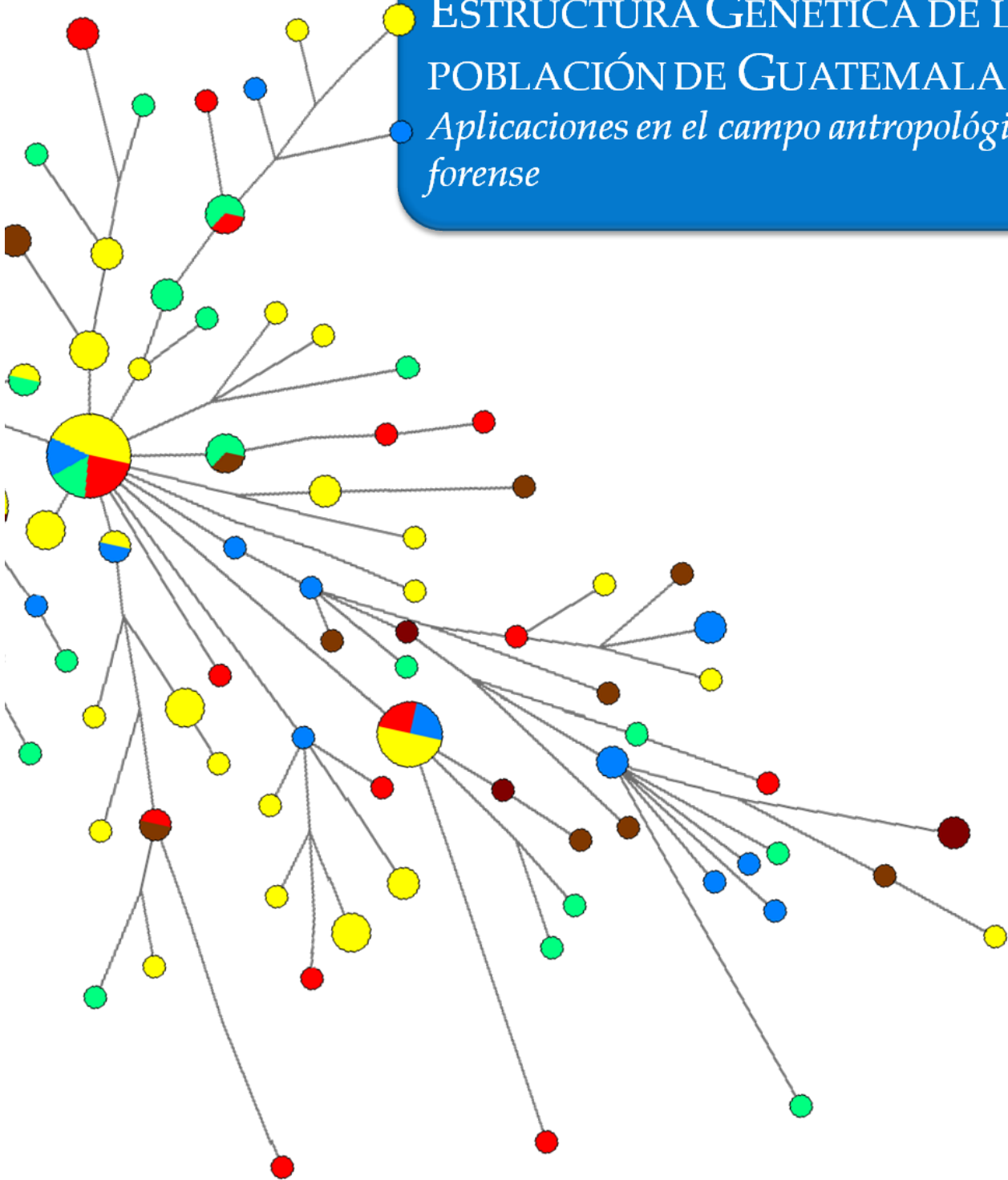


ESTRUCTURA GENÉTICA DE LA POBLACIÓN DE GUATEMALA

Aplicaciones en el campo antropológico y forense



GUATEMALAN GENETICS STRUCTURE – Applications in forensic and anthropological sciences

LUIS JAVIER MARTÍNEZ GONZÁLEZ
TESIS DOCTORAL
GRANADA, 2011

Editor: Editorial de la Universidad de Granada
Autor: Luis Javier Martínez González
D.L.: GR 1570-2012
ISBN: 978-84-9028-031-7



ugr

Universidad
de Granada

Tesis doctoral:

