos y moriscas, éstos no tuvieron repercusión en el legado genético mitocondrial de la población actual.

4. BIBLIOGRAFÍA

- [1] I.E. Scheffler, *Mitochondria*, Wiley, New York, 2001.
- [2] S. Anderson, A.T. Bankier, B.G. Barrell, Sequence and organization of the human mitochondrial genome, *Nature*. 290 (1981) 457–465.
- [3] J.M. Butler, Mitochondrial DNA Analysis, in: *Adv. Top. Forensic DNA Typing*, Elsevier, 2012: pp. 405–456. doi:10.1016/B978-0-12-374513-2.00014-2.
- [4] S. Lutz, H.-J. Weisser, J. Heizmann, S. Pollak, Mitochondrial heteroplasmy among maternally related individuals, *Int. J. Legal Med.* 113 (2000) 155–161.
- [5] O. Handt, S. Meyer, A. Von Haeseler, Compilation of human mtDNA control region sequences, *Nucleic Acids Res.* 26 (1998) 126–129.
- [6] D.A. Clayton, Structure and function of the mitochondrial genome, *J. Inherit. Metab. Dis.* 15 (1992) 439–447.
- [7] R.C. Shuster, A.J. Rubenstein, D.C. Wallace, Mitochondrial DNA in anucleate human blood cells, *Biochem. Biophys. Res. Commun.* 155 (1988) 1360–1365.
- [8] T.J. Parsons, D.S. Muniec, K. Sullivan, N. Woodyatt, R. Alliston-Greiner, M.R. Wilson, et al., A high observed substitutional rate in the human mitochondrial DNA control region, *Nat.Genet.* 15 (1997) 362–367.
- [9] D.C. Wallace, M.D. Brown, M.T. Lott, Mitochondrial DNA variation in human evolution and disease, *Gene*. 238 (1999) 211–230.
- [10] D.C. Wallace, M.D. Brown, S. Melov, B. Graham, M. Lott, Mitochondrial biology, degenerative diseases and aging, *BioFactors*. 7 (1998) 187–190.
- [11] M. Sato, K. Sato, Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA, *Biochim. Biophys. Acta - Mol. Cell Res.* 1833 (2013) 1979–1984. doi:10.1016/j.bbamcr.2013.03.010.
- [12] W.M. Brown, M. George Jr., A.C. Wilson, Rapid evolution of animal mitochondrial DNA, *Proc. Natl. Acad. Sci. U. S. A.* 76 (1979) 1967–1971.
- [13] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, et al., The Sequence of the Human Genome, *Science* (80-.). 291 (2001) 1304–1351. doi:10.1126/science.1058040.
- [14] C.G. Kurland, S.G. Andersson, Origin and evolution of the mitochondrial proteome., *Microbiol. Mol. Biol. Rev.* 64 (2000) 786–820. doi:10.1128/MMBR.64.4.786-820.2000.
- [15] M.W. Gray, Mitochondrial evolution, *Cold Spring Harb. Perspect. Biol.* 4 (2012). doi:10.1101/cshperspect.a011403.
- [16] S.G. Andersson, A. Zomorodipour, J.O. Andersson, T. Sicheritz-Pontén, U.C. Alsmark, R.M. Podowski, et al., The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria., *Nature*. 396 (1998) 133–140. doi:10.1038/24094.
- [17] M. Gray, G. Burger, B.F. Lang, The origin and early evolution of mitochondria, *Genome Biol.* 2 (2001) reviews1018.1–reviews1018.5. doi:10.1186/gb-2001-2-6-reviews1018.
- [18] E.A. O'Brien, Y. Zhang, E. Wang, V. Marie, W. Badejoko, B.F. Lang, et al., GOBASE: An organelle genome database, *Nucleic Acids Res.* 37 (2009). doi:10.1093/nar/gkn819.
- [19] M.E. Gallardo, R. Garesse, Origen y evolución de la mitocondrial: ADN mitocondrial y evolución humana, in: M.J. López Pérez, J.M. Ortiz Melón, A. Doadrio Villarejo (Eds.), *Sist. Mitocondrial Un Reto En La Med. Humana*, Real Academia Nacional de Farmacia, Madrid, 2012: pp. 9–30.

- [20] N. Howell, C.B. Smejkal, Persistent heteroplasmy of a mutation in the human mtDNA control region: Hypermutation as an apparent consequence of simple-repeat expansion/contraction, *Am. J. Hum. Genet.* 66 (2000) 1589–1598.
- [21] R.L. Cann, A.C. Wilson, Length mutations in human mitochondrial DNA, *Genetics.* 104 (1983) 699–711.
- [22] R.E. Giles, H. Blanc, R.M. Cann, D.C. Wallace, Maternal inheritance of human mitochondrial DNA, *Natl Acad Sci.USA.* 83 (1980) 9611–9615.
- [23] M. Denaro, H. Blanc, M.J. Johnson, Ethnic variation in Hpa I endonuclease cleavage patterns of human mitochondrial DNA, *Proc. Natl. Acad. Sci. U. S. A.* 78 (1981) 5768–5772.
- [24] W.M. Brown, Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis, *Proc. Natl. Acad. Sci. U. S. A.* 77 (1980) 3605–3609.
- [25] R.L. Cann, A.C. Stoneking, A.C. Wilson, M. Stoneking, Mitochondrial DNA and human evolution, *Nature.* 356 (1987) 389–390.
- [26] A. Torroni, K. Huoponen, P. Francalacci, M. Petrozzi, L. Morelli, R. Scozzari, et al., Classification of european mtDNAs from an analysis of three European populations, *Genetics.* 144 (1996) 1835–1850.
- [27] A.S. Santachiara Benerecetti, R. Scozzari, O. Semino, A. Torroni, A. Brega, D.C. Wallace, Mitochondrial DNA polymorphisms in Italy. II. Molecular analysis of new and rare morphs from Sardinia and Rome, *Ann. Hum. Genet.* 52 (1988) 39–56.
- [28] R. Scozzari, A. Torroni, O. Semino, G. Sirugo, A. Brega, A.S. Santachiara-Benerecetti, Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms, *Am. J. Hum. Genet.* 43 (1988) 534–544.
- [29] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, W.R. Mayr, N. Morling, et al., DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing., *Forensic Sci. Int. Genet.* 13 (2014) 134–42. doi:10.1016/j.fsigen.2014.07.010.
- [30] M. Li, A. Schönberg, M. Schaefer, R. Schroeder, I. Nasidze, M. Stoneking, Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes, *Am. J. Hum. Genet.* 87 (2010) 237–249. doi:10.1016/j.ajhg.2010.07.014.
- [31] A. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, et al., DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing, *Forensic Sci Int.* 110 (2000) 79–85. doi:DOI: 10.1016/S0379-0738(00)00161-4.
- [32] J.E. Reiner, R.B. Kishore, B.C. Levin, T. Albanetti, N. Boire, A. Knipe, et al., Detection of heteroplasmic mitochondrial DNA in single mitochondria, *PLoS One.* 5 (2010). doi:10.1371/journal.pone.0014359.
- [33] M. Li, R. Schröder, S. Ni, B. Madea, M. Stoneking, Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations, *Proc. Natl. Acad. Sci.* 112 (2015) 201419651. doi:10.1073/pnas.1419651112.
- [34] G. Tully, S.M. Barritt, K. Bender, E. Brignon, C. Capelli, N. Dimo-Simonin, et al., Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts, *Forensic Sci. Int.* 140 (2004) 1–11. doi:10.1016/S0379-0738(03)00181-6.
- [35] A. Torroni, T.G. Schurr, C.-C. Yang, E.J.E. Szathmary, R.C. Williams, M.S. Schanfield, et

- al., Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations, *Genetics*. 130 (1992) 153–162.
- [36] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation., *Hum. Mutat.* 30 (2009) E386–394. doi:10.1002/humu.20921.
- [37] A.W. Röck, A. Dür, M. van Oven, W. Parson, Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA)., *Forensic Sci. Int. Genet.* 7 (2013) 601–9. doi:10.1016/j.fsigen.2013.07.005.
- [38] A. Torroni, T.G. Schurr, M.F. Cabell, M.D. Brown, J. V Neel, M. Larsen, et al., Asian affinities and continental radiation of the four founding Native American mtDNAs, *Am. J. Hum. Genet.* 53 (1993) 563–590.
- [39] D.M. Behar, R. Villems, H. Soodyall, J. Blue-Smith, L. Pereira, E. Metspalu, et al., The Dawn of Human Matrilineal Diversity, *Am. J. Hum. Genet.* 82 (2008) 1130–1140.
- [40] V. Macaulay, C. Hill, A. Achilli, C. Rengo, D. Clarke, W. Meehan, et al., Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes, *Science* (80-.). 308 (2005) 1034–1036.
- [41] D.A. Merriwether, J.A. Hodgson, F.R. Friedlaender, R. Allaby, S. Cerchio, G. Koki, et al., Ancient mitochondrial M haplogroups identified in the Southwest Pacific, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 13034–13039.
- [42] A.M. González, J.M. Larruga, K.K. Abu-Amero, Y. Shi, J. Pestano, V.M. Cabrera, et al., Mitochondrial lineage M1 traces an early human backflow to Africa, *BMC Genomics*. 8 (2007) 223. doi:10.1186/1471-2164-8-223.
- [43] A. Achilli, U.A. Perego, C.M. Bravi, M.D. Coble, Q.P. Kong, S.R. Woodward, et al., The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies, *PLoS One*. 3 (2008) e1764.
- [44] A. Achilli, C. Rengo, C. Magri, V. Battaglia, A. Olivieri, R. Scozzari, et al., 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 (2004) 910–918.
- [45] L. Pereira, M. Richards, A. Goios, A. Alonso, C. Albarrán, O. Garcia, et al., High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium., *Genome Res.* 15 (2005) 19–24. doi:10.1101/gr.3182305.
- [46] P. Grignani, C. Turchi, A. Achilli, G. Peloso, M. Alù, U. Ricci, et al., Multiplex mtDNA coding region SNP assays for molecular dissection of haplogroups U/K and J/T, *Forensic Sci. Int. Genet.* (2009).
- [47] M. Reidla, T. Kivisild, E. Metspalu, K. Kaldma, K. Tambets, H. V Tolk, et al., Origin and diffusion of mtDNA haplogroup X, *Am J Hum Genet.* 73 (2003) 1178–1190.
- [48] S. Paabo, D.M. Irwin, A.C. Wilson, DNA damage promotes jumping between templates during enzymatic amplification, *J. Biol. Chem.* 265 (1990) 4718–4721.
- [49] S. Pääbo, R.G. Higuchi, A.C. Wilson, Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology., *J. Biol. Chem.* 264 (1989) 9709–9712.
- [50] S. Paabo, Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification, *Proc. Natl. Acad. Sci. U. S. A.* 86 (1989) 1939–1943.
- [51] P. Gill, D.J. Werrett, B. Budowle, R. Guerrieri, An assessment of whether SNPs will replace STRs in national DNA databases - Joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the

- Scientific Working Group on DNA Analysis Methods (SWGDM), *Sci. Justice - J. Forensic Sci. Soc.* 44 (2004) 51–53.
- [52] L.L. Cavalli-Sforza, M.W. Feldman, The application of molecular genetic approaches to the study of human diversity, *Nat.Genet.* 33 (2003) 267.
- [53] M.B. Richards, V.A. Macaulay, H.-J.J. Bandelt, B.C. Sykes, Phylogeography of mitochondrial DNA in Western Europe, *Ann.Hum.Genet.* 61 (1998) 241–260.
- [54] M. Richards, V. Macaulay, E. Hickey, E. Vega, B. Sykes, Tracing European founder lineages in the Near Eastern mtDNA pool, *Am J Hum Genet.* 62 (2000) 241–260.
- [55] S. Finnilä, M.S. Lehtonen, K. Majamaa, Phylogenetic network for european mtDNA, *Am. J. Hum. Genet.* 68 (2001) 1475–1484.
- [56] C. Herrnstadt, J.L. Elson, E. Fahy, G. Preston, D.M. Turnbull, C. Anderson, et al., Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups, *Am. J. Hum. Genet.* 70 (2002) 1152–1171.
- [57] Q.-P.P. Kong, Y.-G.G. Yao, C. Sun, H.-J.J. Bandelt, C.-L.L. Zhu, Y.P.-P. Zhang, Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences, *Am. J. Hum. Genet.* 73 (2003) 671–676.
- [58] H.-J. Bandelt, A. Henkman, F. Nicolai, Powers of distance-hereditary graphs, *Discrete Math.* 145 (1995) 37–60.
- [59] H.-J. Bandelt, P. Forster, B.C. Sykes, M.B. Richards, Mitochondrial portraits of human populations using median networks, *Genetics.* 141 (1995) 743–753.
- [60] H.-J. Bandelt, V. Chepoi, Decomposition and l1-Embedding of Weakly Median Graphs, *Eur. J. Comb.* 21 (2000) 701–714.
- [61] H.-J. Bandelt, V. Macaulay, M. Richards, Median networks: Speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA, *Mol. Phylogenet. Evol.* 16 (2000) 8–28.
- [62] A.R. Templeton, Genetics and recent human evolution, *Evolution (N. Y.)* 61 (2007) 1507–1519.
- [63] R.E. Green, A.S. Malaspina, J. Krause, A.W. Briggs, P.L.F. Johnson, C. Uhler, et al., A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing, *Cell.* 134 (2008) 416–426.
- [64] A.W. Briggs, J.M. Good, R.E. Green, J. Krause, T. Maricic, U. Stenzel, et al., Targeted retrieval and analysis of five Neandertal mtDNA genomes, *Science (80-.)* 325 (2009) 318–321.
- [65] J. Krause, Q. Fu, J.M. Good, B. Viola, M. V Shunkov, A.P. Derevianko, et al., The complete mitochondrial DNA genome of an unknown hominin from southern Siberia, *Nature.* 464 (2010) 894–897.
- [66] D. Reich, R.E. Green, M. Kircher, J. Krause, N. Patterson, E.Y. Durand, et al., Genetic history of an archaic hominin group from Denisova Cave in Siberia, *Nature.* 468 (2010) 1053–1060.
- [67] Y.-Y. Ma, T.-F. Wu, Y.-P. Liu, Q. Wang, X.-Y. Li, Y. Ding, et al., A study of 133 Chinese children with mitochondrial respiratory chain complex I deficiency., *Clin. Genet.* 87 (2015) 179–84. doi:10.1111/cge.12356.
- [68] M.A. Sazonova, M.M. Chicheva, A. V Zhelankin, I.A. Sobenin, Y. V Bobryshev, A.N. Orekhov, Association of mutations in the mitochondrial genome with the subclinical carotid atherosclerosis in women., *Exp. Mol. Pathol.* 99 (2015) 25–32. doi:10.1016/j.yexmp.2015.04.003.

- [69] X. Tang, J. Zheng, Z. Ying, Z. Cai, Y. Gao, Z. He, et al., Mitochondrial tRNA(Ser(UCN)) variants in 2651 Han Chinese subjects with hearing loss., *Mitochondrion*. 23 (2015) 17–24. doi:10.1016/j.mito.2015.05.001.
- [70] Y. Tang, L. Wang, M. Zhu, M. Yang, K. Zhong, Q. Du, et al., Association of mtDNA M/N haplogroups with systemic lupus erythematosus: a case-control study of Han Chinese women., *Sci. Rep.* 5 (2015) 10817. doi:10.1038/srep10817.
- [71] B. Kofler, E.E. Mueller, W. Eder, O. Stanger, R. Maier, et al., Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study., *BMC Med Genet.* 10: 35. (2009).
- [72] D. Hansen, M. Healy, K. Reece, C. Smith, G.L. Woods, Repetitive-Sequence-Based PCR Using the DiversiLab System for Identification of *Aspergillus* Species, *J. Clin. Microbiol.* 46 (2008) 1835–1839. doi:10.1128/JCM.00956-07.
- [73] A. Biosystems, DNA Sequencing Analysis Software v5.3.1. Part Numbre 4346366 Rev.B, 2008.
- [74] A. Biosystems, SeqScape(R) Software v2.6 User Guide Part Numbre 4359442 Rev.A, (2007).
- [75] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA, *Nat Genet.* 23 (1999) 147.
- [76] W. Parson, A. Dur, EMPOP-A forensic mtDNA database, *Forensic Sci Int Gene.* 1 (2007) 88–92.
- [77] L. Excoffier, H. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [78] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies, *Genetics.* 164 (2003) 1567–1587.
- [79] M.J. Hubisz, D. Falush, M. Stephens, J.K. Pritchard, Inferring weak population structure with the assistance of sample group information, *Mol. Ecol. Resour.* 9 (2009) 1322–1332.
- [80] M. Jakobsson, N.A. Rosenberg, CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure, *Bioinformatics.* 23 (2007) 1801–1806.
- [81] N.A. Rosenberg, DISTRUCT: a program for the graphical display of population structure, *Mol. Ecol. Notes.* 4 (2004) 137–138.
- [82] A. Kloss-Brandstätter, D. Pacher, S. Schönherr, H. Weissensteiner, R. Binna, G. Specht, et al., HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups, *Hum. Mutat.* 32 (2011) 25–32.
- [83] J.B. Kruskal, Nonmetric multidimensional scaling: A numerical method, *Psychometrika.* 29 (1964) 115–129.
- [84] R. Aboukhalid, K. Sturk-Andreaggi, M. Bouabdellah, D. Squalli, J.A. Irwin, S. Amzazi, Mitochondrial DNA control region variation from samples of the Moroccan population, *Int. J. Legal Med.* 127 (2013) 757–759. doi:10.1007/s00414-012-0813-4.
- [85] A. Bekada, L.R. Arauna, T. Deba, F. Calafell, S. Benhamamouch, D. Comas, Genetic Heterogeneity in Algerian Human Populations., *PLoS One.* 10 (2015) e0138453. doi:10.1371/journal.pone.0138453.