ESTRUCTURA GENÉTICA DE LA POBLACIÓN DE
GUATEMALA

Aplicaciones en el campo antropológico y forense

GUATEMALAN GENETIC STRUCTURE

Applications in forensic and anthropological sciences

Luis Javier Martínez González

Granada 2011



José Antonio Lorente Acosta, Profesor Titular de Universidad, del Departamento de Medicina Legal, Toxicología y Antropología Física de la Universidad de Granada, Director del Laboratorio de Identificación Genética de la Universidad de Granada y Director científico del Centro Pfizer - Universidad de Granada - Junta de Andalucía de Genómica e Investigación Oncológica (GENYO).

Juan Carlos Álvarez Merino, Técnico Superior en Docencia e Investigación del Departamento de Medicina Legal, Toxicología y Antropología Física de la Universidad de Granada.

Cristina Martínez Labarga, Investigadora del "Centro de Antropología Molecular para el Estudio del ADN Antiguo", del Departamento de Biología, de la Universidad de Roma "Tor Vergata".

Certifican que:

Luis Javier Martínez González, Licenciado en Ciencias Biológicas, ha realizado la Tesis Doctoral "ESTRUCTURA GENÉTICA DE LA POBLACIÓN DE GUATEMALA. Aplicaciones en el campo antropológico y forense." en el Laboratorio de Identificación Genética, del Departamento de Medicina Legal, Toxicología y Antropología Física, de la Universidad de Granada y en el Centro de Antropología Molecular para el Estudio del ADN Antiguo, del Departamento de Biología de la Universidad de Roma "Tor Vergata", para aspirar al grado de Doctor en Biología, con mención de "Doctor Europeo", reuniendo las condiciones necesarias para ser presentada y defendida ante el tribunal correspondiente.

Y para que así conste a los efectos oportunos, en cumplimiento de la legislación vigente, firmamos el presente certificado en Granada, a 30 de septiembre de 2011.

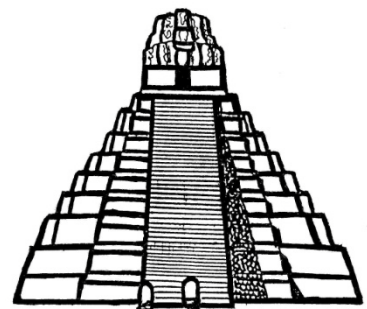
Dr. José Antonio Lorente Acosta

Dra. Cristina Martínez Labarga

Dr. Juan Carlos Álvarez Merino

"La ciencia tiene una gran belleza. Un científico en su laboratorio no es sólo un técnico: es también un niño colocado ante fenómenos naturales que le impresionan como un cuento de hadas".

Marie Curie (Polonia 1867 – Francia 1934)



Índice / *Table of Contents*

OBJETIVOS

CAPÍTULO I: ANTECEDENTES

ABSTRACT.....	3
1. ORGANIZACIÓN DEL GENOMA.....	7
2. CLASIFICACIÓN DE LOS POLIMORFISMOS.....	9
2.1 EL ADN REPETIDO.....	9
2.2 STRs EN CROMOSOMAS AUTOSÓMICOS.....	11
2.3 STRS EN CROMOSOMA Y.....	12
2.4 ORIGEN DE LOS POLIMORFISMOS DE ADN REPETIDO.....	16
2.5 TASAS DE MUTACIÓN EN EL ADN REPETIDO.....	18
3. EL ADN MITOCONDRIAL.....	20
3.1 ESTRUCTURA.....	20
3.2 CARACTERÍSTICAS.....	22
3.3 POLIMORFISMOS EN EL ADNmt.....	27
4. APLICACIÓN DE LOS POLIMORFISMOS.....	29
4.1 APLICACIÓN DE LOS STRs AUTOSÓMICOS.....	29
4.2 POLIMORFISMOS EN CROMOSOMAS SEXUALES.....	30
4.3 APLICACIONES DEL ANÁLISIS DE ADNmt.....	34
5. IDENTIFICACIÓN GENÉTICA EN MEDICINA LEGAL.....	37
6. ESTUDIOS POBLACIONALES Y BASES DE DATOS APLICADOS A LA IDENTIFICACIÓN GENÉTICA.....	38
6.1 PD y PE.....	42
6.1.1 Poder de Discriminación (PD).....	43
6.1.2 Poder de Exclusión a Priori (PE).....	43
7. GENÉTICA FORENSE Y ANTROPOLOGÍA MOLECULAR.....	44
8. ADN Y EL ESTUDIO DE LAS POBLACIONES HUMANAS.....	46
8.1 POLIMORFISMOS.....	47
8.2 MEDIDAS DE DIVERSIDAD.....	50
8.3 HAPLOGRUPOS.....	52
8.4 ANÁLISIS DE FILOGENIAS.....	57
8.4.1 Análisis Multivariados.....	58
8.4.2 Networks.....	60

9. ANTROPOLOGÍA MOLECULAR EN EL NUEVO MUNDO.....	61
9.1 HISTORIA DE AMÉRICA DEFINIDA POR MARCADORES GENÉTICOS.....	62
9.2 MESTIZAJE.....	64
10. LA POBLACIÓN DE GUATEMALA.....	69
10.1 GEOGRAFÍA DE GUATEMALA.....	69
10.2 DISTRIBUCIÓN DE LA POBLACIÓN.....	70
10.3 HISTORIA, CULTURA Y LENGUA.....	72
10.4 PUEBLOS MAYA Y LOS POLIMORFISMOS DEL ADN.....	74
11. BIBLIOGRAFÍA.....	76

CAPÍTULO II: GENOTIPADO DE STRs EN CROMOSOMAS AUTOSÓMICOS

ABSTRACT.....	95
1. INTRODUCCIÓN.....	97
2. MATERIAL Y MÉTODOS.....	101
2.1 LA MUESTRA.....	101
2.2 ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA.....	101
2.2.1 Extracción.....	102
2.2.2 Amplificación.....	102
2.2.3 Electroforesis Capilar.....	103
2.3 ANÁLISIS ESTADÍSTICO DE LA MUESTRA.....	106
3. RESULTADOS.....	111
3.1 ESTUDIO POBLACIONAL.....	111
3.2 DISTANCIAS GENÉTICAS.....	118
3.3 ESTUDIO DEL MESTIZAJE.....	123
4. DISCUSIÓN.....	127
4.1 APORTACIÓN A LA CIENCIA FORENSE.....	127
4.2 APORTACIÓN A LA ANTROPOLOGÍA.....	129
4.3 ESTUDIOS DE MESTIZAJE.....	133
CONCLUSIONS.....	134
5. BIBLIOGRAFÍA.....	135

CAPÍTULO III: GENOTIPADO DE STRs EN EL CROMOSOMA Y

ABSTRACT.....	141
1. INTRODUCCIÓN.....	143
2. MATERIAL Y MÉTODOS.....	146
2.1 LA MUESTRA.....	146
2.2 ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA.....	146
2.2.1 Extracción.....	146

2.2.2 Amplificación.....	146
2.2.3 Electroforesis Capilar.....	148
2.3 ANÁLISIS ESTADÍSTICO DE LA MUESTRA.....	149
3. RESULTADOS Y DISCUSIÓN.....	154
3.1 ESTUDIO HAPLOTÍPICO.....	154
3.2 CALCULO DE HAPLOGRUPOS.....	154
3.2.1 Estudio del Haplogrupo Q en Guatemala.....	157
3.3 DISTANCIAS GENÉTICAS.....	161
3.4 ESTUDIO DE FRECUENCIAS ALÉLICAS.....	163
3.5 DISTANCIAS CON OTRAS POBLACIONES.....	167
3.6 ESTUDIO DEL MESTIZAJE.....	171
3.7 ESTRUCTURA DE LA POBLACIÓN.....	172
4. RESUMEN.....	176
CONCLUSIONS.....	178
5. BIBLIOGRAFÍA.....	180

CAPÍTULO IV: POLIMORFISMOS EN LA SECUENCIA DEL ADN MITOCONDRIAL

ABSTRACT.....	185
1. INTRODUCCIÓN.....	187
1.1 EL ADNmt.....	187
1.2 HAPLOGRUPOS EN EL ADNmt.....	188
1.3 MOVIMIENTOS MIGRATORIOS Y EL ADNmt.....	189
1.4 NUEVAS TÉCNICAS.....	190
2. MATERIAL Y MÉTODOS.....	191
2.1 LA MUESTRA.....	191
2.2 ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA.....	191
2.2.1 Extracción.....	191
2.2.2 Amplificación.....	191
2.2.3 Purificación del Producto de PCR.....	192
2.2.4 Secuenciación.....	193
2.2.5 Purificación del Producto de Secuenciación.....	194
2.2.6 Electroforesis Capilar.....	194
2.3 ANÁLISIS ESTADÍSTICO DE LA MUESTRA.....	195
3. RESULTADOS Y DISCUSIÓN.....	198
3.1 ESTUDIO HAPLOTÍPICO.....	198
3.2 CALCULO DE HAPLOGRUPOS.....	207
3.2.1 Estudio de Redes en la Población de Guatemala.....	211
3.3 DISTANCIAS GENÉTICAS.....	220
3.4 DISTANCIAS CON OTRAS POBLACIONES.....	221