- [86] J.L. Saunier, J.A. Irwin, K.M. Strouss, H. Ragab, K.A. Sturk, T.J. Parsons, Mitochondrial control region sequences from an Egyptian population sample., *Forensic Sci. Int. Genet.* 3 (2009) e97–103. doi:10.1016/j.fsigen.2008.09.004.
- [87] C. Turchi, L. Buscemi, E. Giacchino, V. Onofri, L. Fendt, W. Parson, et al., 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 (2009) 166–172. doi:DOI: 10.1016/j.fsigen.2009.01.014.
- [88] C. Ottoni, C. Martínez-Labarga, E.L. Loogväli, E. Pennarun, A. Achilli, F. De Angelis, et al., First genetic insight into Libyan Tuaregs: A maternal perspective, *Ann. Hum. Genet.* 73 (2009) 438–448. doi:10.1111/j.1469-1809.2009.00526.x.
- [89] J.C. Alvarez, D.L.E. Johnson, J.A. Lorente, E. Martinez-espin, L.J. Martinez-Gonzalez, M. Allard, et al., Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set, *Leg. Med.* 9 (2007) 293–304.
- [90] C.L. Hernández, G. Reales, J.-M. Dugoujon, A. Novelletto, J.N. Rodríguez, P. Cuesta, et al., 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 (2014) 1–16. doi:10.1186/1471-2156-15-11.
- [91] S. Cardoso, M.J. Villanueva-Millán, L. Valverde, A. Odriozola, J.M. Aznar, S. Piñeiro-Hermida, et al., Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country., *Forensic Sci. Int. Genet.* 6 (2012) e106–8. doi:10.1016/j.fsigen.2011.12.007.
- [92] S.L. Marques, A. Goios, A.M. Rocha, M.J. Prata, A. Amorim, L. Gusmão, et al., Portuguese mitochondrial DNA genetic diversity-An update and a phylogenetic revision., *Forensic Sci. Int. Genet.* 15 (2015) 27–32. doi:10.1016/j.fsigen.2014.10.004.
- [93] F. Brisighelli, V. Álvarez-Iglesias, M. Fondevila, A. Blanco-Verea, Á. Carracedo, V.L. Pascali, et al., Uniparental Markers of Contemporary Italian Population Reveals Details on Its Pre-Roman Heritage, *PLoS One.* 7 (2012). doi:10.1371/journal.pone.0050794.
- [94] L. Chaitanya, M. van Oven, S. Brauer, B. Zimmermann, G. Huber, C. Xavier, et al., High-quality mtDNA control region sequences from 680 individuals sampled across the Netherlands to establish a national forensic mtDNA reference database., *Forensic Sci. Int. Genet.* 21 (2015) 158–167. doi:10.1016/j.fsigen.2015.12.002.
- [95] M. Stoljarova, J.L. King, M. Takahashi, A. Aaspõllu, B. Budowle, Whole mitochondrial genome genetic diversity in an Estonian population sample, *Int. J. Legal Med.* (2015) 67–71. doi:10.1007/s00414-015-1249-4.
- [96] M. Mikkelsen, L. Fendt, A.W. Röck, B. Zimmermann, E. Rockenbauer, A.J. Hansen, et al., Forensic and phylogeographic characterisation of mtDNA lineages from Somalia, *Int. J. Legal Med.* 126 (2012) 573–579. doi:10.1007/s00414-012-0694-6.
- [97] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, *Genetics.* 164 (2003) 1567–1587. doi:Article.
- [98] L.J. Martinez-Gonzalez, *Estructura Genética de la población de Guatemala. Aplicaciones en el campo antropológico y forense*, Universidad de Granada, 2011.
- [99] O. García, R. Fregel, J.M. Larruga, V. Álvarez, I. Yurrebaso, V.M. Cabrera, et al., Using mitochondrial DNA to test the hypothesis of a European post-glacial human

- recolonization from the Franco-Cantabrian refuge, *Heredity* (Edinb). 106 (2010) 37–45. doi:10.1038/hdy.2010.47.
- [100] C.L. Hernández, G. Reales, J.-M. Dugoujon, A. Novelletto, J.N. Rodríguez, P. Cuesta, et al., 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 (2014) 11. doi:10.1186/1471-2156-15-11.
- [101] D.M. Behar, M. Van Oven, S. Rosset, M. Metspalu, E.-L. Loogväli, N.M. Silva, et al., A “copernican” reassessment of the human mitochondrial DNA tree from its root, *Am. J. Hum. Genet.* 90 (2012) 675–684.
- [102] B. Malyarchuk, T. Grzybowski, M. Derenko, M. Perkova, T. Vanecek, J. Lazur, et al., Mitochondrial DNA phylogeny in eastern and western Slavs, *Mol. Biol. Evol.* 25 (2008) 1651–1658. doi:10.1093/molbev/msn114.
- [103] N. Maca-Meyer, A.M. González, J. Pestano, C. Flores, J.M. Larruga, V.M. Cabrera, et al., Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography, *BMC Genet.* 4 (2003) 15. doi:10.1186/1471-2156-4-15.
- [104] E. Pennarun, T. Kivisild, E. Metspalu, M. Metspalu, T. Reisberg, J.-P. Moisan, et al., Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa., *BMC Evol. Biol.* 12 (2012) 234. doi:10.1186/1471-2148-12-234.
- [105] C. Coudray, A. Olivieri, A. Achilli, M. Pala, M. Melhaoui, M. Cherkaoui, et al., The complex and diversified mitochondrial gene pool of Berber populations, *Ann. Hum. Genet.* 73 (2009) 196–214.
- [106] A.M. González, A. Brehm, J.A. Pérez, N. Maca-Meyer, C. Flores, V.M. Cabrera, Mitochondrial DNA affinities at the Atlantic fringe of Europe., *Am. J. Phys. Anthropol.* 120 (2003) 391–404. doi:10.1002/ajpa.10168.
- [107] M.J. Casas, E. Hagelberg, R. Fregel, J.M. Larruga, A.M. González, Human mitochondrial DNA diversity in an archaeological site in al-Andalus: Genetic impact of migrations from North Africa in Medieval Spain, *Am. J. Phys. Anthropol.* 131 (2006) 539–551.
- [108] L. Alvarez, C. Santos, A. Ramos, R. Pratdesaba, P. Francalacci, M.P. Aluja, Mitochondrial DNA patterns in the Iberian Northern plateau: population dynamics and substructure of the Zamora province., *Am. J. Phys. Anthropol.* 142 (2010) 531–9. doi:10.1002/ajpa.21252.
- [109] E. Ruiz-Pesini, M.T. Lott, V. Procaccio, J.C. Poole, M.C. Brandon, D. Mishmar, et al., An enhanced MITOMAP with a global mtDNA mutational phylogeny, *Nucleic Acids Res.* 35 (2007) D823–D828.
- [110] M. Pala, A. Olivieri, A. Achilli, M. Accetturo, E. Metspalu, M. Reidla, et al., Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia, *Am. J. Hum. Genet.* 90 (2012) 915–924. doi:10.1016/j.ajhg.2012.04.003.

VARIABILIDAD GENÉTICA EN LA POBLACIÓN GMA Y SU RELACIÓN
CON LOS ANTIGUOS POBLADORES DEL NORTE DE ÁFRICA

GENETIC VARIABILITY IN THE GMA POPULATION AND ITS
RELATION WITH THE FORMER INHABITANTS OF NORTH AFRICA

7

ABSTRACT

After the study of 245 samples from the provinces of Granada, Malaga and Almeria both autosomal STRs and autosomal INDELS, Y chromosome STRs and mitochondrial DNA control region, a great homogeneity is observed in the results despite the different heritage of the markers studied. Very high discrimination power values have been achieved for autosomal STRs, autosomal INDELS and Y chromosome STRs, as far as haplotype diversity and discrimination capacity in lineage markers.

The populations of the provinces of Granada, Malaga and Almeria behave as a single population in regards to their genetic structure. Y chromosome and mitochondrial DNA haplogroups found in the GMA population indicate that the possible male influence and female of the North African inhabitants could have been remained in the population of the south of the peninsula has not influenced the genetic bequest of population of this region. The analyses of allelic frequencies, mitochondrial haplogroup frequencies and genetic distances confirm that the GMA population is strongly linked to Mediterranean populations, especially the Spanish ones.

1. INTRODUCTION

For the nearly 800 years of the Arab invasion in the Peninsula, the North African groups spread through the country except the Basque Country, Galicia, Cantabria, Asturias and most of the Pyrenees; but it was in the south of the Peninsula where they had a greater influence [1]. The study of mtDNA and Y chromosome has identified geographic regions with 8-10% of North African genetic influence [2-4]. However, the study of autosomal markers yields lower values of North African influence (2.1-9%) [2]. Crossbreeding studies based on the characterization of ALU sequences found traces of sub-Saharan genes in north Mediterranean populations, suggesting continuous contacts between both coasts [5]. The fact that genetic traits and certain specific haplotypes have been detected along the north coast of the Mediterranean confirms the hypothesis that gene flow in this region is linked to the first trans-Mediterranean sailings and remained homogeneous while trade slave lasted, until the late seventeenth century; rather than a reflection of Islamic expansion (S. VII to S. XV) [5]. The analysis of genome-wide SNP data from over 2,000 individuals has allowed the characterization of broad clinal patterns of recent gene flow between Europe and Africa that have a substantial effect on genetic diversity of European populations, especially in the southwest European populations [6]. Y chromosome studies have described the Iberian Peninsula genetic structure calculating the different contributions of Muslims and Jews to the current population of the Peninsula [3] mainly focused on the study of E3b2 haplogroup, common in northern Africa and present in a 5.6% in the Peninsula [7]. The distribution of Y chromosome haplogroup E3b2-M81 in the Iberian Peninsula suggests a genetic flow of North Africa during this period [3]. High levels of patrilineal descent from North Africa and Sephardic Jews in the current population of the Iberian Peninsula have been detected [3]. However, contrary to what might be expected based on historical data that would favor a gradient from south to north of North African genetic influence, most genetic influence is found in Galicia and northern Castilla (>20 %). The main North African gene frequencies gradient is observed between the west and east, the latter being the one with smaller proportions. After the Reconquest, the Moors were distributed homogeneously in the Peninsula, but their final expulsion in 1609 was much more effective in some regions of Spain, Valencia and western Andalusia, while in Galicia and Extremadura the population dispersed and integrated into the society [3].

After the study of 245 samples from the provinces of Granada, Malaga and Almeria both autosomal STRs and autosomal INDELS, Y chromosome STRs and mitochondrial DNA control region, a great homogeneity is observed in the results despite the different heritage of the markers studied.

2. FORENSIC ANALYSIS

Unsurprisingly, the most informative markers for GMA population are autosomal STRs due to its high degree of polymorphism and heterozygosity (Table 7.1). The combination of a not very large number of autosomal STRs (15 markers) allows the differentiation between individuals with a reliability >0.9999999999999999 . If additionally, the use of autosomal STRs is combined with autosomal INDELS and Y chromosome STRs (in males), the power of discrimination greatly increases. On the other hand, a good characterization of the mutations found in the mitochondrial DNA control region and the determination of mitochondrial haplogroups lead to a better characterization of populations. Thanks to this, the analysis of mitochondrial DNA from a forensic sample can target the geographic origin of the individual. In addition, the characterization of not only Y chromosome STRs, but also male lineages defining SNPs will give a better identification of the sample, especially to infer its geographical origin.

Table 7.1: Forensic summary statistics

	n	n mark	MDP	MEP	MPIC	MMP	MHo	CDP	CEP	HD	DC
a-STRs	245	15	0,921	0,555	0,741	0,083	0,773	$1 - 5,554 \cdot 10^{-18}$	0,999	-	-
INDELS	143	38	0,581	0,136	0,339	0,430	0,422	$1 - 3,550 \cdot 10^{-15}$	0,996	-	-
Y-STRs	146	17	0,629	-	0,574	0,371	-	$1 - 6,172 \cdot 10^{-8}$	-	0,9974	95,21%
RC mit	191	1121	-	-	-	-	-	-	-	0,9973	94,09%

n, number of samples; *n mark*, number of markers; *MDP*, Mean power of discrimination; *MEP*, Mean power of exclusion; *MPIC*, Mean polymorphic informative content, *MMP*, Mean matching probability, *MHo*, Mean observed heterozygosity, *CDP*, Combined power of discrimination, *CEP*, Combined power of exclusion; *HD*, Haplotype diversity; *DC*, Discrimination capacity.

These results underscore the importance of a good genetic characterization of populations, both for further forensic and genetic population structure studies and to determine the genetic relationships that have been established between them through their history.

3. GMA POPULATION STRUCTURE

As determined in all chapters, the populations of the provinces of Granada, Malaga and Almeria behave as a single population in regards to their genetic structure. STRUCTURE analyses have identified two different distribution patterns of the ancestral components found in the population.

While autosomal markers uniquely identify a single ancestral component within the sample placing all individuals in the center of the triangular diagram (figures 7.1a and 7.1b), lineage markers identify a large number of ancestral components, generally related to the haplogroups present in the GMA population (figures 7.1c

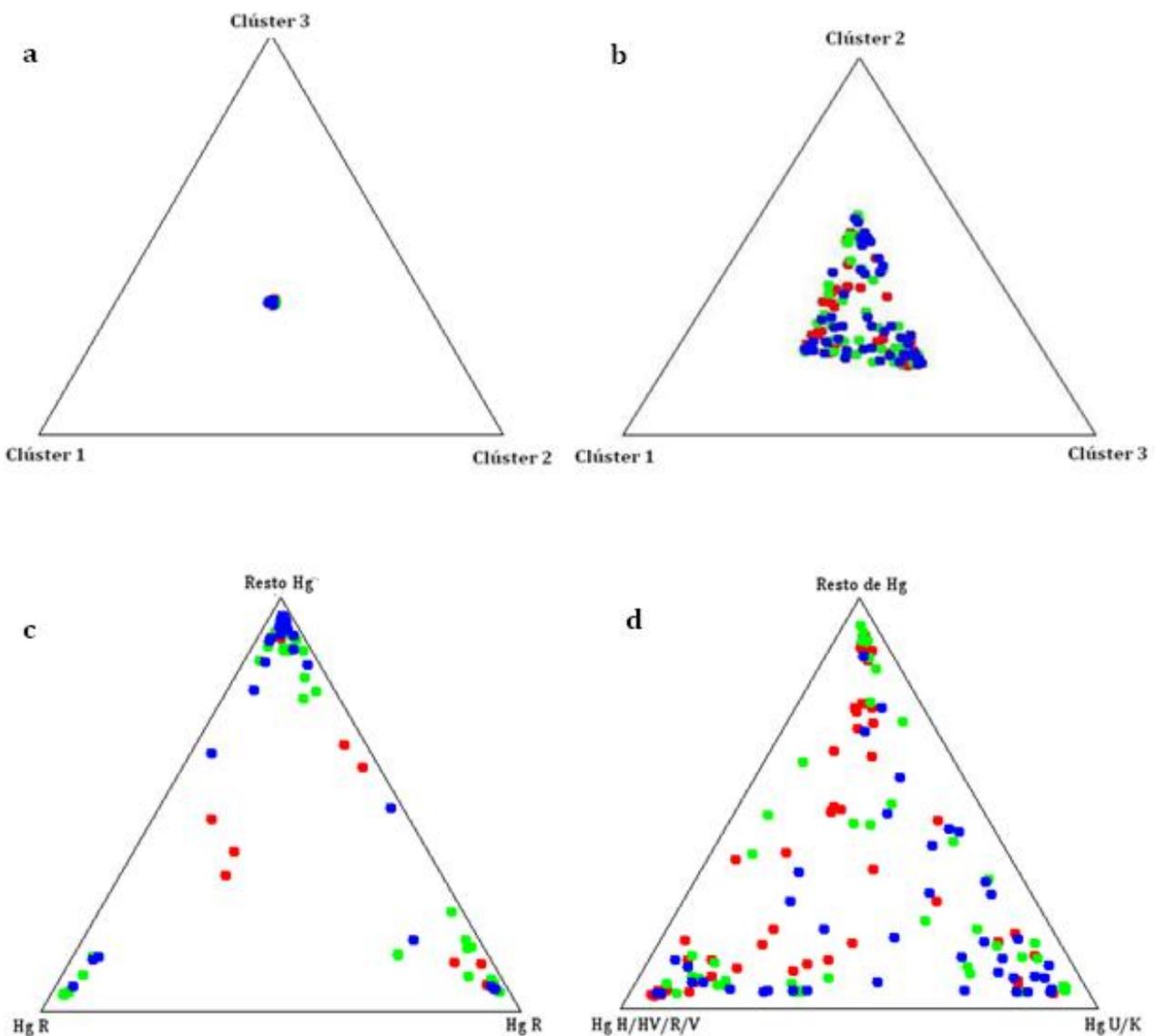


Figure 7.1: Triangular diagrams generated with STRUCTURE for the genetic homology grouping of the individuals of the three provinces; *a*, aSTRs; *b*, INDELs; *c*, Y-STRs; *d*, mtDNA. Ancestry components are represented at each vertex of the triangle and provinces are presented by a different colour.

and 7.1d). As it has been previously stated in chapters 5 and 6, STRUCTURE analysis using uniparental markers show similar results as haplogroup studies. As it can be seen in the figure, although individuals are distributed into different components, these components are allocated similarly in the populations of each of the provinces favoring a single common node for the three populations studied. These results are in line with the historical and cultural expectations as they share the same origin plus a large geographical proximity.

4. HAPLOGROUPS

Both the Y chromosome and mitochondrial DNA haplogroups described, clearly show a big difference between North African populations and the Spanish ones, more specifically with the GMA population (Figure 7.2). The main haplogroups found in North African populations are E (Y chromosome) and L (mtDNA). Both are found in the Spanish population but at much lower frequencies and always similar to those observed in other European populations [8,9]. Analyzing the second most common Y chromosome haplogroup among North African populations (haplogroup J), it can be seen as the GMA population has high values of this haplogroup. However, a detailed evaluation of J subhaplogroups shows that the prevalent in the GMA population is lineage J2a4, found along the north coast of the Mediterranean and the Caucasus; while the predominant in the North African populations are J1 lineages, found at low frequencies in subpopulations of southern Europe. A similar phenomenon occurs with U mtDNA haplogroup. Although similar frequencies are observed among all populations, U5 lineage is predominantly in Spanish (8.90% in the GMA population) while U6 lineage predominates in North Africa. The U6 lineage is considered as a marker of female lineage in North Africa, especially in the Maghreb [10,11]. The frequencies of this haplogroup in North African populations vary between 11% (Moroccan Berbers) [12] and 6.5% (Algeria) [13]. However, it is only present in 2.62% of individuals of the GMA population (Table 6.7). These data are similar to those found in other Spanish populations (Galicia 2.20% [14]; Córdoba 0.90% [15]; Zamora 0.50% [16]; Basque Country 0.94% [17] such as the Mediterranean basin (Italy 0.86% [18]).

Y chromosome and mitochondrial DNA haplogroups found in the GMA population indicate that the possible male influence and female of the North African inhabitants could have been remained in the population of the south of the peninsula has not influenced the genetic bequest of population of this region. After

the Reconquest of the region by the Catholic Monarchs, the region was repopulated with entire families from the rest of the peninsula very effectively [19]. Although many of the Moriscos who inhabited the region were converted to Christianity and although mixed marriages might have been happend, the lineages of both Y chromosome and mitochondrial DNA indicate that this phenomenon might have occurred in such low frequency that did not generate a great influence able to survive the 600 years that have passed since the dissolution of the Kingdom of Granada.