3.5 ESTUDIO DEL MESTIZAJE.....	227
3.6 ESTRUCTURA DE LA POBLACIÓN.....	229
CONCLUSIONS.....	232
4. BIBLIOGRAFÍA.....	233

CAPÍTULO V: PROBLEMAS DE IDENTIFICACIÓN EN POBLACIONES PEQUEÑAS Y AISLADAS

ABSTRACT.....	243
1. INTRODUCCIÓN.....	244
1.1 PROBLEMAS DE IDENTIFICACIÓN.....	245
1.2 PODER DE DISCRIMINACIÓN, EXCLUSIÓN Y HOMOCIGOSIDAD.....	247
1.3 PROBLEMAS EN BASES DE DATOS DE IDENTIFICACIÓN CIVIL.....	248
1.3.1 Personas Desaparecidas (Programa FÉNIX).....	249
1.3.2 Tráfico de Seres Humanos (Iniciativa DNA-PROKIDS).....	250
2. MATERIALES Y MÉTODOS.....	251
2.1 POBLACIONES Y MUESTRAS.....	251
2.2 ANÁLISIS GENÉTICO.....	254
2.3 ANÁLISIS ESTADÍSTICO.....	255
3. RESULTADOS Y DISCUSIÓN.....	256
CONCLUSIONS.....	261
4. BIBLIOGRAFÍA.....	262

CAPÍTULO VI: VARIABILIDAD GENÉTICA DE LA POBLACIÓN DE GUATEMALA

ABSTRACT.....	267
1. INTRODUCCIÓN.....	269
2. DIFERENCIAS ENTRE LOS MARCADORES ESTUDIADOS.....	270
2.1 DISTRIBUCIÓN DE LA VARIABILIDAD EN LA POBLACIÓN MESTIZA DE GUATEMALA Y SU IMPORTANCIA EN LOS ESTUDIOS DE ASOCIACIÓN.....	273
3. VARIABILIDAD EN POBLACIONES NATIVAS AMERICANAS.....	276
4. BIBLIOGRAFÍA.....	277

ABREVIATURAS.....	279
--------------------------	------------

ANEXOS.....	283
--------------------	------------

Objetivos / Aims of the thesis

Los objetivos de la presente tesis doctoral son:

1. Caracterizar la estructura genética de la población de Guatemala, mediante el genotipado de los marcadores genéticos: D3S1358, TH01, D21S11, D18S51, D2S1338, D5S818, D13S317, D7S820, D16S539, CSF1PO, D19S433, vWA, D8S1179, TPO X y FGA, presentes en cromosomas autosómicos; DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385, DYS393, DYS391, DYS439, DYS635, DYS392, YGATAH4, DYS437, DYS438, y DYS448, localizados en el cromosoma Y y la caracterización de las líneas mitocondriales a través de la secuenciación de dos regiones de alta variación HV1 y HV2 del D-loop del ADN mitocondrial
2. Calcular el grado de variabilidad de cada marcador, mediante sus frecuencias en la población y calcular su poder de discriminación en la población. Discutir la viabilidad de estos marcadores para su uso en identificación genética en la población de Guatemala. Así como estudiar sus posibles aplicaciones en el campo forense.
3. Estudiar el origen de la población de Guatemala, en particular la componente genética que ha formado el pool genético de los mestizos a través del análisis del mestizaje.
4. Estudiar las relaciones genéticas de las poblaciones de Guatemala y su relación genética con otras poblaciones nativas americanas y describir los procesos genéticos que han dado lugar a su estructura actual.

The objectives of this thesis are:

1. Characterize the genetic structure of the Guatemalan population through the genotyping of genetic markers: D3S1358, TH01, D21S11, D18S51, D2S1338, D5S818, D13S317, D7S820, D16S539, CSF1PO, D19S433, vWA, D8S1179, TPO X y FGA, in autosomal chromosomes; DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385, DYS393, DYS391, DYS439, DYS635, DYS392, YGATAH4, DYS437, DYS438, y DYS448, located in Y chromosome and the characterization of mitochondrial profiles through the sequencing of two hypervariable regions, HV1 and HV2 of the mitochondrial DNA D-loop.
2. Calculate the variability degree of each marker through their frequencies and estimate their discrimination power in the populations. Discuss the viability of these markers in genetics identification in the Guatemalan population as the different applications in forensic field.
3. Study the origin of the Guatemalan population, especially the genetic component that has created the genetic pool of Mestizos with crossbreeding studies.
4. Determine the different genetic relationships between the Guatemalan populations as with other Native American populations and describe the genetic processes that have developed the current structure.

Capítulo I: Antecedentes

Chapter I: Antecedents



ABSTRACT

Forensic Genetic and Molecular Anthropology are state-of-the-art fields of knowledge. They share the same resources to facilitate for their own development, which includes new techniques of Molecular Biology.

Using these new techniques it has been possible to study different regions within the DNA. Some of these regions can be very interesting for both Forensic Genetics and Molecular Anthropology, since they are distinctive markers to distinguish between individuals, polymorphic regions and their inheritance.

This chapter talks about the human genome, and the most interesting regions to make a profoundly genetic characterization of a population, from the Forensic and Anthropologic points of view. DNA polymorphism and single mutations will be looked at in detail and different characteristics will be described depending on where they are located in the genome: autosomic or sexual chromosome, and mitochondrial DNA. All their characteristics, inheritance and applications are described.

Initially, and in order to carry out these studies in the genome, we should know how to interpret the results obtained by molecular analysis. Secondly, we should also know perfectly about the sample we are processing.

In summary, in this chapter we will briefly review the Statistic Analysis for this genetic data to obtain all the information for these markers. So, firstly we will

analyze all the information and obtain data like Hardy-Weinberg, heterocigosity and linking. We will study their variability to know if they are useful for identification purposes (power of discrimination and exclusion...). Finally, we will evaluate the differences among the data obtained in the different population.

Our study will be focused on a population from Guatemala; due to many previous articles written about this theme, we will know what to expect from this within this population. We also need to know the origin and contributions of other populations to our population studied.

Finally, we will describe the historical and geographical environment where our population is located, so we could interpret all the results obtained in the markers.

La genética forense como la antropología molecular son dos ciencias relativamente jóvenes. Ambas coinciden en las herramientas que han facilitado su desarrollo, las nuevas técnicas de biología molecular.

Desde que Wyman y White en 1980 descubrieron el primer locus polimórfico [1], al ver la gran variabilidad que presentaban dichos loci entre distintos individuos, fueron distintas ciencias las que rápidamente aplicarían este nuevo conocimiento con distintos fines. El estudio de estos loci se vio acentuado sobre todo a partir de 1989 que sería cuando Mullis, encontraría la técnica para poder estudiar cualquier loci polimórfico que se encontrara en el genoma humano [2,3].

La variabilidad que presentaban estos loci hizo que rápidamente se aplicaran a la identificación humana, lo que supuso una gran revolución para la biología forense. Y por tanto, se convirtieron rápidamente en un arma muy importante dentro de la Medicina Legal, aplicándose por primera vez en un caso judicial en 1985.

Estos loci además presentaban otra característica muy importante, su heredabilidad. Esto no sólo fue importante para la identificación humana, sino que otras ciencias rápidamente supieron aprovechar esta nueva herramienta. Simultáneamente la genética de poblaciones supo encontrar en estos marcadores un instrumento muy útil para poder explicar fenómenos de variabilidad genética. La ausencia de selección natural, en estas zonas no codificantes del ADN, podía resolver y sustentar muchas teorías relacionadas con la herencia. El estudio de estos fenómenos hizo que la antropología sufriera un gran cambio. Poder estudiar las distintas poblaciones mediante caracteres que no sufrían selección y cuya herencia era bien conocida, permitió comenzar a sustentar teorías evolutivas, migratorias y de mestizaje.

El estudio de estos mismos marcadores desde distintos puntos de vista ha hecho enriquecerse a los profesionales que trabajan en estas ciencias, pudiendo colaborar en sus estudios o emergiendo y diferenciándose en los distintos campos que el apasionante mundo de la genética humana ofrece.