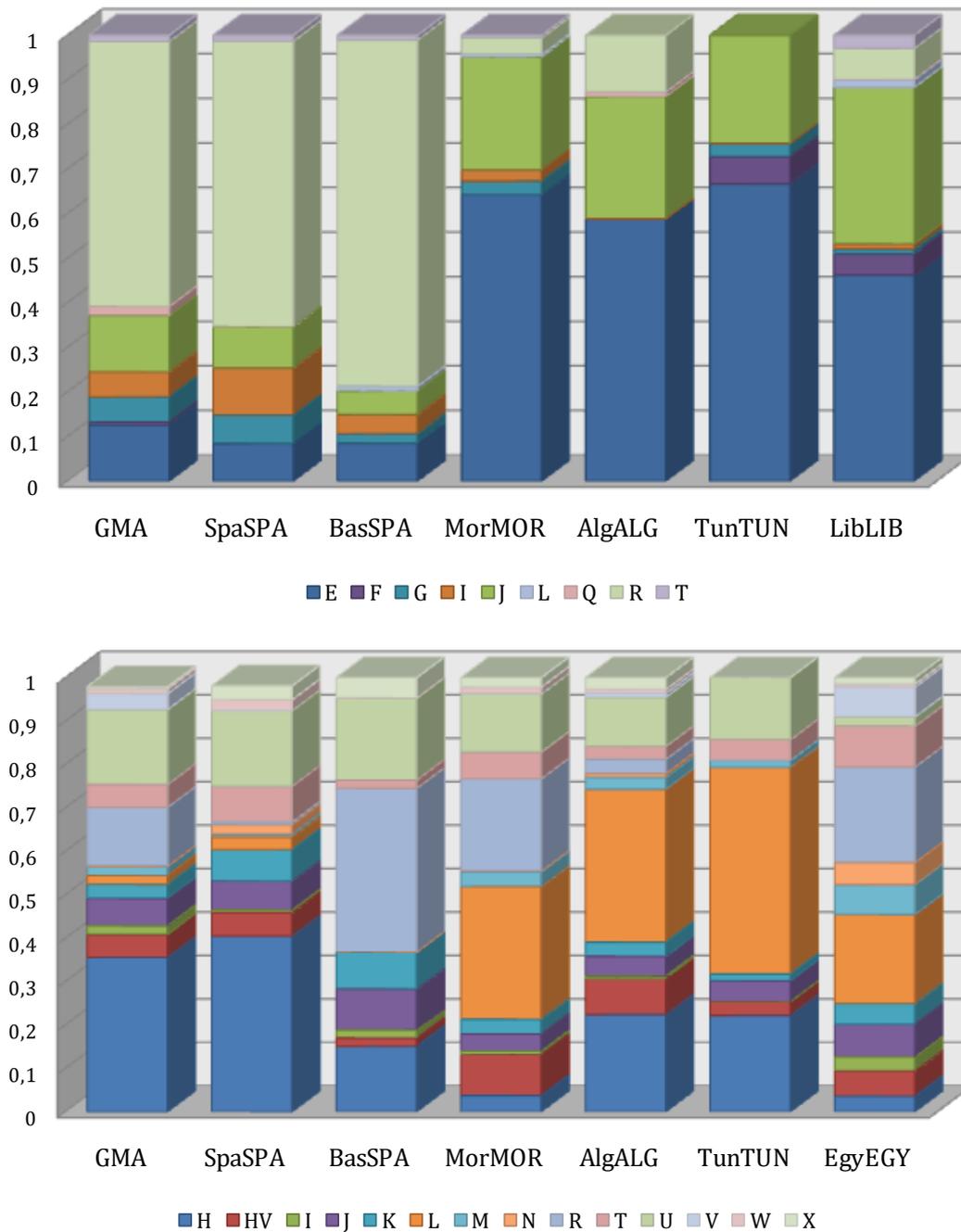


Figure 7.2: Histograms of the haplogroup frequencies in GMA, Spanish and North African populations. *Top* Y chromosome haplogroups; *down* mitochondrial DNA haplogroups.

5. ALLELIC AND HAPLOGROUP FREQUENCIES

The analysis of allelic frequencies of different STRs markers, both autosomal and Y chromosome; and frequencies of mitochondrial DNA haplogroups clearly group European populations on the positive X axis and North African populations on the negative side of the X axis (Figure 7.3). As it has already observed through the previous chapters, the GMA population is strongly linked to European populations.

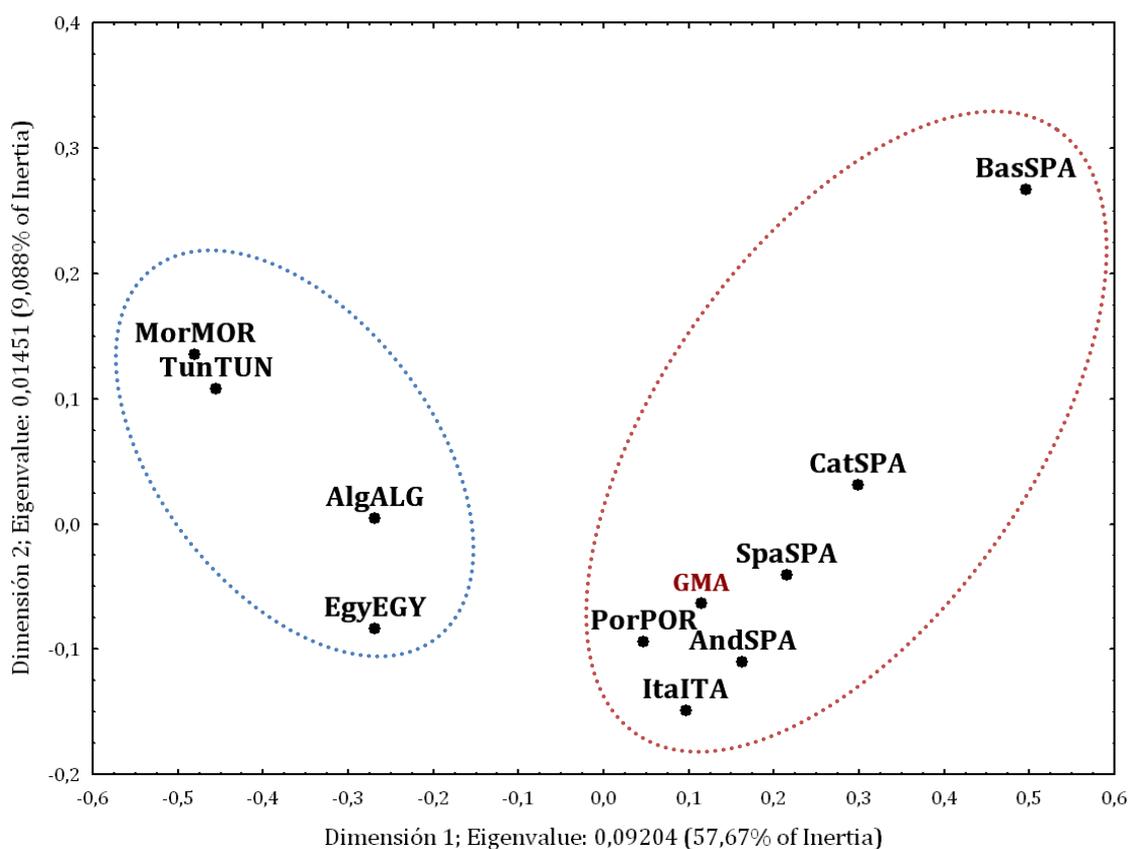


Figure 7.3: Correspondence analysis showing the relationship between allele frequencies of aSTRs and Y-STRs markers, the frequencies of mtDNA haplogroups and the populations studied.

The big difference between the frequencies of mtDNA haplogroup L (African origin) and H (European origin) generates these two populations association. The distance of the Moroccan (MorMOR) and Tunisian (TunTUN) populations regarding the Algerian (AlgALG) and Egyptian (EgyEGY) populations is caused by the frequencies of allele 13 in DYS19 Y chromosomal marker.

6. GENETIC DISTANCES

The results of genetic distances corroborate previously results obtained through haplogroup and genetic frequencies analyses (figure 7.4). As can be seen in all the studied genetic markers, Mediterranean populations are genetically closest to the GMA populations, especially the Spanish ones. The only exception is the population of the Basque Country (BasSPA). As already mentioned above, this population contains a large number of ancestral Palaeolithic alleles [20]. Furthermore, there are many mtDNA lineages shared with the European maternal pool. However, there are some lineages that are virtually unique to this population [21].

On the other hand, in general all the African populations are within the same range of distances with GMA population. Moroccan Berber (BAsMOR and BBoMOR) and Tunisian (BKeTUN) populations, in the case of autosomal markers (figure 7.3 *above*), and Lebanese Tuareg population (TuaLIB) in the case of mitochondrial DNA (figure 7.3 *below*) should be highlighted. As it happened with Muslim groups in the Iberian Peninsula after the conquest, the Berbers were isolated in the regions of North Africa after the Arab conquest of the territories they inhabited during the S. VII [22] fact that makes that large genetic differences exist between these groups and populations geographically close to them. Lebanese Tuareg population has two major mitochondrial components. It has a strong European influence determined by the frequencies of the haplogroups H1 (61.24%) and V. It also presents a typical sub-Saharan African component present as a higher frequency of L0 haplogroup (6.20%) compared to the rest of North African populations [23]. The high presence of sub-Saharan mitochondrial DNA haplogroups in the Algerian population (AlgALG) [13] locates it in the same range of genetic distances as Lebanese Tuaregs (figure 7.3 *below*).

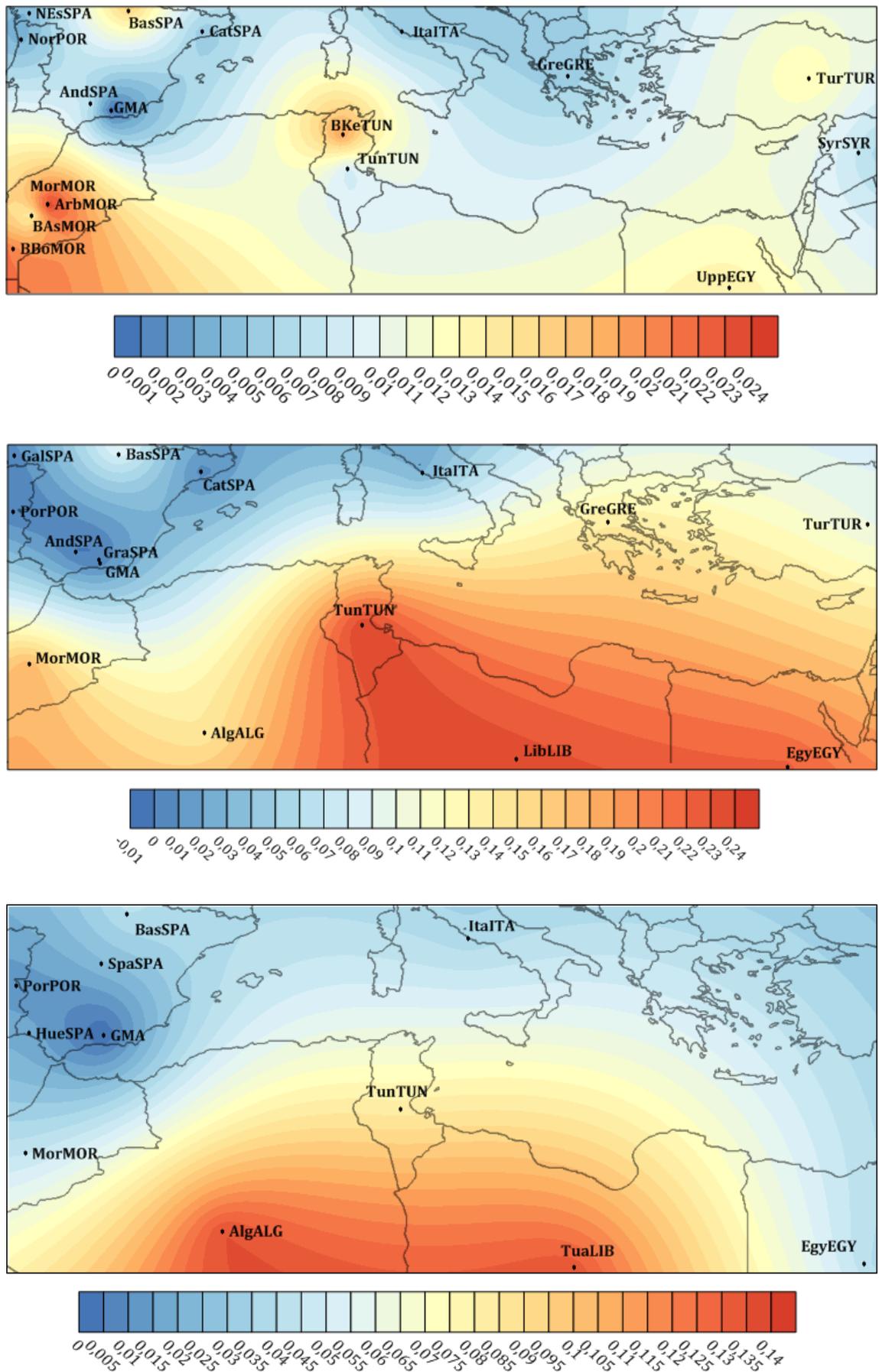


Figura 7.4: Map of genetic distances between GMA population, Mediterranean and North African populations plotted with Surfer 13 using the Kriging method. Autosomal STRs Cavalli-Sforza genetic distances (*top*). Y STRs *RST* genetic distances (*center*). mtDNA haplogroup *FST* genetic distances (*below*). Dark blue corresponds to lower values and dark red to higher values

7. FINAL CONSIDERATIONS

After the characterization of all populations for the studied genetic markers, it has been able to determine the extremely high discrimination power of these markers, which makes it virtually impossible to find two people with the same genetic information.

The rapid development of genotyping new technologies has increased the degree of detail with which one can infer the biogeographical origin of the samples. Possible future studies based on the analysis of Y chromosome and mitochondrial DNA coding region SNPs will allow a much better characterization of haplogroups. The characterization of a large number of AIM markers will lead to a better establishment of the origin of the population.

As this study is an anthropological population study, it has been taken into account the geographical origin of the samples as well as the origin of three previous generations. It makes that data are not biased by recent migratory flows. To have in mind at least three generations residing in the study area helps to verify the population genetic structure, to interpret it correctly through improving the representativeness of the sample and to estimate more accurately the data of the observed genetic events. Generally, these data are not taken into account in forensic studies, in which studied individuals are residents of the study area and easily encompass individuals whose ancestors come from other areas.

There have been many cultural traits that the people of the kingdom of Granada left in this region as evidenced by still used daily customs, names of towns and cities, expressions and words, architecture and gastronomy. However, as has been seen throughout all chapters, both the study of markers separately and their joint study indicates that the current population of the provinces of Granada, Malaga and Almeria has no significant genetic influence of the inhabitants of the ancient Kingdom of Granada compared to the rest of Spanish or southern Europe populations. After the Reconquest of the region by the Catholic Monarchs, the region was repopulated with entire families from the rest of the peninsula so effectively [19] that any possible North African genetic influence was totally eliminated.

The vast majority of the surnames present in the population show higher frequencies in the South of the Iberian Peninsula although they come from the North and the Center of the Peninsula. Both Y chromosome and mitochondrial DNA haplogroups, clearly show big differences between the North African populations

and the Spanish ones, more specifically with the studied population. Furthermore, the data obtained from the study of allelic frequencies and mitochondrial DNA haplogroups as far as genetic distances analysis claim that the current population of the provinces of Granada, Malaga and Almeria show no significant genetic influence of the inhabitants of the ancient Kingdom of Granada compared to other Spanish populations or southern Europe ones.

Given that the territory occupied by the provinces of Granada, Malaga and Almeria was repopulated with a limited number of families from specific regions of the Iberian Peninsula [19], a large genetic variability observed among individuals studied is detected thanks to microevolution phenomena.

8. REFERENCES

- [1] M. Brion, A. Salas, A. González-Neira, M. V Lareu, A. Carracedo, Insights into Iberian population origins through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite, *Am. J. Phys. Anthropol.* 122 (2003) 147–161. doi:10.1002/ajpa.10231.
- [2] M. Pino-Yanes, A. Corrales, S. Basaldúa, A. Hernández, L. Guerra, J. Villar, et al., North African influences and potential bias in case-control association studies in the Spanish population., *PLoS One.* 6 (2011) e18389. doi:10.1371/journal.pone.0018389.
- [3] S.M. Adams, E. Bosch, P.L. Balaesque, S.J. Ballereau, A.C. Lee, E. Arroyo, et al., The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula., *Am. J. Hum. Genet.* 83 (2008) 725–36. doi:10.1016/j.ajhg.2008.11.007.
- [4] N. Maca-Meyer, P. Sánchez-Velasco, C. Flores, J.-M. Larruga, A.-M. González, A. Oterino, et al., Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain), *Ann. Hum. Genet.* 67 (2003) 329–339.
- [5] E. González-Pérez, E. Esteban, M. Via, M. Gayà-Vidal, G. Athanasiadis, J.M. Dugoujon, et al., Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems), *Am. J. Phys. Anthropol.* 141 (2010) 430–9. doi:10.1002/ajpa.21161.
- [6] L.R. Botigué, B.M. Henn, S. Gravel, B.K. Maples, C.R. Gignoux, E. Corona, et al., Gene flow from North Africa contributes to differential human genetic diversity in southern Europe., *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 11791–6. doi:10.1073/pnas.1306223110.
- [7] C. Flores, N. Maca-Meyer, A.M. González, P.J. Oefner, P. Shen, J.A. Pérez, et al., Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography., *Eur. J. Hum. Genet.* 12 (2004) 855–63. doi:10.1038/sj.ejhg.5201225.
- [8] L. Alvarez, C. Santos, R. Montiel, B. Caeiro, A. Baali, J.-M. Dugoujona, et al., Y-chromosome variation in South Iberia: insights into the North African contribution., *Am. J. Hum. Biol.* 21 (2009) 407–9. doi:10.1002/ajhb.20888.
- [9] C.L. Hernández, G. Reales, J.-M. Dugoujon, A. Novelletto, J.N. Rodríguez, P. Cuesta, et al., 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 (2014) 11. doi:10.1186/1471-2156-15-11.
- [10] N. Maca-Meyer, A.M. González, J. Pestano, C. Flores, J.M. Larruga, V.M. Cabrera, et al., Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography, *BMC Genet.* 4 (2003) 15. doi:10.1186/1471-2156-4-15.
- [11] E. Pennarun, T. Kivisild, E. Metspalu, M. Metspalu, T. Reisberg, J.-P. Moisan, et al., Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa., *BMC Evol. Biol.* 12 (2012) 234. doi:10.1186/1471-2148-12-234.
- [12] C. Coudray, A. Olivieri, A. Achilli, M. Pala, M. Melhaoui, M. Cherkaoui, et al., The complex and diversified mitochondrial gene pool of Berber populations, *Ann. Hum.*

- Genet. 73 (2009) 196–214.
- [13] A. Bekada, L.R. Arauna, T. Deba, F. Calafell, S. Benhamamouch, D. Comas, Genetic Heterogeneity in Algerian Human Populations., *PLoS One*. 10 (2015) e0138453. doi:10.1371/journal.pone.0138453.
- [14] A.M. González, A. Brehm, J.A. Pérez, N. Maca-Meyer, C. Flores, V.M. Cabrera, Mitochondrial DNA affinities at the Atlantic fringe of Europe., *Am. J. Phys. Anthropol.* 120 (2003) 391–404. doi:10.1002/ajpa.10168.
- [15] M.J. Casas, E. Hagelberg, R. Fregel, J.M. Larruga, A.M. González, Human mitochondrial DNA diversity in an archaeological site in al-Andalus: Genetic impact of migrations from North Africa in Medieval Spain, *Am. J. Phys. Anthropol.* 131 (2006) 539–551.
- [16] L. Alvarez, C. Santos, A. Ramos, R. Pratdesaba, P. Francalacci, M.P. Aluja, Mitochondrial DNA patterns in the Iberian Northern plateau: population dynamics and substructure of the Zamora province., *Am. J. Phys. Anthropol.* 142 (2010) 531–9. doi:10.1002/ajpa.21252.
- [17] S. Cardoso, M.J. Villanueva-Millán, L. Valverde, A. Odriozola, J.M. Aznar, S. Piñeiro-Hermida, et al., Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country., *Forensic Sci. Int. Genet.* 6 (2012) e106–8. doi:10.1016/j.fsigen.2011.12.007.
- [18] F. Brisighelli, V. Álvarez-Iglesias, M. Fondevila, A. Blanco-Verea, Á. Carracedo, V.L. Pascali, et al., Uniparental Markers of Contemporary Italian Population Reveals Details on Its Pre-Roman Heritage, *PLoS One*. 7 (2012). doi:10.1371/journal.pone.0050794.
- [19] P. Bueno, *El reino de Granada (De orígenes a 1936)*, Don Quijote Editorial, Granada, 2004.
- [20] O. García, I. Yurrebaso, I. Uriarte, J.A. Pérez, R. Peñas, S. Alonso, et al., Distribution of Y-chromosomal haplotypes in the Basque Country autochthonous population using a 17-locus multiplex PCR assay, *Int. Congr. Ser.* 1288 (2006) 319–321.
- [21] D.M. Behar, C. Harmant, J. Manry, M. van Oven, W. Haak, B. Martinez-Cruz, et al., The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times., *Am. J. Hum. Genet.* 90 (2012) 486–93. doi:10.1016/j.ajhg.2012.01.002.
- [22] C. Coudray, E. Guitard, C. Keyser-Tracqui, M. Melhaoui, M. Cherkaoui, G. Larrouy, et al., Population genetic data of 15 tetrameric short tandem repeats (STRs) in Berbers from Morocco, *Forensic Sci. Int.* 167 (2007) 81–86.
- [23] C. Ottoni, C. Martínez-Labarga, E.L. Loogväli, E. Pennarun, A. Achilli, F. De Angelis, et al., First genetic insight into Libyan Tuaregs: A maternal perspective, *Ann. Hum. Genet.* 73 (2009) 438–448. doi:10.1111/j.1469-1809.2009.00526.x.