1. ORGANIZACIÓN DEL GENOMA

Las células contienen el ADN con la información necesaria para realizar sus tareas establecidas. En su núcleo, confinan casi la totalidad de la información para realizar las tareas que programarán y permitirán el desarrollo de la célula. En realidad, esto no es del todo cierto; no todas las células poseen núcleo, hay algunas que lo pierden cuando maduran o envejecen e incluso otras que lo tienen solamente durante pequeñas fases vitales. Por otro lado, no todo el ADN se encuentra en los núcleos de las células sino que también en orgánulos como las mitocondrias. Según su localización y estructura se diferencian dos tipos de ADN, el ADN nuclear y el ADN mitocondrial. Ambos tienen interés tanto desde un punto de vista forense como para la antropología molecular; y poseen características que los complementan tanto en la identificación e investigación de relaciones familiares, como en los estudios de antropología.

El porcentaje de **ADN Codificante** se encuentra en torno al 23,5%, aunque el directamente encargado de codificar proteínas no alcanza el 2%. Este tipo de ADN se llama ADN codificante o expresivo. Pero no todo el ADN que conforma el genoma es codificante.

Una de las características del ADN codificante es su escasa variabilidad entre individuos, con la excepción de ciertas regiones (el sistema HLA).

Más del 75% del ADN humano corresponde a **ADN no codificante**, sin expresión génica. Aunque en un principio se pensó que este tipo de ADN no tenía ningún tipo de función, o simplemente función estructural, se ha comprobado la necesidad de él a nivel promotor y regulador. Un 60% de este tipo de ADN corresponde a secuencias de copia única y un 40% a secuencias repetitivas de ADN.

Clasificación de ADN no codificante

El **ADN de copia única** está compuesto por secuencias que se encuentran representadas una o muy pocas veces en el genoma. Dentro de las secuencias de copia única se encuentran los “pseudogenes”, secuencias que fueron genes en estadios evolutivos anteriores, pero que han perdido su funcionalidad. También se encuentra ADN espaciador entre regiones codificantes del genoma, como los intrones, que no contienen información útil para la síntesis de la cadena polipeptídica y son eliminados durante el proceso de “splicing”.

Por otro lado, hay dos tipos diferentes de **ADN de copia múltiple**, (secuencias repetitivas): repeticiones que se van intercalando en el genoma y otras que se repiten de forma concatenada, en tándem, localizadas en centrómeros y en telómeros.

- ADN Repetitivo Disperso (15-20% del genoma): regiones donde las unidades de repetición no se agrupan, aparecen dispersas a lo largo del genoma. Está representado por 2 familias: SINEs y LINEs.
- ADN en tándem (10% del genoma): regiones donde un pequeño número de nucleótidos (unidades de repetición) aparecen repetidas en bloque un número concreto de veces. Carecen de una función conocida y son capaces de propagarse y extenderse a través del genoma. En función del número de nucleótidos que contengan estas unidades se puede diferenciar ADN satélite, minisatélites y microsatélites.

2. CLASIFICACIÓN DE LOS POLIMORFISMOS

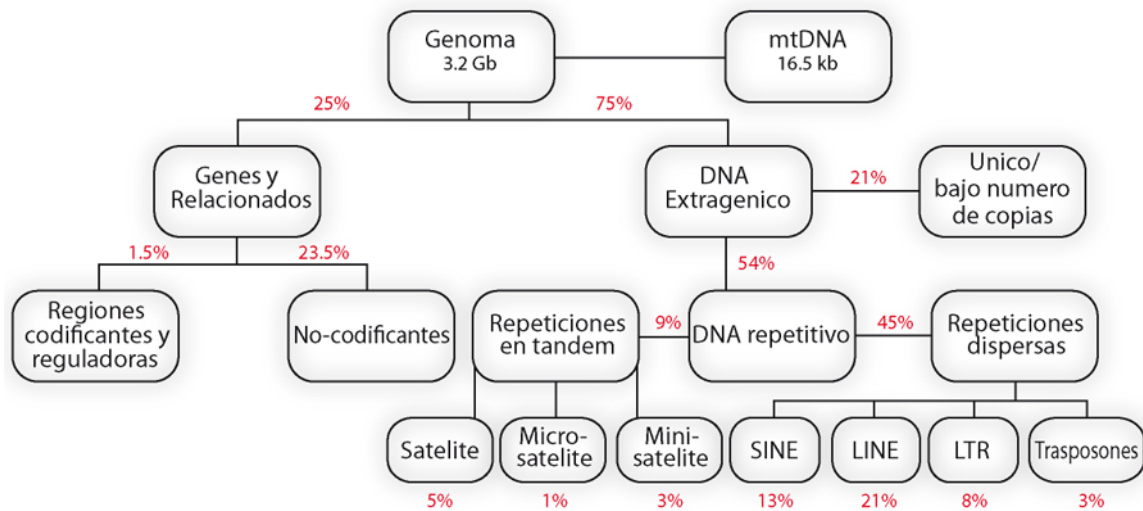


Figura I 1. Esquema de la organización del genoma humano.