CONCLUSIONES

CONCLUSIONS

Primera. Tras analizar una gran batería de marcadores genéticos, se afirma que la población de las provincias de Granada, Málaga y Almería se comporta como una única población en lo que a su estructura genética se refiere.

Segunda. El poder de discriminación combinado para STRs autosómicos, INDELS autosómicos y STRs de cromosoma Y es de $1 - 5,554 \cdot 10^{-18}$, $1 - 3,550 \cdot 10^{-15}$ y $1 - 6,172 \cdot 10^{-8}$, respectivamente. Tanto los STRs como INDELS autosómicos estudiados se encuentran en equilibrio Hardy-Weinberg y no se encuentran ligados entre sí en la población analizada. La diversidad haplotípica del cromosoma Y es 0,9974 y la del ADN mitocondrial es 0,9973. La capacidad de discriminación del cromosoma Y y del ADN mitocondrial es de 95,21% y 94,09%, respectivamente.

Tercera. Se han observado 266 apellidos diferentes en el total de la población estudiada (165 si sólo se tiene en cuenta el primer apellido). La gran mayoría de los apellidos tienen origen en las regiones norte (35%) y centro de la Península (37%). Sin embargo, se observa que un 37% de los apellidos están distribuidos por Andalucía, 22% en Castilla y 20% en el norte de la Península. Un 12,57% de los apellidos con mayor número de individuos nacidos en el sur de la Península son de origen gallego, asturiano o cántabro, el 18,03% provienen del centro de la Península, Castilla León y Castilla la Mancha y 7,65% son de origen navarro-aragonés. La mayoría de los apellidos de los varones están asociados al haplogrupo de cromosoma Y R1b1b2-M269. Ninguno de los individuos que presentan haplotipos de cromosoma Y compartidos tienen el mismo apellido.

Cuarta. Los haplogrupos de cromosoma Y mayoritarios en la población GMA son los haplogrupo R1b1b2 (58,22%) y E1b1b1 (11,64%). Los haplogrupos mitocondriales mayoritarios en la población GMA son el haplogrupo H (36,13%) y el haplogrupo U (17,28%). Tanto para los haplogrupos descritos para cromosoma Y como para ADN mitocondrial, se observa claramente una gran diferencia entre

las poblaciones del norte de África y las poblaciones españolas y más concretamente con la población de estudio.

Quinta. Los datos obtenidos del estudio de frecuencias alélicas y de haplogrupos de ADN mitocondrial así como las distancias genéticas afirman que la población actual de las provincias de Granada, Málaga y Almería no presenta una influencia genética significativa de los pobladores del antiguo Reino de Granada comparado con el resto de poblaciones españolas o del sur de Europa.

First. After analyzing a large battery of genetic markers, it can be affirmed that the population of the provinces of Granada, Malaga and Almeria behave as a single population concerning to genetic structure.

Second. The combined discrimination power for autosomal STRs, autosomal INDELS and Y chromosome STRs are $1 - 5.554 \cdot 10^{-18}$, $1 - 3.550 \cdot 10^{-15}$ and $1 - 6.172 \cdot 10^{-8}$, respectively. No deviations from Hardy-Weinberg equilibrium are observed in both autosomal STRs and INDELS and there are not any signs of linkage between loci. Y chromosomal haplotypic diversity is 0.9974 and 0.9973 for mitochondrial DNA. Discrimination capacity for Y chromosome and mitochondrial DNA is 95.21% y 94.09%, respectively.

Third. 266 different surnames have been observed in the GMA population (165 when only the first surname is considered). The vast majority originate in the North (35%) and Center (37%) of the Iberian Peninsula. However, 37% of surnames are mainly distributed in Andalusia, 22% in Castilla and 20% in the North of the Peninsula. 12.57% of the surnames mainly distributed in the South of the Peninsula are original from Galicia, Asturias and Cantabria, 18.03% come from the Center of the Peninsula, Castilla León and Castilla la Mancha and 7.65% have Navarro-Aragonese origin. Most names are associated with the Y chromosome haplogroup R1b1b2-M269. None of the individuals with shared Y chromosome haplotypes have the same surname.

Fourth. Principal Y chromosome haplogroups in GMA population are R1b1b2 (58.22%) and E1b1b1 (11.64%). Main mitochondrial haplogroups in GMA population are H (36.13%) and U (17.28%). Both Y chromosome and mitochondrial DNA haplogroups, clearly show big differences between the North African populations and the Spanish ones, more specifically with the studied population.

Fifth. The data obtained from the study of allelic frequencies and mitochondrial DNA haplogroups as far as genetic distances analysis claim that the current population of the provinces of Granada, Malaga and Almeria show no significant genetic influence of the inhabitants of the ancient Kingdom of Granada compared to other Spanish populations or southern Europe ones.

PROCOLOS DE EXTRACCIÓN Y CUANTIFICACIÓN

EXTRACTION AND QUANTIFICATION PROTOCOLS



Protocolo de extracción orgánica para muestras mediante fenol/cloroformo-alcohol isoamílico

1. Cortar el extremo del hisopo con un bisturí esterilizado y depositarlo en un tubo eppendorff.
2. Añadir a cada muestra 300µl de buffer de extracción (10 mM Tris-ClH, 100 mM NaCl, 39mM DTT, 10 mM EDTA, 2% SDS), 7,5µl de proteinasa K (10 mg/ml) y 12µl de DTT (1M).
3. Agitar a baja velocidad durante un segundo y centrifugar en una nanofuga durante 2 segundos. A continuación incubar a 56°C toda la noche (12-18horas).
4. Una vez terminada la incubación, se elimina el algodón del hisopo con una punta de pipeta, presionando contra las paredes del eppendorf para no arrastrar nada de muestra.
5. En una campana extractora, se añaden 300µl de fenol:cloroformo:isoamilalcohol 24:25:1 (saturado con 10 mM Tris, pH 8,0, 1 mM EDTA, Sigma[®]) agitar hasta conseguir una emulsión de consistencia lechosa y centrifugar en microcentrífuga 3 minutos a máximas revoluciones (13.000 rpm).
6. Una vez separada la fase orgánica y la acuosa se recupera la acuosa que contiene el ADN y se transfiere a un tubo Amicon-100 (Amicon Ultra MWCO 100kDa, Millipore[®]) al cual previamente se le han añadido 100µl de H₂O estéril.
7. Centrifugar a 3500 rpm durante 20 minutos, o en su defecto hasta que filtre.
8. Eliminar el filtrado y añadir 150 µl de H₂O estéril para limpiar la membrana.
9. Centrifugar a 3500 rpm durante 20 minutos o hasta que filtre.
10. Repetir pasos 8-9.

11. Añadir 100µl de H₂O estéril, resuspender varias veces con la pipeta, pasar a un tubo limpio invertir la membrana y centrifugar a 5500 rpm 5 minutos. Descartar la membrana y cerrar el tubo.

Protocolo de cuantificación mediante electroforesis horizontal en gel de agarosa

La electroforesis en geles de agarosa se trata de una electroforesis horizontal donde la fase estacionaria es el gel, en el que se introducirán las muestras, y la fase móvil es un buffer, conductor, que tendrá igual concentración que el gel y las muestras. Para poder visualizar los resultados se debe de añadir Gel Red™. Éste forma parte de una nueva generación de reveladores de geles de ADN y está diseñado para minimizar la interacción entre el colorante y el ADN de las células vivas.

Se han utilizado geles con concentraciones del 0,8%, y 30µl de Gel Red para 30ml de agarosa. Para poder cuantificar se han introducido muestras control de casas comerciales de concentración conocida con concentraciones entre 0,4ng/µl y 4ng/µl intentando abarcar una gama de concentraciones que nos permita obtener el valor de ADN testado en cada caso. Dentro de este umbral, intentaremos aproximarnos de la forma más precisa a la concentración de nuestra muestra.

FUNDAMENTOS DE LA ELECTROFORESIS CAPILAR

PRINCIPLES OF CAPILLARY ELECTROPHORESIS

II

La separación de los fragmentos de ADN amplificados se realiza por medio de electroforesis capilar, técnica que se emplea desde mediados de los años 90 en el campo clínico y en genética forense.

La electroforesis capilar es una técnica de separación de fragmentos de ADN por tamaño y carga basada en la diferente velocidad de migración de las distintas moléculas cargadas bajo la acción de un campo eléctrico. La separación se lleva a cabo en un capilar de sílice fundido de muy pequeño diámetro y longitud variable.

A un lado y al otro del capilar se sitúan los electrodos que van a crear la diferencia de potencial que generará el movimiento de los fragmentos de ADN. Para que pueda llevarse a cabo el análisis por electroforesis capilar es necesario que el ADN sea amplificado utilizando un par de primers o más (en el caso de los multiplex) marcados en el extremo 5' con unas moléculas llamadas fluorocromos los cuales emiten fluorescencia a diferentes longitudes de onda cuando son excitados por láser. En los kits de amplificación los fluorocromos van ligados a uno de los primers o a los nucleótidos (STRs y secuenciación respectivamente). El campo eléctrico al que se somete el producto amplificado nos permite la caracterización de fragmentos de ADN marcados con fluorocromos mediante su tiempo de retención en una matriz líquida. El capilar se rellena con un polímero a través del cual pasan los fragmentos amplificados. La aplicación de altos voltajes permite que la separación de las muestras se dé en tiempos muy rápidos, aproximadamente unos 15 minutos. Los fragmentos se separan, en función de su tamaño, en el capilar y poco a poco van saliendo por el extremo de salida donde se encuentra un detector. La fluorescencia emitida por los fragmentos, al ser excitados con un rayo de argón, es detectada a través de una cámara de detección CCD que se encarga de convertir la información proporcionada por la fluorescencia en información electrónica que se transfiere al ordenador que contiene el software necesario para traducir la fluorescencia emitida (RFUs) en variables alélicas o en secuencias de ADN que posteriormente podremos interpretar [1] (Figura II.1).



Figura II.1: El ADN se introduce dentro del capilar por donde migra. Equipos de electroforesis capilar donde se han llevado a cabo los análisis, ABI Prism 3130 (centro) y ABI Prism 310 (derecha)

Condiciones del analizador

El capilar escogido se situará dentro de un bloque donde se encuentra el polo positivo, cercano a la ventana de lectura del láser. A lo largo del capilar se creará una diferencia de potencial, DDP, que producirá el movimiento de los fragmentos a través de éste. La DDP, se establece cuando el capilar está sumergido en el buffer (10x Genetic Analyzer buffer with EDTA, Applied Biosystems®).

El polímero utilizado para el análisis de fragmentos rellenará el capilar y el bloque de metacrilato por donde pasaran los fragmentos y será la fase estacionaria que se encargará de retener las moléculas de ADN dependiendo de su tamaño.

DETECCIÓN DE FRAGMENTOS POR ELECTROFORESIS CAPILAR

Para la correcta detección de los fragmentos de ADN nuclear amplificados, se han de preparar las muestras antes de separarlas por electroforesis capilar. A cada muestra se le ha de añadir formamida y estándar de tamaño.

La formamida es una sustancia con una alta viscosidad y una alta temperatura de ebullición y de fusión. Debido a sus propiedades internas, permite evitar fenómenos de homo o heteroconjugación de la molécula disuelta en ella [2].

El estándar de tamaño evita errores en la asignación del tamaño de los fragmentos analizados debidos a diferencias en la movilidad electroforética entre las distintas inyecciones (bien por variaciones de temperatura, de concentración de polímero o en las condiciones de electroforesis). Con el estándar utilizado es posible analizar fragmentos de entre 75 y 500 pares de bases (Figura II.2).

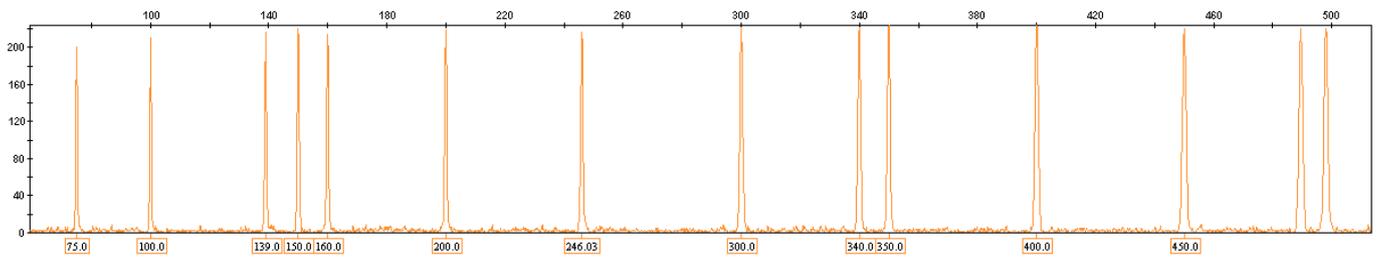


Figura 2. Estándar de tamaño LIZ 500

Al igual que se preparan las muestras, se prepara una escalera de tamaño o ladder en el que en vez añadir una muestra pondremos 1µl de ladder con todas las posibilidades alélicas que presenta el kit.

Programación de los parámetros de análisis

Entre los parámetros de mayor interés y que pueden afectar más a la hora de interpretar los resultados se encuentran, por un lado, el rango de análisis, y por otro la intensidad de los picos.

En el rango de análisis es importante especificar el rango de tamaños (en pares de bases o en puntos de datos) de los fragmentos que se van a analizar, por lo tanto este parámetro deberá de ser programado en función de los marcadores que se vayan a estudiar.

Para la intensidad de picos, normalmente se recomiendan 150 RFU, unidades de fluorescencia, para evitar picos de menor intensidad que podrían corresponder a un ADN extraño. No obstante, cuando los datos van a ser genotipados con GeneMapper se recomienda bajar la intensidad hasta 50 RFU para que este programa pueda identificar señales bajas de manera automática.

- **Hoja de muestras** ("Sample Sheet"): Al poner el listado de muestras que vamos a analizar debemos señalar los colores que queremos analizar. En este estudio (azul, verde, amarillo, rojo y naranja) y en éste último es donde se encuentra el estándar de tamaño (naranja).
- **Hoja de inyección**: En ella debemos especificar el módulo de análisis, que contiene las funciones específicas ejecutadas para el proceso de análisis de las muestras. Estos parámetros variarán en función del tipo de marcador analizado y el modelo del analizador genético empleado.

Análisis de los resultados y asignación de alelos

Una vez recogido el dato bruto o “raw data” (dato en bruto), el software *GeneMapper ID-X 1.1* [3] procesa la información para proporcionar una cuantificación en pares de bases de todos los picos. Una vez definido el estándar, se comprueba que el estándar interno de cada una de las muestras coincida. El método por el que el programa determina los tamaños de los fragmentos se denomina Southern Local (“Local Southern Method”). Dicho método utiliza tres puntos de la curva generada con los puntos del estándar: los dos que están por debajo en tamaño del fragmento a determinar y uno por encima, con dichos valores realiza una interpolación y asigna el valor del tamaño. También podemos variar el recorrido y altura cambiando las ordenadas de los ejes en los que comienza el análisis.

El software *GeneMapper ID-X 1.1* también asigna los alelos a los fragmentos analizados a través de su comparación con *ladder* alélicos. Una vez asignados los alelos a cada uno de los picos ya tenemos los resultados de las muestras que pueden ser introducidas en bases de datos para el posterior análisis de estos. El *ladder* contiene todas las posibilidades alélicas más usuales, aunque pueden aparecer nuevas expresiones de STRs, sobre todo en estudios poblacionales de zonas aisladas donde persisten.

DETECCIÓN DE SECUENCIAS POR ELECTROFORESIS CAPILAR

Al igual que para la detección de fragmentos, en la detección de secuencias las muestras han de ser tratadas antes de ser separadas por electroforesis capilar. En este caso, únicamente añadiremos formamida a la muestras cuando preveamos que puede pasar mucho tiempo entre la preparación de la muestras y el análisis. Si el análisis va a ser inmediato, se puede analizar directamente el producto de secuenciación purificado sin necesidad de añadir formamida.

Programación de los parámetros de análisis

- Hoja de muestras ("Sample Sheet"): Al poner el listado de muestras que vamos a analizar debemos señalar los colores que queremos analizar. En este estudio (azul, verde, amarillo, rojo) cada uno correspondiente a un nucleótido; guanina, adenina, citosina y timina respectivamente.

- Hoja de inyección: En ella debemos especificar el módulo de análisis, que contiene las funciones específicas ejecutadas para el proceso de análisis de las muestras. En el caso de análisis de secuencias el módulo seleccionado es el P4Rapidseq(1ml)E para el analizador genético ABI 310 y SeqSanger_POP7_50cm_KB para el analizador genético ABI 3130.

Análisis de los resultados

Una vez recogido el dato bruto o “raw data”, el software *Sequencing Analysis 5.3.1* [4] nos permite analizar, representar y editar este dato bruto. Mediante este software se asignarán los nucleótidos a cada posición en las secuencias obtenidas. Nos permite observar la calidad de la secuenciación y mejorar la lectura de la secuencia. Una vez los datos han sido analizados, el software *SeqScape® Software for Mutation Profiling v2.6* [5] permite la comparación de las secuencias a través de alineamiento simple con secuencias conocidas. Los resultados se comparan con la región control completa de la secuencia de Anderson [6] que nos permite determinar los polimorfismos que presenta cada individuo.

BIBLIOGRAFÍA

- [1] H. Shintani, J. Polonsky, Handbook of Capillary Electrophoresis Applications, Blackie Acad. Prof. (1997).
- [2] S.P. Porras, E. Kenndler, Formamide as solvent for capillary zone electrophoresis., Electrophoresis. 25 (2004) 2946–58. doi:10.1002/elps.200406040.
- [3] A. Biosystems, GeneMapper® ID-X Software Mixture Analysis Tool, v1.1 (2008).
- [4] A. Biosystems, DNA Sequencing Analysis Software v5.3.1. Part Numbre 4346366 Rev.B, 2008.
- [5] A. Biosystems, SeqScape® Software v2.6 User Guide Part Numbre 4359442 Rev.A, (2007).
- [6] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA, Nat Genet. 23 (1999) 147.

PowerStats v12

PowerStats v12 es una hoja de cálculo para el cálculo de las frecuencias alélicas así como de diversos parámetros forenses:

- Poder de discriminación.
- Poder de exclusión.
- Índice de contenido polimórfico.
- Probabilidad de coincidencia.
- Índice típico de paternidad.
- Frecuencias de homocigosidad y heterocigosidad.

Permite el análisis de hasta 600 individuos y 50 alelos, así como de distintas razas a la vez. Muestra gráficos comparativos entre las frecuencias alélicas observadas en las razas estudiadas [1].

Arlequin 3.5.1.3

Arlequin 3.5.1.3 es un paquete informático que integra un grupo de métodos básicos y avanzados para el análisis de los datos de genética de poblaciones. Éste es un programa muy versátil que permite el análisis de múltiples tipos de datos, RFLPs, secuencias de ADN, microsatélites, frecuencias alélicas o genotipos de locus múltiples [2,3].

Arlequin 3.5.1.3 permite el análisis de patrones de diversidad genética dentro y entre poblaciones.

Métodos intrapoblacionales:

- Diferentes índices genéticos estándar.
- Estimación de máxima similitud entre frecuencias alélicas y entre frecuencias haplotípicas.
- Estimación de la fase gamética de genotipos multilocus.
- Estimación de parámetros de expansiones demográficas y espaciales.

- Cálculo de medidas de desequilibrio de ligamiento y de asociación no aleatoria entre alelos de diferentes loci.
- Desviación del equilibrio de Hardy-Weinberg.
- Test de parámetros evolutivos y de neutralidad selectiva.

Métodos interpoblacionales

- Búsqueda de haplotipos compartidos entre poblaciones.
- Análisis de subdivisión de la población a través de AMOVA y cálculo de estadísticos *F*.
- Cálculo de distancias genéticas entre poblaciones.
- Test exacto de diferenciación de poblaciones.

Este programa se presenta con una interfaz compatible con el sistema Windows facilitando su manejo. Además, se han desarrollado dos versiones en formato línea de comandos, ambos para Windows y Linux, que permiten el análisis de un mayor número de muestras.

Los resultados se generan en formato XML que permite un mejor manejo de los datos, compatible con la gran mayoría de exploradores de internet; así como su análisis con funciones en R [2,3].

Phylip v3.69

PHYLIP v3.69 (*PHYLogeny Inference Package*) es un paquete de programas para inferir filogenias. Está disponible de forma gratuita en internet y apto para todo tipo de sistemas operativos [4].

A través de los distintos programas incluidos dentro del paquete se pueden realizar cálculos de parsimonia, matrices de distancias y métodos de verosimilitud como los árboles consenso y enraizados. Estos programas permiten el análisis de múltiples tipos de datos: secuencias genéticas, frecuencias genéticas, fragmentos y sitios de restricción, matrices de distancias y caracteres discretos.

Todos los programas se ejecutan a través de un sistema de línea de comandos desde donde el usuario establece las opciones de análisis.

- 1. GENDIST:** Permite el cálculo de tres matrices de distancias genéticas a través de los datos de frecuencias génicas o alélicas. Las fórmulas empleadas para el cálculo de las distancias genéticas son las de Nei [5], Cavalli-Sforza [6] y Reynolds [7]. La primera es idónea para el estudio de datos en la que las mutaciones ocurren bajo un modelo de mutación neutro. Las otras dos siguen un patrón sin mutaciones y con una deriva genética pura.

- 2. NEIGHBOR:** Se trata de una implementación de los métodos "Neighbor Joining" y UPGMA (Average Linkage clustering) de Saitou y Nei [8] por Mary Kuhner y John Yamato. Neighbor Joining es un método que produce un árbol no enraizado sin asumir un reloj molecular mientras que el método UPGMA sí lo asume. El tamaño de las ramas no está optimizado por el criterio de los mínimos cuadrados pero ambos métodos son muy rápidos y permiten el manejo de un gran número de muestras.

Disponible en <http://evolution.genetics.washington.edu/phylip.html>

IBM SPSS Statistics y STATISTICA 9.1

IBM SPSS Statistics es un software de análisis estadístico que incluye un amplio rango de procedimientos y técnicas para realizar el proceso analítico de principio a fin. SPSS Statistics permite de una forma sencilla la visualización de bases de datos y los resultados generados. Además posee herramientas para el análisis multivariante, así como de modelación y agrupación.

STATISTICA 9.1 es otro software estadístico que ofrece una gran variedad de herramientas para el análisis de datos, el manejo y visualización de estos y la obtención de toda la información posible de ellos. Incluye una gran selección de técnicas de modelación predictiva, clasificación, agrupación y análisis de los datos.

Con estos dos programas se pueden visualizar los datos derivados del estudio de distancias genéticas a través de análisis de escalamiento multidimensional y análisis de frecuencias alélicas entre poblaciones a través de análisis de correspondencias y de análisis de componentes principales.

Treewiew 1.6.6

TREEVIEW es un programa para la visualización y edición de filogenias. Este programa lee un gran número de archivos y permite representar las filogenias en múltiples formatos; radiales, cladogramas rectangulares y oblicuos y filogramas. Permite mostrar información adicional como el tamaño de las ramas e información nodal. Se puede modificar el orden de los taxones terminales así como volver a enraizar el árbol [9].

Disponible en <http://taxonomy.zoology.gla.ac.uk/rod/treewiew.html>

Structure 2.3.4

El programa STRUCTURE 2.3.4 está basado en el uso de información genética para detectar la subestructura poblacional en un grupo de individuos. Para ello se diseñó una aproximación Bayesiana, un algoritmo Monte Carlo basado en cadenas de Markov; donde se asumen un modelo de K poblaciones, cada una de las cuales está caracterizada por un grupo de frecuencias alélicas para cada locus [10]. Dado un número de clústeres (K) y asumiendo el equilibrio Hardy-Weinberg y la independencia de los loci; STRUCTURE 2.3.4 estima las frecuencias alélicas en cada clúster y la coeficiente de pertenencia a la población de cada individuo [11].

Este programa permite el análisis de múltiples tipos de datos: microsatélites, RFLPs, AFLPs o SNPs. El programa permite el análisis de loci que no se encuentran ligados así como de aquellos en los que se observa desequilibrio de ligamiento. Además, es posible asumir dos modelos de mestizaje: entre poblaciones en las que ha habido fenómenos de mestizaje como entre aquellas en las que no ha habido intercambio genético [12]. A lo largo de los años se han ido incluyendo modificaciones en las que se incorpora información sobre la localización geográfica de las muestras en las simulaciones [11].

STRUCTURE 2.3.4 permite inferir la presencia de distintas poblaciones, asignar los individuos a las poblaciones, estudiar zonas híbridas, identificar individuos migrantes y mestizos y estimar las frecuencias alélicas poblacionales en las situaciones donde mucho de los individuos han migrado o son mestizos.

El programa consiste en diversas partes. La parte computacional está escrita en C y los creadores distribuyen tanto el código fuente como ejecutables para varias plataformas. El ejecutable en C lee un archivo que ha de ser generado por el usuario, pero se han desarrollado un interfaz en Java que facilita el uso del programa así como la visualización de los resultados (Figura III.1). Disponible en <http://pritchardlab.stanford.edu/structure.html>

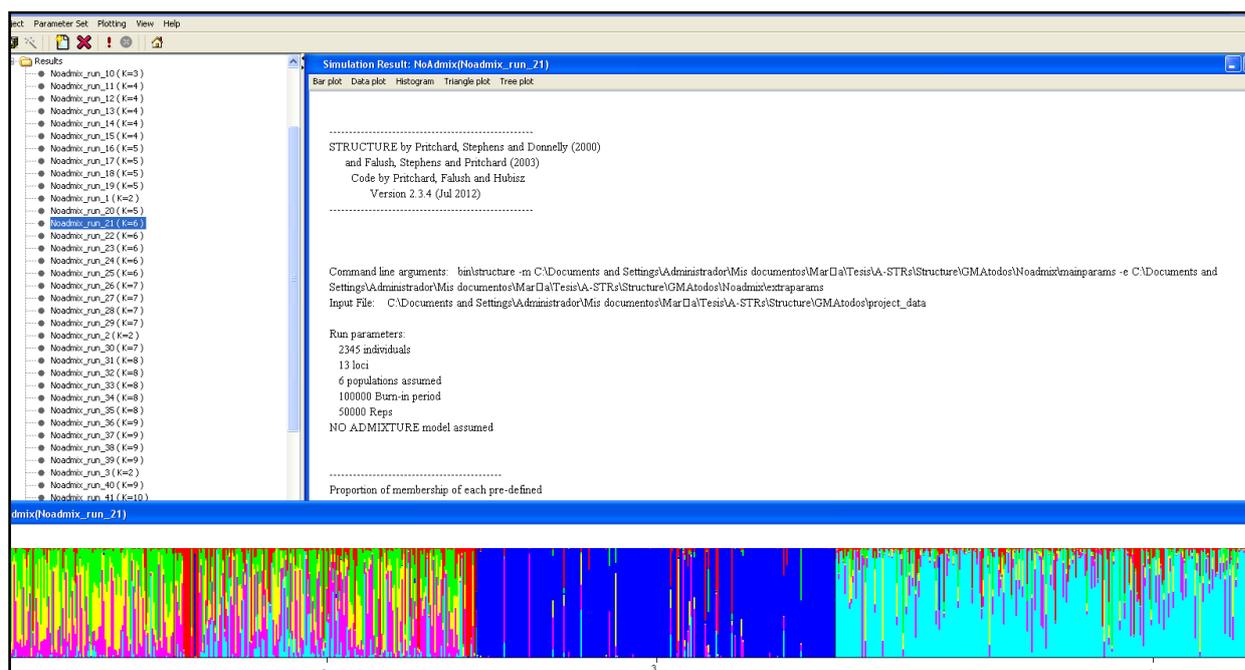


Figura III.1: Resultados de múltiples repeticiones para cada clúster en STRUCTURE 2.3.4

Clumpp 1.1.2

Actualmente se han desarrollado diversos algoritmos para asignar clústeres de individuos a través de datos genéticos poblacionales. Ya que los algoritmos de agrupamiento incorporan simulaciones estocásticas, diferentes análisis de los mismos datos generarán resultados distintos aún usando las mismas condiciones iniciales. Por ello, es necesario un método para la interpretación de los resultados de diversas réplicas [13].

Se ha desarrollado un programa, CLUMPP 1.1.2; que busca el alineamiento óptimo para todas los clústeres replicados para los mismos datos. Este programa toma como dato inicial todas las réplicas de las matrices de coeficientes de pertenecía estimados para cada clúster dando como resultado una única matriz media de las matrices estimadas en todas las réplicas [13].

El archivo de entrada es similar los resultados obtenidos por STRUCTURE 2.3.4 [10] y los resultados que arroja pueden ser usados directamente en un programa de visualización de clústeres como DISTRUCT 1.1 [14].

El programa se ejecuta en línea de comandos en todos los sistemas operativos. El programa necesita un archivo creado por el usuario con las matrices resultantes del programa STRUCTURE 2.3.4 y otro donde se definen los parámetros necesarios para que el programa funcione. Entre otros se han de especificar si los datos proporcionados son sobre poblaciones o individuos, el número de K asumidos en

la simulación, el número de poblaciones y de individuos estudiados así como el número de réplicas que se han realizado.

Disponible en <https://web.stanford.edu/group/rosenberglab/clumpp.html>

Distruct 1.1

DISTRUCT 1.1 es un programa diseñado para ser usado junto con STRUCTURE 2.3.4 y CLUMPP 1.1.2 ya que los archivos resultados de ambos programas son archivos entrada de DISTRUCT 1.1.

Una forma ideal para representar los resultados obtenidos en STRUCTURE 2.3.4 y CLUMPP 1.1.2 es una línea segmentada. Este segmento está dividido en K colores, en función de los componentes asignados; que representa los coeficientes de pertenencia de cada individuo a los clústeres estudiados. Para ello, DISTRUCT 1.1 sólo necesita los coeficientes de pertenencia a cada K para cada individuo y un identificador de grupo predefinido [14].

El programa se ejecuta en línea de comandos en todos los sistemas operativos. El programa necesita un archivo creado por el usuario con las matrices resultantes del programa STRUCTURE 2.3.4, CLUMPP 1.1.2 o cualquier otro programa de agrupamiento; y otro donde se definen los parámetros necesarios para que el programa funcione. Con este programa se pueden especificar el esquema de colores y el orden en el que los grupos o los individuos se representan así como los rótulos de datos en la parte superior e inferior del gráfico y la dirección de éstos. El resultado se da en formato PostScript y la figura resultante se puede visualizar mediante programas como GhostView [14].

Disponible en <https://web.stanford.edu/group/rosenberglab/distruct.html>

SNPAnalyzer 2.0

SNPAnalyzer 2.0 realiza análisis de desequilibrio de ligamiento específicos de muestra e implementa análisis de asociación genética a través de múltiples modelos genéticos. Se pueden analizar cientos de miles de SNPs y miles de muestras en un tiempo razonable comparado con otros programas similares [15].

Los genotipos de los SNPs bialélicos se han de codificar en un formato específico del programa indicando el nombre del marcador, el cromosoma en el que se encuentra, la localización cromosómica y el número *dbSNP rs* de cada SNP.

Cuando los datos se cargan en el programa se realiza un preprocesado de las muestras donde se indican los errores (SNPs monomórficos) y las frecuencias

alélicas y genotípicas. También se indican los resultados del test de equilibrio Hardy-Weinberg y la corrección de Bonferroni aplicada al test HWE.

Además, el programa ofrece la posibilidad de calcular diferentes parámetros:

- Estimación del haplotipo a través de algoritmos EM y PL-EM.
- Test de desequilibrio de ligamiento a través de los índices D' , r^2 , LOC score y test de cuatro gametos. Los resultados se representan en forma de bloque LD.
- Análisis de asociación entre SNPs y enfermedades a través del test de X^2 y entre haplotipos y enfermedades a través de test específicos. .

El programa tiene una interfaz en JAVA, muy cómoda para el usuario; y corre a través de un navegador de internet [15] (Figura III.2).

Disponible en <http://snp.istech21.com/snpanalyzer/2.0/>

* Marker_ID	Chr_No	Chr_Pos	dbSNP_rs	1	2	3	4	5	6	7	8	9	10	11	12
B1	20	301455066	rs34541393	T/T	A/A	A/T	A/A	T/T	A/T	A/T	A/A	A/T	A/T	A/A	A/A
B2	2	234481130	rs136624	A/T	A/T	A/T	A/A	T/T	T/T	A/A	A/A	T/T	A/T	A/T	A/T
B3	19	48896180	rs2307689	A/T	T/T	T/T	T/T	A/T	A/T	T/T	A/A	T/T	A/A	T/T	A/A
B4	8	76681235	rs35769550	A/T	A/A	T/T	A/A	A/T	A/A	T/T	A/T	A/T	A/T	A/T	T/T
B5	22	25120901	rs2307700	A/T	A/T	A/T	A/T	A/T	A/A	A/A	A/A	T/T	A/A	A/T	T/T
B6	10	6027167	rs140809	T/T	T/T	T/T	T/T	A/T	A/T	T/T	A/T	A/T	T/T	T/T	T/T
B7	1	161077452	rs3047269	A/T	A/T	T/T	T/T	A/A	A/T	A/T	A/T	A/T	A/T	T/T	A/T
B8	11	125794082	rs33972805	A/A	A/T	T/T	T/T	A/T	T/T	A/T	A/A	T/T	T/T	A/T	T/T
B9	20	11643625	rs33917182	T/T	T/T	T/T	T/T	A/T	A/A	A/A	A/A	T/T	A/A	A/A	A/T
B10	9	38396788	rs16402	A/T	A/T	T/T	A/T	A/T	T/T	A/T	A/T	T/T	A/T	A/T	A/A
G1	5	171020572	rs1610871	A/T	A/T	A/T	A/T	A/A	T/T	A/T	A/T	A/T	A/T	A/A	A/A
G2	12	113772931	rs2067238	A/T	A/A	A/T	A/A	A/A	A/T	A/T	A/A	A/T	A/T	A/T	A/A
G3	9	70504941	rs2067294	A/A	A/T	A/T	A/T	A/T	A/T	T/T	T/T	T/T	T/T	T/T	A/T
G4	6	47929222	rs2307710	A/T	T/T	A/T	A/T	A/T	A/T	A/A	A/A	A/T	A/T	T/T	A/T
G5	3	8591709	rs2308242	A/T	T/T	T/T	A/T	A/T	T/T	T/T	A/T	T/T	T/T	T/T	A/T

Allele Frequency				Genotype Frequency				Sample Types	
Allele	T	A	Total	Genotype	T/T	T/A	A/A	N/N	Total
Count	212	74	286	Obs.Count	80	52	11	0	143
Frequency	0.741	0.259	1	Exp.Count	78.5	54.9	9.6	(0)	143

HWE Test			
Chi Square	DOF	P-value	Call rate
0.387	1	0.533942	143

* Flag out: condition: 0.5 < Sample Missing, 0.1 < SNP Missing, 0.0 > SNP MAF, 0.0 > SNP HWE p-value.

Figura III.2: Visualización y preprocesado de las muestras en SNPAnalyzer 2.0

23-Haplogroup Beta Program y YPredictor by Vadim Urasin v15.0

23-Haplogroup Beta Program (On-line Software) [16,17] es una aplicación web para calcular el haplogrupo de cromosoma Y de un individuo partiendo de su perfil de STRs, sin necesidad de realizar el estudio de SNPs. Este cálculo es una aproximación, a partir de los datos reales incluidos en una base de datos extendida. Infieren a partir de su base de perfiles los resultados esperados.

Esta aplicación se basa en una aproximación Bayesiana basada en las frecuencias alélicas para predecir los haplogrupos. Gracias a este método se obtiene la probabilidad de que cada haplotipo pertenezca a un haplogrupo. Este tipo de aproximación comienza con el análisis de las frecuencias de cada

haplogrupo en el área geográfica a la que pertenece el haplotipo [17]. Estas frecuencias se llaman “probabilidades a priori”. Estas frecuencias varían entre el noroeste de Europa y Europa del Este, por lo que se puede seleccionar las probabilidades a priori que el programa utiliza para realizar las predicciones.

Disponible en <http://www.hprg.com/hapest5/>.

YPredictor by Vadim Urasin v15.0. Las predicciones se basan en los árboles filogenéticas de cada haplogrupo, calculando una probabilidad que depende de los alelos de cada marcador, la tasa de mutación del marcador y el tiempo de divergencia del nodo parental.