2.1 EL ADN REPETIDO

Las secuencias repetidas en tándem suponen el 5 - 10 % del genoma. Se caracterizan por la presencia de una secuencia común repetida en tándem de manera continua en un fragmento de ADN [4]. Son características de los centrómeros y de los telómeros y parece que son importantes para determinar la estabilidad y la posición correcta de los cromosomas. Se encuentran distribuidas en el genoma de dos maneras:

-Tipo I: En grandes bloques caracterizados por una disposición repetitiva en tándem de diversas unidades.

Se distinguen tres tipos de secuencias repetidas dependiendo de su densidad (gradiente de $\text{Ag-Cs}_2\text{SO}_4$) y longitud (en pb):

El "ADN satélite" o **familias satélite** hace alusión a grandes bloques de ADN altamente repetitivo organizadas en tándem. La longitud de las secuencias de este grupo varía y puede alcanzar hasta miles de pares de bases. Hay cuatro grandes clases de satélites: I, II, III y IV[4]. Constituyen un 2-6% del genoma. La

localización de estas secuencias es básicamente la heterocromatina estructural de las regiones centroméricas.

El **satélite alfoide**, sería el segundo tipo, con una densidad similar a la de los satélites II y III, sin embargo su secuencia consenso es de 340 pb. El origen de esta secuencia estaría en la formación de un dímero de 340 pb a partir de dos unidades de 171pb.

Cuando las secuencias poseen una longitud total comprendida entre 250 y 500 Kb, están repetidas en tándem, son de copia única y altamente polimórficos se denominan **midisatélites**.

-Tipo II: Secuencias repetidas en pequeños bloques, con la disposición en tándem formada por la misma unidad básica y dispersos por todo en genoma.

Cuando los tándem están constituidos por una secuencia consenso de 9 a 100 pb se denominan **minisatélites** [5]. Estos suelen encontrarse en regiones teloméricas, aunque se han encontrado intercalados en otras posiciones. Cuando se encuentra en los extremos de los telómeros se denomina "ADN telomérico" y consta de un bloque de 10 a 15 Kb de repeticiones en tándem de una secuencia de 6 nucleótidos. Los hexanucleótidos son añadidos a los extremos de los telómeros por una enzima llamada telomerasa y parece que actúan como amortiguadores de la degradación de los extremos de los cromosomas.

Los minisatélites son altamente polimórficos entre los distintos individuos, presentando índices de heterocigosidad superiores al 50% llegando a veces al 100% [5]. Estas secuencias muy utilizadas en los comienzos de la genética forense se les denominaban abreviadamente VNTR (Variable Number of Tandem Repeat).

Los **microsatélites** contienen una secuencia consenso de 1 a 6 pb repetida en tándem. Son también secuencias altamente polimórficas y en este caso a los polimorfismos de estas secuencias se les llama STR (Short Tandem Repeat). Cuando las repeticiones son de un solo nucleótido suelen ser las bases A y T

(0.3% del genoma). En repeticiones de dos bases, dinucleótidos, son comunes las de CA (0.5%), seguidas de CT/AG (0.2%) y suelen ser muy polimórficas. Las repeticiones de trinucleótidos o tetranucleótidos se encuentran en menos ocasiones pero poseen un alto grado de polimorfismo, por lo que son en este momento la herramienta consenso de los laboratorios de identificación genética.

2.2 STRs EN CROMOSOMAS AUTOSÓMICOS

Si se hace un screening por hibridación para detectar la presencia de STRs en el genoma, éste permitiría identificar un locus STR en la mayoría de los segmentos genómicos, cada 6-10 Kb [6]. Restringiendo la búsqueda a los STRs triméricos y tetraméricos, se ve que se distribuyen cada 15 Kb, existiendo aproximadamente 200.000 loci de estas características en el genoma humano [7,8].

Los STRs son numerosos por todo el genoma 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. Estos loci suelen estar asociados con el extremo 3' de las inserciones Alu, resultando, muy probablemente, de la degeneración de la secuencia de la cola poli-A [8,9].