Disponible en <http://predictor.ydna.ru/>

Hay que tener cuidado con este tipo de programas ya que al ser un cálculo estadístico ofrece un error [18,19].

Haplogrep 2.0

HaploGrep 2.0 es una aplicación web para determinar haplogrupos de ADN mitocondrial. Está basada en Phylotree [20], una representación filogenética perfectamente estructurada del árbol de la variación humana de ADN mitocondrial. La aproximación que emplea Haplogrep para determinar el haplogrupo más probable incluye el cálculo del peso filogenético de cada mutación encontrada.

Cualquier fragmento del genoma del ADN mitocondrial se puede emplear para determinar los haplogrupos. Para cada individuo se muestran los diez haplogrupos con mejor clasificación así como la posición filogenética del mismo y el porqué de ese resultado. Además, se genera una imagen interactiva de los resultados dando recomendaciones sobre los polimorfismos que deberían de ser analizado de forma adicional para obtener mejores determinaciones. El programa da un ranking de calidad del haplogrupo predicho así como los polimorfismos esperados para cada haplogrupo encontrados en el individuo y los que se observan pero no son característicos del haplogrupo [21] (Figura III.3).

Los resultados se pueden exportar en formato texto o para ser usados posteriormente en el software de filogenias *Network*.

Actualmente existe una versión *beta* que incluye la determinación de los haplogrupos de forma automática al cargar el archivo inicial, la indicación de errores y datos a tener en cuenta en el análisis de los datos y la exportación de la secuencia completa de cada muestra en formato fasta así como una representación en forma de árbol filogenético de todas las muestras donde se puede escoger entre

los *hotspots* definidos en Phylotree o los cambios de aminoácidos para SNPs mitocondriales no usados.

Disponible en <http://haplogrep.uibk.ac.at/>

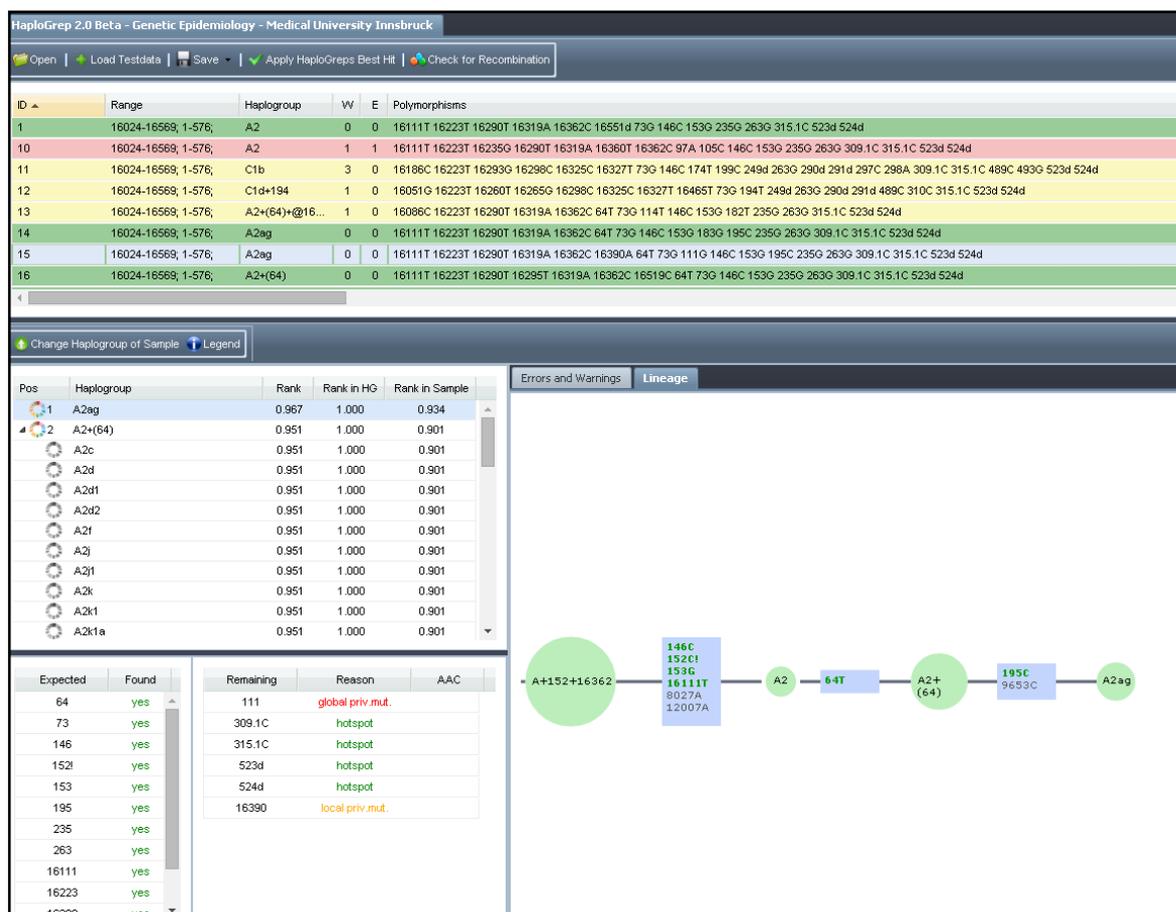


Figura III.3: Predicción de los haplogrupos mitocondriales mediante Haplogrep 2.0

Network 4.6.1.3

Network se emplea para la reconstrucción de redes y árboles filogenéticos, inferir elementos ancestrales, ramas y variantes evolutivas y estimar datas.

Los algoritmos empleados por el programa están designados para biomoléculas no recombinantes; ADN mitocondrial, STR de cromosoma Y, amonoácidos, ARN, ADN vírico, ADN bacterianos y datos no biomoleculares como datos lingüísticos.

El programa se desarrolló para la reconstrucción de todos los tipos de árboles filogenéticos de *máxima parsimonia* (MP). El programa emplea dos algoritmos diferentes en la reconstrucción filogenética:

- Algoritmo *Reduced Median* (RM) que requiere datos binarios [22].

- Algoritmo *Median-Joining* (M) que permite el análisis de datos multivariantes [23].

Disponible en <http://www.fluxus-engineering.com/sharenet.htm>

Surfer® 13

Surfer es un programa empleado en el mapeo de superficies y contornos pensado para que científicos, ingenieros, educadores puedan general mapas de forma rápida y fácil. Este programa permite modelar terrenos, generar mapas topográficos bajo el agua, visualizar paisaje, analizar superficies, mapear contornos, mapear en 3 dimensiones tanto la superficie terrestre como acuática, analizar áreas visibles, etc. Surfer transforma cualquier dato en formato XYZ en mapas de representación (Figura III.4).

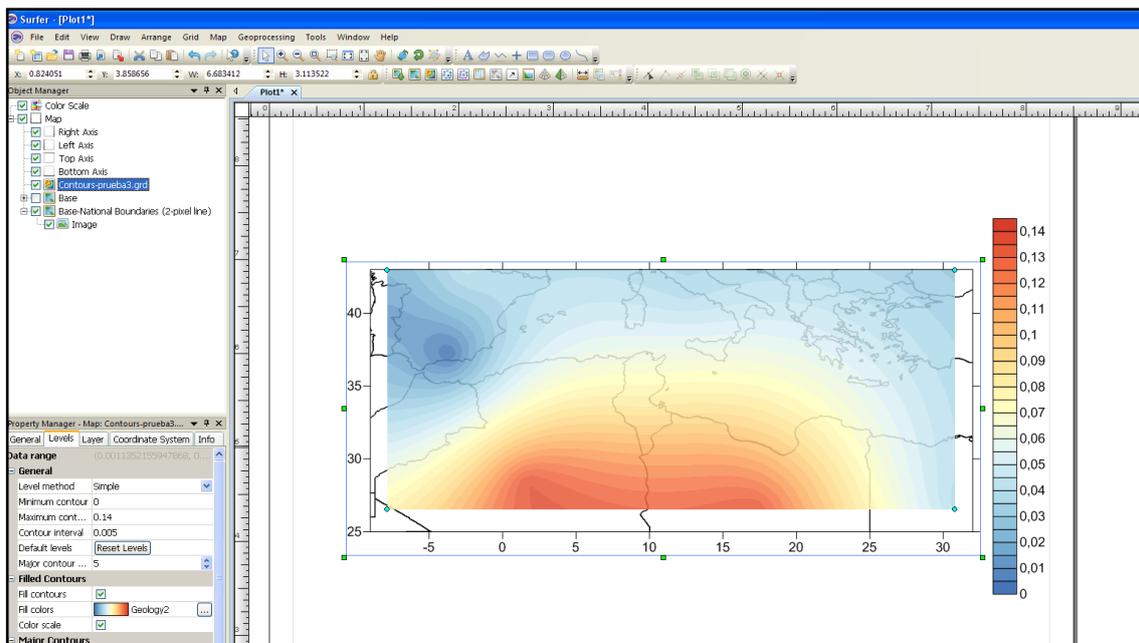


Figura III,4: Interfaz de usuario de Surfer® 13, superposición de mapa base y un mapa de contornos

BIBLIOGRAFÍA

- [1] A. Tereba, Tools for Analysis of Population Statistics, Profiles DNA. (2001).
- [2] L. Excoffier, H. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows , Mol. Ecol. Resour. 10 (2010) 564–567.
- [3] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: An integrated software package for population genetics data analysis, Evol. Bioinform. Online. 1 (2005) 47.
- [4] J. Felsenstein, PHYLIP (Phylogeny Inference Package), version 3.62 [computer program], Dep. Genome Sci. Univ. Washingt. (2004).
- [5] M. Nei, Effect of Linkage of the Genetic Load Manifested Under Inbreeding,, Genetics. (1965).
- [6] L.L. Cavalli-Sforza, M.W. Feldman, The application of molecular genetic approaches to the study of human diversity, Nat.Genet. 33 (2003) 267.
- [7] J. Reynolds, B.S. Weir, C.C. Cockerham, Estimation for the coancestry coefficient: Basis for a short-term genetic distance, Genetics. 105 (1983) 769–779.
- [8] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees., Mol. Biol. Evol. 4 (1987) 406–425.
- [9] R.D.M. Page, TREEVIEW: an application to display phylogenetic trees on personal computers, Comput.Appl.Biosci. 12 (2001) 357–358.
- [10] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data., Genetics. 155 (2000) 945–959. doi:10.1111/j.1471-8286.2007.01758.x.
- [11] M.J. Hubisz, D. Falush, M. Stephens, J.K. Pritchard, Inferring weak population structure with the assistance of sample group information, Mol. Ecol. Resour. 9 (2009) 1322–1332.
- [12] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies, Genetics. 164 (2003) 1567–1587.
- [13] M. Jakobsson, N.A. Rosenberg, CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure, Bioinformatics. 23 (2007) 1801–1806.
- [14] N.A. Rosenberg, DISTRUCT: a program for the graphical display of population structure, Mol. Ecol. Notes. 4 (2004) 137–138.
- [15] J. Yoo, Y. Lee, Y. Kim, S.Y. Rha, SNPAnalyzer 2.0: A web-based integrated workbench for linkage disequilibrium analysis and association analysis, BMC Bioinformatics. 9 (2008) 290.
- [16] T.W. Athey, Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach, J. Genet. Geneal. 2 (2006) 34–39.
- [17] T.W. Athey, Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach, J. Genet. Geneal. 1 (2005) 1–7.
- [18] M. Muzzio, V. Ramallo, J.M.B. Motti, M.R. Santos, J.S. López Camelo, G. Bailliet, Software for Y-haplogroup predictions: a word of caution, Int. J. Legal Med. (2010) 1–5.
- [19] W. Athey, Comments on the article, “Software for Y Haplogroup Predictions, a Word of Caution,” Int. J. Legal Med. (2010) 1–3.

- [20] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation., *Hum. Mutat.* 30 (2009) E386–394. doi:10.1002/humu.20921.
- [21] A. Kloss-Brandstätter, D. Pacher, S. Schönherr, H. Weissensteiner, R. Binna, G. Specht, et al., HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups, *Hum. Mutat.* 32 (2011) 25–32.
- [22] H.-J. Bandelt, A. Henkman, F. Nicolai, Powers of distance-hereditary graphs, *Discrete Math.* 145 (1995) 37–60.
- [23] H.J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, *Mol. Biol. Evol.* 16 (1999) 37–48.

Publicaciones de los datos derivados de esta tesis doctoral

- **M Saiz**, MJ Alvarez-Cubero, LJ Martinez-Gonzalez, JC Alvarez, JA Lorente. POPULATION GENETIC DATA OF 38 INSERTION-DELETION MARKERS IN SOUTH EAST SPANISH POPULATION. *Forensic Science International: Genetics*. 13 (2014) 236-238

Otras publicaciones donde se emplean datos derivados de esta tesis doctoral

- **M Saiz**, MJ Alvarez-Cubero, JC Alvarez, LJ Martinez-Gonzalez, JA Lorente. ACTION PROTOCOLS IN DNA IDENTIFICATION OF ISOLATED POPULATIONS. *Journal of Forensic Research*. 5 (2014) 10000218



Forensic Population Genetics - Letter to the Editor

Population genetic data of 38 insertion–deletion markers in South East Spanish population

Dear Editor,

Insertion–deletion polymorphisms (INDELs) are usually diallelic markers derived from a single mutation event. INDELs can combine characteristics of both SNPs and STRs, as an extended distribution throughout the genome [1,2]; the possibility of the use of reduced amplicon sizes improving amplification success of compromised samples [3–5]; origination from a single mutation event occurring at a low frequency and unlikely to present recurrent mutations; and the effortlessness of analysis through PCR and capillary electrophoresis, as STRs [6–8]. These beneficial characteristics make them suitable for forensic and parentage testing [6,8–10]. Furthermore, the significant differences in allele frequencies among geographically separated population groups [11,12] make some INDELs good ancestry informative markers.

According to the Spanish Statistics Office 2011 census, the South East Spanish population consists of 3,205,644 inhabitants, of which 28.76% are found in the province of Granada, 49.75% in the province of Málaga and 21.48% in the province of Almería (<http://www.ine.es/>). These three provinces occupy an area of 28,613 km², 5.67% of the Spanish territory.

The purpose of this study was to characterize the diversity of 38 INDEL markers in the South East Spanish population, provinces of Granada, Málaga and Almería; and to assess their usefulness both in forensic casework and population genetics.

Buccal cell swabs from a total of 143 unrelated adults from the South East Spanish population, [Almería (45), Granada (53), Málaga (45)] were collected under informed consent and the study was approved by the ethics committee of the University of Granada.

Genomic DNA was isolated using an organic extraction procedure by phenol/chloroform/isoamyl alcohol and proteinase K. It was purified by Amicon[®] 100 (Millipore). Quantification of extracted DNA was performed by 0.8% agarose gel.

The samples were amplified using a 38 INDELs multiplex, according to Pereira et al. [13]. PCR was carried out in a GeneAmp[®] PCR System 2720 thermal cycler (Applied Biosystems, California) under the following conditions: initial incubation at 95 °C for 15 min; 10 cycles at 94 °C for 30 s, 60 °C for 90 s and 72 °C for 60 s; 20 cycles at 94 °C for 30 s, 58 °C for 90 s and 72 °C for 60 s; with a final extension at 72 °C for 60 min.

Amplification products were separated by electrophoresis in an Applied Biosystems ABI 3130 genetic analyzer and allele calls were determined automatically with GeneMapperX v1.1 software (Applied Biosystems) using the bin sets for INDELs provided by Pereira et al. [13]. Typing quality and allele designation were warranted by simultaneous electrophoretic analysis of a control

sample of known genotype (profile information available in Supplementary Table 1).

Supplementary Table 1 related to this article can be found, in the online version, at [doi:10.1016/j.fsigen.2014.08.009](https://doi.org/10.1016/j.fsigen.2014.08.009).

Allelic frequencies, heterozygosity (H), polymorphism information content (PIC), power of discrimination (PD), power of exclusion (PE) and deviation probability from the Hardy–Weinberg equilibrium (p) as well as genetic differentiation and genetic distance (F_{st}) coefficients for the population of Granada, Málaga and Almería were calculated using the Arlequin v3.11 software [14]. Matching probability (MP) and typical paternity index (TPI) were calculated for each locus using the PowerStats v12 software [15]. The statistical significance of linkage disequilibrium among loci pairs was measured by exact tests to estimate p values using Arlequin 3.11 software and SNPAnalyzer 2 [16].

In order to assess the South East Spanish genetic substructure, Arlequin v3.11 software was also used to test the hypothesis of a random distribution of individuals between the pairs of populations with an exact test of population differentiation and to perform AMOVA [17]. STRUCTURE 2.3.4 [18–20] analysis was performed to infer clusters of individuals on the basis of genetic data from unlinked loci alone. For all simulations and calculations, an admixture and non-admixture model was assumed with a burn-in period of 50,000 iterations followed by an additional 100,000 iterations ($K = 1–4$) within the South East Spanish population. Each run was carried out 10 times and posterior probabilities for each K were computed for each set of runs.

Allele frequencies and statistic parameters for the 38 INDEL loci of the South East Spanish population are shown in Table 1. All loci analyzed were in Hardy–Weinberg equilibrium (HWE) in the studied population ($p > 0.05$), with three exceptions (rs36040336, rs10688868 and rs2308242). After applying Bonferroni's correction, the genotype frequency distribution showed no deviations for the HWE ($p > 0.00132$). The forensic parameter averages calculated were: polymorphism information content (PIC) = 33.85%; power of discrimination (PD) = 58.11%; power of exclusion (PE) = 13.57% and observed heterozygosity (H_o) = 42.19%. The rs3047269, rs33972805 and rs1610871 markers showed the highest level of heterozygosity (53.85%) and the rs36040336 marker showed the lowest (24.48%). The combined power of discrimination and the power of exclusion for the 38 INDELs studied were $1–3.5496 \times 10^{-15}$ and 99.630617% respectively. Linkage disequilibrium test showed no significant allelic association between all 38 markers, so they could be treated as independent ones.

AMOVA analysis shows no significant substructure within the three provinces of the South East Spanish population ($p > 0.05$; see Supplementary Table 2). Furthermore, these results were corroborated by the pairwise genetic distances analysis (see Supplementary Table 3), as well as the exact tests of population differentiation

Table 1

Forensically relevant diversity parameters for 38 INDELS in the Granada, Málaga and Almería population.

rs number	Code	Short allele	Long allele	H_o	H_e	PIC	PD	PE	MP	TPI	p
rs34541393	B01	0.4056	0.5944	0.5175	0.4864	0.3659	0.5980	0.2033	0.4020	1.0362	0.4939
rs16624	B02	0.7448	0.2552	0.3706	0.3600	0.3079	0.5429	0.1048	0.4571	0.8125	0.8172
rs2307689	B03	0.3427	0.6573	0.4476	0.4564	0.3490	0.5976	0.1456	0.4024	0.9051	0.8590
rs35769550	B04	0.3671	0.6329	0.4196	0.4644	0.3567	0.6183	0.1309	0.3817	0.8720	0.2817
rs2307700	B05	0.5769	0.4231	0.4126	0.4919	0.3690	0.6418	0.1309	0.3582	0.8720	0.0573
rs140809	B06	0.2483	0.7517	0.3427	0.3746	0.3036	0.5398	0.0826	0.4602	0.7606	0.3690
rs3047269	B07	0.4476	0.5524	0.5385	0.4955	0.3722	0.5937	0.2305	0.4063	1.1000	0.3189
rs33972805	B08	0.4790	0.5210	0.5385	0.5009	0.3746	0.6027	0.2234	0.3973	1.0833	0.4063
rs33917182	B09	0.6014	0.3986	0.4476	0.4839	0.3645	0.6265	0.1456	0.3735	0.9051	0.3884
rs16402	B10	0.2587	0.7413	0.3636	0.3849	0.3100	0.5489	0.0933	0.4511	0.7857	0.5136
rs1610871	G01	0.5350	0.4650	0.5385	0.4993	0.3738	0.6011	0.2234	0.3989	1.0833	0.4015
rs2067238	G02	0.6434	0.3566	0.4615	0.4605	0.3536	0.6009	0.1560	0.3991	0.9286	1.0000
rs2067294	G03	0.2972	0.7028	0.4196	0.4220	0.3305	0.5750	0.1218	0.4250	0.8512	1.0000
rs2307710	G04	0.3881	0.6119	0.5245	0.4766	0.3622	0.5868	0.2098	0.4132	1.0515	0.2855
rs2308242	G05	0.2203	0.7797	0.2727	0.3447	0.2845	0.5069	0.0582	0.4931	0.7010	0.0277
rs2307580	G06	0.4930	0.5070	0.4685	0.5017	0.3750	0.6392	0.1614	0.3608	0.9408	0.5109
rs1160956	G07	0.8252	0.1748	0.3217	0.2895	0.2469	0.4550	0.0728	0.5450	0.7371	0.2508
rs34511541	G08	0.4545	0.5455	0.4615	0.4976	0.3729	0.6379	0.1560	0.3621	0.9286	0.4054
rs2307978	G09	0.2273	0.7727	0.3427	0.3525	0.2896	0.5178	0.0826	0.4822	0.7606	0.8128
rs3051300	Y01	0.3636	0.6364	0.4126	0.4663	0.3557	0.6183	0.1263	0.3817	0.8614	0.2018
rs10629077	Y02	0.2483	0.7517	0.3566	0.3815	0.3036	0.5391	0.0896	0.4609	0.7772	0.5114
rs10688868	Y03	0.2797	0.7203	0.3217	0.4044	0.3218	0.5694	0.0728	0.4306	0.7371	0.0228
rs2067208	Y04	0.2902	0.7098	0.3566	0.4134	0.3271	0.5778	0.0896	0.4222	0.7772	0.1068
rs2307579	Y05	0.5175	0.4825	0.4755	0.5011	0.3747	0.6295	0.1785	0.3705	0.9795	0.6275
rs2308020	Y06	0.6538	0.3462	0.4546	0.4543	0.3502	0.5973	0.1507	0.4027	0.9167	1.0000
rs3080855	Y07	0.3077	0.6923	0.4266	0.4248	0.3353	0.5776	0.1357	0.4224	0.8827	1.0000
rs1610919	Y08	0.5280	0.4720	0.5175	0.4998	0.3742	0.6180	0.1968	0.3820	1.0214	0.7381
rs2307839	Y09	0.2413	0.7587	0.3636	0.3710	0.2991	0.5307	0.0970	0.4693	0.7944	0.8248
rs2308137	R01	0.2832	0.7168	0.4615	0.4163	0.3236	0.5506	0.1507	0.4494	0.9167	0.2219
rs36040336	R02	0.8077	0.1923	0.2448	0.3117	0.2624	0.4655	0.0432	0.5345	0.6620	0.0147
rs1160886	R03	0.4266	0.5734	0.4476	0.4909	0.3696	0.6363	0.1456	0.3637	0.9051	0.3073
rs2308026	R04	0.4091	0.5909	0.4406	0.4852	0.3666	0.6329	0.1406	0.3671	0.8938	0.3066
rs2307526	R05	0.3147	0.6853	0.3706	0.4302	0.3383	0.5950	0.1009	0.4050	0.8034	0.1187
rs34811743	R06	0.6259	0.3741	0.4825	0.4700	0.3587	0.6016	0.1727	0.3984	0.9662	0.8576
rs2308189	R07	0.3741	0.6259	0.4965	0.4700	0.3587	0.5950	0.1845	0.4050	0.9931	0.5915
rs5895447	R08	0.3392	0.6608	0.4546	0.4498	0.3478	0.5929	0.1507	0.4071	0.9167	1.0000
rs2308171	R09	0.2448	0.7552	0.3846	0.3746	0.3014	0.5334	0.1009	0.4666	0.8034	0.8224
rs35605984	R10	0.3042	0.6958	0.3566	0.4248	0.3337	0.5879	0.0970	0.4121	0.7944	0.0776

H_o , observed heterozygosity; H_e , expected heterozygosity; PIC, polymorphism information content; PD, power of discrimination; PE, power of exclusion; MP, matching probability; TPI, typical paternity index; p , p value for Hardy–Weinberg equilibrium.

($p > 0.05$) and the clustering program STRUCTURE 2.3.4. There was no evidence of presence of significant genetic structure within the South East Spanish population corresponding to the geographical clusters related to provinces. The model with the highest posterior probability value was $K = 1$ ($\ln P(D) = -6822.76$) implying that the genetic data favored a single cluster containing all the South East Spanish population.

Supplementary Tables 2 and 3 related to this article can be found, in the online version, at [doi:10.1016/j.fsigen.2014.08.009](https://doi.org/10.1016/j.fsigen.2014.08.009).

This study presents a new database using 38 indels for the South East Spanish population of Granada, Málaga and Almería. According to the results obtained it can be stated that these markers represent a useful new tool in forensic analysis in this population. It is recommended to conduct similar studies in other worldwide populations in order to assess the usefulness of these markers in molecular anthropology studies.

This paper follows the publication guidelines of population data requested by the journal [21].

Acknowledgements

The content of this article is part of the Ph.D. thesis of María Saiz which was conducted at the University of Granada under the doctoral program “Biomedicine”. Furthermore, the authors thank all the participants who donated buccal swabs and all the people who helped in the sample collection, namely María Luisa Aceituno Villalva, Leticia Olga Rubio Lamia and Verónica Pérez Delgado.

References

- [1] D.M. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, A. Chakravarti, A.G. Clark, et al., An integrated map of genetic variation from 1,092 human genomes, *Nature* 491 (2012) 56–65.
- [2] S.B. Montgomery, D.L. Goode, E. Kvikstad, C.A. Albers, Z.D. Zhang, X.J. Mu, et al., The origin, evolution, and functional impact of short insertion–deletion variants identified in 179 human genomes, *Genome Res.* 23 (2013) 749–761.
- [3] R. Pereira, L. Gusmão, Capillary electrophoresis of 38 noncoding biallelic mini-indels for degraded samples and as complementary tool in paternity testing, *Methods Mol. Biol.* 830 (2012) 141–157.
- [4] C. Romanini, M.L. Catelli, A. Borosky, R. Pereira, M. Romero, M. Salado Puerto, et al., Typing short amplicon binary polymorphisms: supplementary SNP and indel genetic information in the analysis of highly degraded skeletal remains, *Forensic Sci. Int. Genet.* 6 (2012) 469–476.
- [5] C. Hollard, F. Mendisco, C. Keyser, E. Crubézy, B. Ludes, First application of the investigator DIPplex indels typing kit for the analysis of ancient DNA samples, *Forensic Sci. Int. Genet. Suppl. Ser. 3* (2011) e393–e394.
- [6] R. Pereira, C. Phillips, C. Alves, A. Amorim, A. Carracedo, L. Gusmão, Insertion/deletion polymorphisms: a multiplex assay and forensic applications, *Forensic Sci. Int. Genet. Suppl. Ser. 2* (2009) 513–515.
- [7] J. Huang, H. Luo, W. Wei, Y. Hou, A novel method for the analysis of 20 multi-indel polymorphisms and its forensic application, *Electrophoresis* 35 (2014) 487–493.
- [8] K. Oka, M. Asari, T. Omura, M. Yoshida, C. Maseda, D. Yajima, et al., Genotyping of 38 insertion/deletion polymorphisms for human identification using universal fluorescent PCR, *Mol. Cell. Probes* 28 (2014) 13–18.
- [9] A. Zidkova, A. Horinek, V. Kebrdlova, M. Korabecna, Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population, *Int. J. Legal Med.* 127 (2013) 7–10.
- [10] F. Manta, A. Caiafa, R. Pereira, D. Silva, A. Amorim, E.F. Carvalho, et al., Indel markers: genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material, *Forensic Sci. Int. Genet.* 5 (2012) 658–661.

- [11] R. Pereira, C. Phillips, N. Pinto, C. Santos, S.E.B. dos Santos, A. Amorim, et al., Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing, *PLOS ONE* 7 (2012) e29684.
- [12] N.P.C. Santos, E.M. Ribeiro-Rodrigues, Â.K.C. Ribeiro-dos-Santos, R. Pereira, L. Gusmão, A. Amorim, et al., Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel, *Hum. Mutat.* 31 (2010) 184–190.
- [13] R. Pereira, C. Phillips, C. Alves, A. Amorim, A. Carracedo, L. Gusmão, A new multiplex for human identification using insertion/deletion polymorphisms, *Electrophoresis* 30 (2009) 3682–3690.
- [14] L. Excoffier, H. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [15] A. Tereba, Tools for Analysis of Population Statistics, Profiles in DNA, 2001.
- [16] J. Yoo, Y. Lee, Y. Kim, S.Y. Rha, Y. Kim, SNPAnalyzer 2.0: a web-based integrated workbench for linkage disequilibrium analysis and association analysis, *BMC Bioinformatics* 9 (2008) 290.
- [17] M. Raymond, F. Rousset, An exact test for population differentiation, *Evolution* 41 (1995) 823–833.
- [18] M.J. Hubisz, D. Falush, M. Stephens, J.K. Pritchard, Inferring weak population structure with the assistance of sample group information, *Mol. Ecol. Resour.* 9 (2009) 1322–1332.
- [19] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, *Genetics* 164 (2003) 1567–1587.
- [20] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, *Genetics* 155 (2000) 945–959.
- [21] A. Carracedo, J.M. Butler, L. Gusmão, W. Parson, L. Roewer, P.M. Schneider, Publication of population data for forensic purposes, *Forensic Sci. Int. Genet.* 4 (2010) 145–147.

María Saiz^{*}
*Laboratory of Genetic Identification, Department of Legal Medicine,
 Toxicology and Physical Anthropology, Faculty of Medicine, University
 of Granada, Avd. Madrid 11, 18071 Granada, Spain*

Maria Jesus Alvarez-Cubero
*Centre of Molecular Anthropology for Ancient DNA Studies,
 Dipartimento di Biologia, Università di Roma "Tor Vergata",
 Rome, Italy*

Luis Javier Martinez-Gonzalez
*Pfizer-University of Granada-Junta de Andalucía Centre for Genomics
 and Oncological GENYO, Biomedical Research Centre, Avda. del
 Conocimiento s/n, 18100 Armilla, Granada, Spain*

Juan Carlos Alvarez
*Laboratory of Genetic Identification, Department of Legal Medicine,
 Toxicology and Physical Anthropology, Faculty of Medicine, University
 of Granada, Avd. Madrid 11, 18071 Granada, Spain*

José Antonio Lorente^{a,b}
^a*Laboratory of Genetic Identification, Department of Legal Medicine,
 Toxicology and Physical Anthropology, Faculty of Medicine, University
 of Granada, Avd. Madrid 11, 18071 Granada, Spain*
^b*Pfizer-University of Granada-Junta de Andalucía Centre for
 Genomics and Oncological GENYO, Biomedical Research Centre, Avda.
 del Conocimiento s/n, 18100 Armilla, Granada, Spain*

^{*}Corresponding author. Tel.: +34 958249950; fax: +34 958246107
 E-mail addresses: msaiz@ugr.es,
marisaiz85@gmail.com (M. Saiz).

Received 22 May 2014
 12 August 2014 Accepted 14 August 2014

Action Protocols in DNA Identification of Isolated Populations

Maria Saiz MS^{1#}, Maria J Alvarez-Cubero^{1#}, Juan C Alvarez^{1,2}, Luis J Martinez-Gonzalez^{2*} and Jose A Lorente^{1,2}¹Laboratory of Genetic Identification, Dept of Legal Medicine, Universidad de Granada, Granada, Spain²GENyO Centre for Genomics and Oncological Research: Pfizer - Universidad de Granada - Junta de Andalucía, Spain[#]Authors contributed equally to this work

Abstract

This paper presents the basic problems and difficulties that can be found when working with isolated populations as far as propose some approaches in the analysis of this type of populations. 13 autosomal STRs have been analyzed and statistical forensic parameters, such as observed heterozygosity and power of discrimination have been determined, in samples from isolated and non-isolated populations. Samples were amplified with AmpFISTR®Identifiler® kit (Applied Biosystems) and PowerPlex 16®kit (Promega). For DNA typing, an ABI PRISM 310 Genetic Analyzer was used, and the analysis was performed with GeneMapper ID-X Software v1.1. PowerStats software and SPSS v15.0 were used to calculate forensic and other parameters. The analysis is based on the comparison of three main population groups (large, medium, and small), and an estimation of the forensic parameters, such as discrimination power (PD), observed heterozygosity (Ho), and combined PD, was obtained. The obtained results reveal that heterozygosity and PD are lower in aboriginal populations than in other populations. This research contributes to defining the decrease in allele presence caused by the lower size of a population as far as endogamy processes. To obtain an efficient human identification, it is necessary to generate independent databases for all the different populations, including small isolated groups. Action protocols to these kinds of populations have to be adapted: type as many markers as possible, not only autosomal STR but also mitochondrial DNA and sexual chromosome markers and characterize those one that better describe the population, and also get genetic information of close relatives as far as direct references of the individual.

Keywords: Discrimination power; Identification; Isolated populations; Observed heterozygosity; STRs

Introduction

The most common use of STRs is in forensic cases [1]. However, STR population data has also been used for admixture estimation and structure inference in different populations [2,3]. When a sample is identified with a specific genetic profile, the probability of discrimination can differ depending on the laboratory where it was processed and on the STR markers chosen and their population frequency. Heterozygosity and homozygosity are both parameters that show the polymorphism and efficacy of a genetic marker.

Even today, some parts of the world are still isolated due to geographic, linguistic, or cultural factors. Endogamy is characteristic of isolated groups which creates a problem in forensic casework. Loss of heterozygosity and the presence of same haplotypes are found at a higher proportion in these populations because of aboriginal identity preservation. Most relationships occur among the community members, and this phenomenon of endogamy usually continues within the next generations. Furthermore, there is a strong tradition of endogamy and a preference for consanguineous unions. Although there are a lot of similarities in terms of anthropology, geographic ethnicity, or culture, the populations may still exhibit major genetic differentiation [4]. When working with forensic samples coming from isolated and small populations, some problems in genetic identification can be found even when a direct relative is available.

Materials and Methods

Subjects and samples

Three kinds of populations were used in this essay: large populations living in a vast territory, medium or large populations living in small territory, and small isolated populations.

This project was carried out using the genetic data of 13 autosomal short tandem repeats (STRs) polymorphisms in three classes of populations (small isolated populations, medium populations, and large populations). The samples were analyzed with different kits; hence, the project was reduced to 13 CODIS markers (D3S1358, TH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1PO, vWA, D8S1179, TPOX, and FGA). The small isolated populations are Chakma, Čakovec, Shiite Muslim, OtomiIxmiquilpan, Huastecos, AdiPanggi, AdiKomkar, Kaqchiquel, Kiche, Mam, and Qeqchi. The medium populations are compounded by Liaoning, Kuala Lumpur, Dobruja, Northern Poland, Karamoja, Rio de Janeiro, East Timor, Granada, and Ovambo. The big populations are Macedonia, Morocco, Somalia, Guatemalan Mestizo, Spain, Mexico-Chihuahua, Paraguay, Costa Rica, Brazil, and Jalisco. Some of these populations, such as Granada, Guatemalan Mestizo, Spain, Paraguay, Ecuador, Costa Rica, Brazil, Mexico-Chihuahua, Kaqchiquel, Kiche, Mam, and Qeqchi, were analyzed and processed in the Laboratory of Genetic Identification (University of Granada, Spain). The others were collected from previously published frequency data (For more information about the populations see supplementary data Table S1).

***Corresponding author:** Luis J Martinez-Gonzalez, GENyO. Centre for Genomics and Oncological Research: Pfizer - Universidad de Granada - Junta de Andalucía, Avd. Ilustración 114 Granada, Granada 18016, Spain, Tel: +34958715501; Fax: +34 958 637 071; E-mail: luisjavier.martinez@genyo.es

Received December 30, 2013; **Accepted** February 07, 2014; **Published** February 14, 2014

Citation: Maria Saiz MS, Alvarez-Cubero MJ, Alvarez JC, Martinez-Gonzalez LJ, Lorente JA (2014) Action Protocols in DNA Identification of Isolated Populations. J Forensic Res 5: 218 doi:[10.4172/2157-7145.1000218](http://dx.doi.org/10.4172/2157-7145.1000218)

Copyright: © 2014 Maria Saiz MS, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Sample collection followed the recommendations of GEP-ISFG, and all the samples were collected under the informed consent of the donors. Each single project used in this article has been approved by the Ethical Committee.

Populations

Special focus is given to isolated populations, which are all characterized by cultural, linguistic, and/or geographical factors (Figure 1).

Some population groups, like Kaqchiquel, Kiche, Mam, Qeqchi, and Huastecos, are Mayan tribes. There are more than twenty Mayan languages and a huge variety of populations groups that explains the high grade of isolation present in these Mayan groups [5]. Most of these groups belong to the same language family, Mayan Quichean-Mamean, although they preserve their native language [6]. K'iche', Kaqchikel and Q'eqchi' are into the Greater Quichean language group while Mam language is situated in the Greater Mamean group and Huastecos, another Mexican tribe, call their language "Teenek" [7].

The other isolated populations also have their own characteristics that help to preserve their aboriginal identity. Chakmas, an ethnic group from Bangladesh, speaks Changma [8]; OtomiIxmiquilpan, a Mexican tribe, continues to use their own language, Hñahñu [7]; and AdiPanggi and AdiKomkar are dialects of Adi, a language spoken in India, China, and Bhutan [9]. Although they abandoned some traditional customs and trades, the Romani ethnic minority Čakovec kept a distinct archaic dialect of the Romanian (*Limba d' bjaš*) as a predominant language [10]. Shiite Muslims from Tamilnadu, India, besides speaking the local languages, have their own language called *Lisānu l-Dā'wat*, "The language of the Dā'wat". All of these populations also have their own culture and customs, and profess the same religion. Geographical isolation is also present in these and other similar groups. While Mayan tribes reside in the highlands of Guatemala, Chakmas live near the foothills of the Himalayas; AdiPanggi and AdiKomkar are part from little populations of Tibet, and OtomiIxmiquilpan inhabit a semi-deserted area at an altitude of 1700-1800 m.

Besides linguistic, cultural, and geographical studies, genetic analyses have also confirmed that aboriginal and isolated populations

are closely related and are more similar in terms of population diversity [11].

DNA analysis

The samples were amplified with AmpFlSTR®Identifiler® kit (Applied Biosystems, Foster City, CA) [12] and a GeneAmp System 2400 thermal cyclers. Some populations, like Paraguay, Costa Rica, and Brazil, were amplified with PowerPlex 16®kit (Promega) [13]. A total of 13 markers were chosen for the statistical analysis (D3S1358, TH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1PO, vWA, D8S1179, TPOX, and FGA). Genotyping of nuclear STRs was achieved using an ABI PRISM® 310 Genetic Analyzer (Applied Biosystems). Fragment sizes were analyzed by GeneScan v3.1 [14] and GeneMapper® ID-X Software v1.1 (Applied Biosystems) [15].

Statistic analysis

Allele frequencies were determined by direct counting. Gene diversity and haplotype diversity were calculated according to Nei [16,17]. All the forensic statistical parameters were recalculated with the PowerStats software [18]. Other statistic parameters, such as power of discrimination (PD; medium, maximum, and minimum), observed heterozygosity (Ho; maximum and minimum), and combined PD (-log 10 (1-combined PD)), were calculated by SPSS v15.0.

Results

In order to compare the three classes of populations, the forensic parameters were estimated. High values of PD were obtained in large (0.915 ± 0.004) and medium populations (0.916 ± 0.007) in contrast to small isolated populations (0.847 ± 0.055). Moreover, higher Ho values were found in big (0.783 ± 0.007) and medium (0.786 ± 0.013) populations compared to those in small isolated populations (0.7662 ± 0.030). These results were not as different as those in PD, but isolated populations still had the lowest values. The differences in PD and Ho between small isolated populations, and medium and large populations are significant (Figure 2), more details in Table 2.

The values of the minimum and maximum Ho achieved the biggest differences in isolated populations, whereas in large and medium populations, these values had lesser differences between them

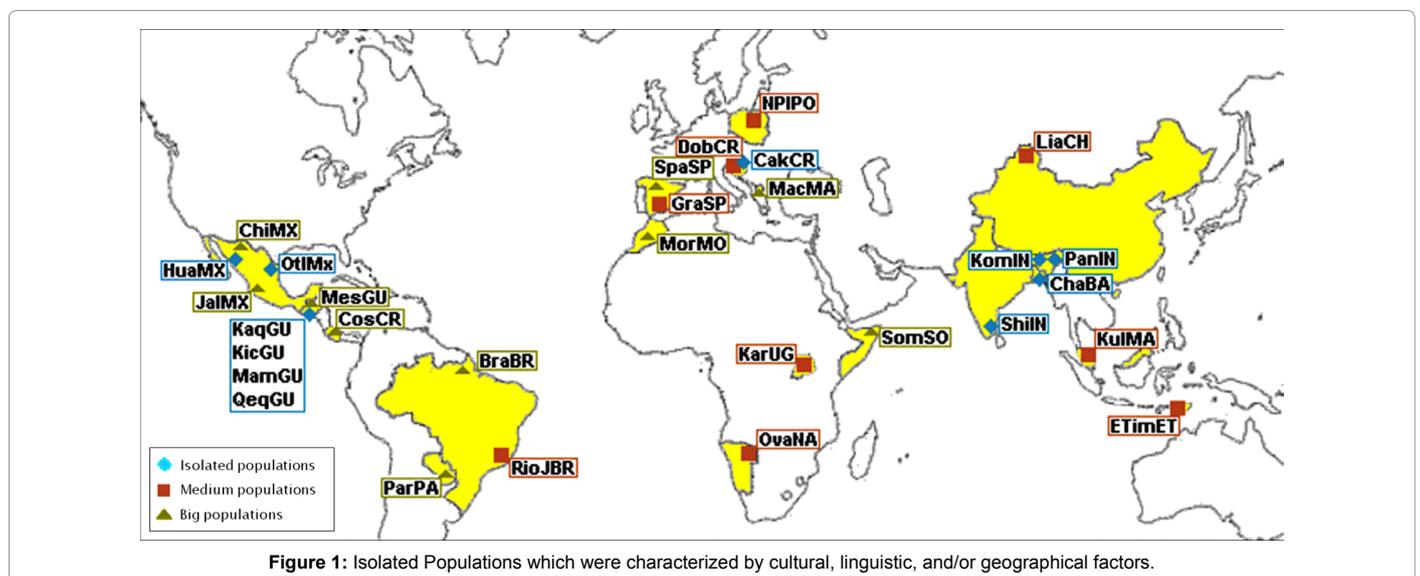


Figure 1: Isolated Populations which were characterized by cultural, linguistic, and/or geographical factors.

(Figure 3). Similar results have been obtained in the analysis of PD. Studies about these types of populations have revealed an increase in homozygosity and a decrease in genotypic variability. Thus, the power of discrimination, typical paternity index, and power of exclusion are lower, too (data not shown).

If we present combined PD, the differences can be also noticed. As

in previous analyses and representations, the lowest values are found in isolated populations, while the highest values are achieved in medium and big populations (Figure 4).

Furthermore, the gene diversity is also different among these populations. If PD is taken into account, whereas TH01 has the less discrimination capacity and FGA the highest in isolated populations,

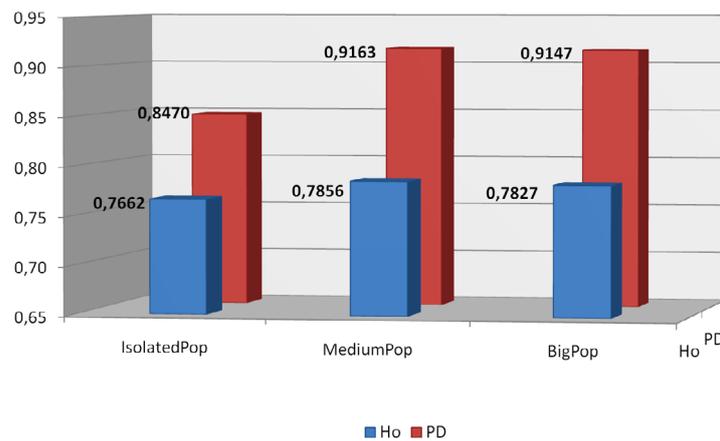


Figure 2: The differences in PD and Ho between small isolated populations, and medium and large populations.

Sample Info	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH0	D13S31	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Case 1 Mother	12,13	29,29	12,12	9,11	14,15	6,6	9,12	9,11	19,22	13,13	16,17	8,11	14,14	9,12	23,25
Case 1 Minor	12,14	29,31.2	12,12	11,12	14,15	6,9	9,9	9,11	22,23	13,15	16,17	8,11	14,15	11,12	24,25
Case 1 Father	11,12	30,31	12,12	11,12	15,18	9,9	9,11	9,13	22,23	13,13.2	13,17	11,12	14,16	12,13	22,25
Case 2 Mother	14,15	29,31.2	9,10	7,10	15,15	6,7	8,11	11,12	19,23	13,14	16,18	8,12	12,16	11,12	18.2,25
Case 2 Minor	13,14	29,31.2	10,11	10,12	15,15	6,6	11,12	11,12	22,23	13,15.2	15,16	12,12	16,21	11,12	18.2,23
Case 2 Father	12,13	29,33.2	11,11	9,10	15,16	6,9.3	11,12	10,12	20,23	13,13.2	16,18	11,12	15,18	11,11	21,23

In italics and bold, markers that exclude the father when studying minor and father. Only in bold, those markers that exclude the father when studying mother, minor and father

Table 1: Genetic profiles of mother, minor and father in two paternity cases.

Sample Info	PD p-value	OR(95%CI)	Ho p-value	OR(95%CI)
Isolated vs. large populations	0.0001*	-0.0676(-0.0839- -0.0513)	0.0753	-
Isolated vs. medium populations	0.001*	-0.0693(-0.0859- -0.0526)	0.0474*	-0.0194(-0.0385- -0.0002)
Medium vs. large populations	0.0737	-	0.7303	-

*This difference is considered statistically significant.

vs: Versus

OR: Odds Ratio

CI: Confidence Interval

Table 2: PD and Ho statistical analysis.

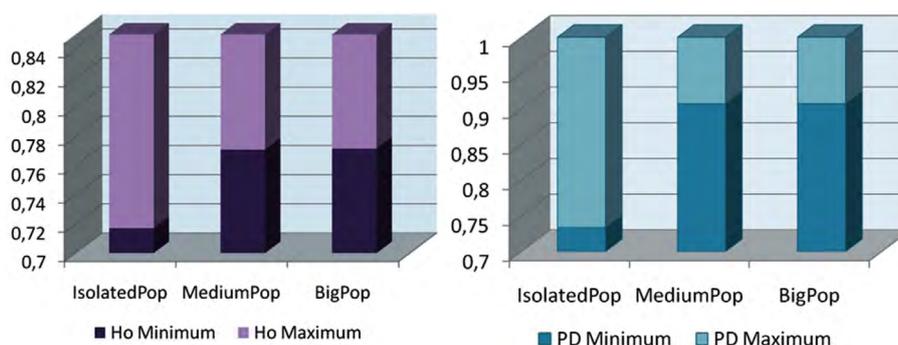


Figure 3: The values of the minimum and maximum Ho achieved the biggest differences in isolated populations.

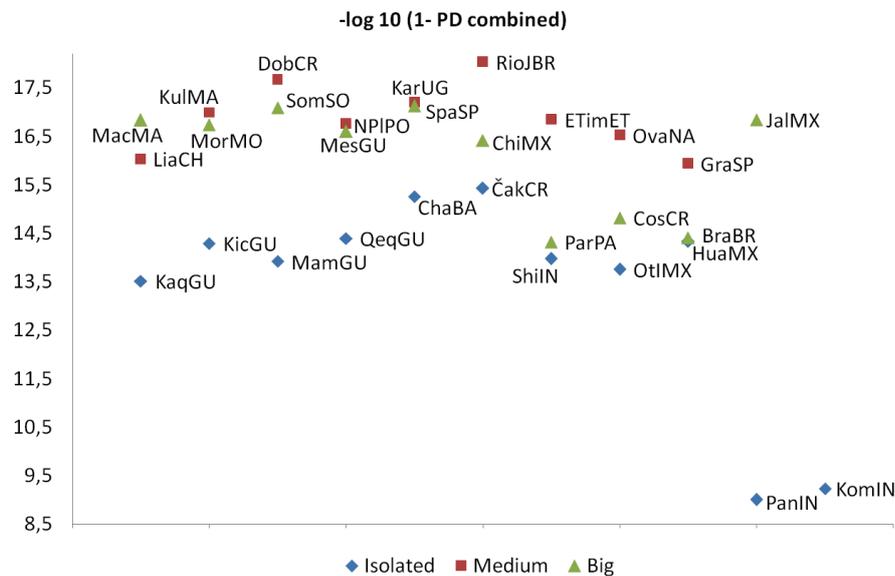


Figure 4: Graphical analysis of Discrimination power (PD).

in medium populations are D5S818 and D18S51 respectively; and TPOX and D18S51 in large size populations. When studying H_o , the gene diversity also differs. In isolated populations the loci with the least discrimination capacity is TH01, while the one with the highest is FGA. However, in both medium and large populations, the lowest values are in the loci TPOX and the highest in D18S51.

Discussion

As can be seen, forensic parameters, such as PD and H_o , and the analysis of the 13 STR loci are useful tools for individual and paternity testing in a population. However, in this study we compared how these values varied depending on the type of population. It is common to find loss of heterozygosity and presence of similar haplotypes at a higher proportion in small isolated populations because of aboriginal identity preservation.

The aim of this work is to evaluate the forensic efficiency of some statistical parameters. Some forensic cases in isolated populations require modifications in analytical procedures. In daily lab work, it is easy to find single-parent paternity cases. In many cases, a positive result is given in the single-parent situation, but when the other progenitor is included the positive result turns into a negative one. This is even more frequent in case in which the individuals come from an isolated and endogamic population. Table 1 shows a case in which a fatherhood relation was established given a positive result. Later, when the mother of the minor was included in the study as a reference sample, the fatherhood relation turns negative. That is why it is very important to know the ethnic origin of the individuals for creation of genetic databases of isolated populations to better calculate paternity indexes and statistics.

Besides the previously mentioned problems; analyzing the results, it can be deduced that there is a decrease in the value of the power of discrimination (PD) by around 10% in isolated populations compared to medium and large populations. This finding is also confirmed when Y-STRs are studied, as can be seen in the Mayan and Guatemalan Mestizo populations [19], where medium PD values are lower in

isolated populations (Mayan) (0.527) than in big populations (0.654). The H_o value was also lower in isolated population than in medium and large size populations, but in with a less pronounced difference than the previous one.

As can be observed, H_o and PD are lower in all isolated populations than in the other populations. Linguistic and geographical isolation as well as endogamous processes have contributed to the genetic isolation of aboriginal populations. The fact that the isolated populations reside in zones with geographical barriers and rarely interact with other people has limited their genetic pool. This can be a problem in forensic casework.

The use of general genetic frequencies databases derives in erroneous results when samples from isolated populations are studied. There is progressively more genetic information of isolated populations that let the creation of specific genetic frequency tables for each population. If the ethnic origin of the sample is well known and a specific genetic frequency table is available for that population, better statistical results will be given in paternity and criminology testing. The availability of different commercial kit that cover a large number of aSTR markers (IdentifilerPlus, NGM SE, Powerplex 21, PowerplexFusion, Investigator ESSplex SE Plus Kit...) lets a better characterization of isolated population. The more genetic information available of these populations, the better results in the identification of the individuals will be obtained. Therefore, frequency tables can be built with those markers that better characterize the population and give more information (better discrimination power, more heterozygosity, better exclusion power...).

Furthermore, it is important to add to the routine analysis lineage markers; mitochondrial DNA, X-STR markers and Y-chromosome STR.

One of the most important activities toward achieving efficient human identification is that of obtaining reference samples. In contrast with other populations, the study of endogamous populations requires an appropriate family references (direct relatives, parents, or children)

and direct references, such as personal effects or antemortem biological specimens like biopsies and saved bloodstain cards, because of the reduced genetic diversity. In some cases, even when there is only one direct relative, the analyst has to be careful in giving a false positive result. That is why it is important to adapt action protocols when working with these kinds of populations.

Conclusions

It is common knowledge that small and isolated populations have a reduced gene pool which is why each case has to be studied on its own. Action protocols have to be adapted in the daily work with these populations and that is why some recommendations have been proposed in the paper:

- Perform genetic population studies using the highest number of autosomal STR available and determine the ones that give the most genetic information possible and better characterize each population.
- Add complementary lineage information with the use of X and Y STR and mitochondrial DNA.
- Use of appropriate family references (direct relatives, parents, or children) and direct references, such as personal effects or antemortem biological specimens like biopsies and saved bloodstain cards.

Acknowledgements

The authors thank the DNA donors from the city of Granada, Spain, as well as Paraguay, Ecuador, Costa Rica, Brazil, Mexico-Chihuahua, and Guatemala. Furthermore, the authors want to thank the staff of the DNA Laboratory of the National Police of Paraguay, the Criminalistics - DNA Laboratory of the National Police of Ecuador, the OIJ - Forensic Science Laboratory of San Joaquín de Flores (Costa Rica), the Fleury Medical Diagnostic Center - Molecular Biology Section of Sao Paulo (Brazil), the DNA Laboratory of P.G.J. del Estado de Chihuahua (Mexico), and the INACIF (Instituto Nacional de Ciencias Forenses de Guatemala) for kindly providing the DNA samples and for their great effort in collecting information about the origin and ethnicity of each individual in the study. Thanks to all of them, population studies can be performed.

References

1. Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* 51: 253-265.
2. Meyer E, Wiegand P, Brinkmann B (1995) Phenotypic differences of STRs in 7 human populations. *Int J Legal Med* 107: 314-322.
3. Bosch E, Calafell F, Pérez-Lezaun A, Clarimón J, Comas D, et al. (2000) Genetic structure of north-west Africa revealed by STR analysis. *European Journal of Human Genetics* 8: 360-366.
4. Zhivotovskiy LA, Ahmed S, Wang W, Bittles AH (2001) The forensic DNA implications of genetic differentiation between endogamous communities. *Forensic Sci Int* 119: 269-272.
5. Sharer RJ (1998) *La civilización maya*. Fondo de Cultura.
6. Coe MD (2005) *The Maya*.
7. Barrot C, Sánchez C, Ortega M, González-Martín A, Brand-Casadevall C, et al. (2005) Characterisation of three Amerindian populations from Hidalgo State (Mexico) by 15 STR-PCR polymorphisms. *Int J Legal Med* 119: 111-115.
8. Gazi NN, Easwarkhanth M, Ferdous A, Qadri SS, Singh L, et al. (2010) Genetic polymorphisms of 15 autosomal STR loci in three isolated tribal populations of Bangladesh. *Forensic Science International: Genetics* 4: 265-266.
9. *Ethnologue: Languages of the World* (2009) SIL International, Dallas, Texas.
10. Novokmet N, Pavčec Z (2007) Genetic polymorphisms of 15 AmpFISTR Identifier loci in Romani population from Northwestern Croatia. *Forensic Sci Int* 168:e43-e46.
11. Wu FC, Ho CW, Pu CE, Hu KY, Liu DH (2009) Genetic polymorphisms of 17 Y-chromosomal short tandem repeat loci in Atayal population of Taiwan. *Croat Med J* 50: 313-320.
12. Collins PJ, Hennessy LK, Leibel CS, Roby RK, Reeder DJ, et al. (2004) Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: The AmpFISTR Identifier PCR amplification kit. *J Forensic Sci* 49: 1265-1277.
13. Finis CJ (2001) Megaplex STR Analysis from a Single Amplification: Validation of the PowerPlex® 16 System. *Profiles in DNA*: 3.
14. GeneScan Analysis Software Version 3.1 © Copyright 1998, The Perkin-Elmer Corporation.
15. Applied Biosystems (2008) GeneMapper® ID-X Software Mixture Analysis Tool v1.1.
16. Nei M, Roychoudhury AK (1974) Sampling Variances of Heterozygosity and Genetic Distance. *Genetics* 76: 379-390.
17. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361-372.
18. Tereba A (2001) Tools for Analysis of Population Statistics. *Profiles in DNA*.
19. Martínez-González LJ, Saiz M, Álvarez-Cubero MJ, Gómez-Martín A, Álvarez JC, et al. (2012) Distribution of Y-chromosomal STRs loci in Mayan and Mestizo populations from Guatemala. *Forensic Science International: Genetics* 6: 136-142.

Citation: Maria Saiz MS, Alvarez-Cubero MJ, Alvarez JC, Martinez-Gonzalez LJ, Lorente JA (2014) Action Protocols in DNA Identification of Isolated Populations. *J Forensic Res* 5: 218 doi:[10.4172/2157-7145.1000218](https://doi.org/10.4172/2157-7145.1000218)

Submit your next manuscript and get advantages of OMICS Group submissions

Unique features:

- User friendly/feasible website-translation of your paper to 50 world's leading languages
- Audio Version of published paper
- Digital articles to share and explore

Special features:

- 300 Open Access Journals
- 25,000 editorial team
- 21 days rapid review process
- Quality and quick editorial, review and publication processing
- Indexing at PubMed (partial), Scopus, EBSCO, Index Copernicus and Google Scholar etc
- Sharing Option: Social Networking Enabled
- Authors, Reviewers and Editors rewarded with online Scientific Credits
- Better discount for your subsequent articles

Submit your manuscript at: <http://www.omicsonline.org/submission>