El alto grado de polimorfismo de los loci STRs se debe a la variación en el número de unidades de repetición [10,11]. 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, junto con mayor o menor variación de las secuencias.

Los loci STR se pueden diferenciar en dos grupos principales. Un grupo sería los que poseen un pequeño número de alelos bien diferenciados (< 12) y otro los que poseen un número elevado (> 35) aunque su diferenciación entre alelos es compleja.

Los STRs del segundo grupo son mucho más polimórficos, y más complejos; poseen alelos que difieren en un número incompleto de repeticiones, presentando mayor heterogeneidad, pero, estas diferencias entre alelos, de una sola base o dos, requieren un análisis muy preciso. Dentro del primer grupo lo más usual es que la diferencia entre dos alelos sea de una unidad de repetición, aunque puede haber algunas excepciones, por ejemplo en el locus TH01 existen delecciones en la región de repetición de los alelos 8, 9 y 10, apareciendo las variantes 8.3, 9.3 y 10.3 (son 7, 8 o 9 repeticiones del tándem AATG más un trinucleótido ATG). Realmente tras observar las frecuencias del alelo 9.3, se puede ver que las otras dos variantes alélicas se han generado a partir de este alelo, por delección o inserción de una unidad de repetición. El alelo 9.3 se puede observar con frecuencia en todas los estudios poblacionales excepto en los aborígenes australianos. También se puede ver en estos estudios que los alelos 8.3 y 10.3 únicamente aparecían en la población alemana [12]. Estudios actuales de otras poblaciones europeas ponen de manifiesto la propagación a otras regiones de estas variables alélicas [13].

2.3 STRS EN CROMOSOMA Y

Antes de describir estos sistemas en el cromosoma Y, y sus características, hay que conocer generalidades de este cromosoma así como su herencia.

El **cromosoma Y** es específico del sexo masculino, haploide y su herencia es de padre a hijo.

Al ser haploide, no tiene homólogo, con lo cual su recombinación está en su mayoría impedida (en un 90%). Por tanto, se tiene que hablar de zonas recombinantes y no recombinantes en los cromosomas sexuales, o lo que es lo mismo zonas pseudoautosómicas. A pesar de su distinta naturaleza morfológica, los cromosomas X e Y pueden aparearse durante la meiosis e intercambiar información. Este fenómeno ocurre en pequeñas regiones que contienen secuencias homólogas presentes en ambos cromosomas, denominadas regiones pseudoautosómicas porque en ellas las secuencias de

ADN no muestran una herencia ligada al sexo en sentido estricto. En los cromosomas sexuales humanos hay dos de estas regiones, denominadas PAR1 y PAR2, situadas respectivamente en los extremos terminales del brazo corto y largo de dichos cromosomas.

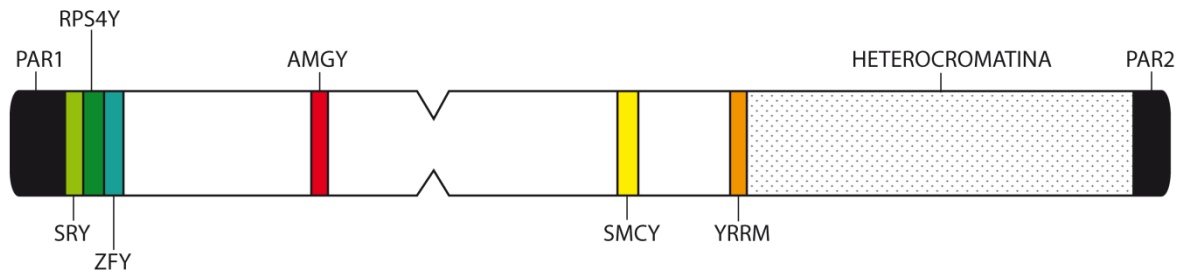


Figura I.2. Estructura y distribución del genoma en el cromosoma Y

La región pseudoautosómica mayor (PAR1) del cromosoma Y tiene un tamaño aproximado de 2.6 Mb. La recombinación a nivel de PAR1 es necesaria para una segregación normal de los cromosomas X e Y en la meiosis [14,15] y, por tanto, recombina siempre. Pueden existir alteraciones, como en el síndrome de Klinefelter, cuyo cariotipo 47 XXY es producto de la no-disyunción, no produciéndose recombinación a nivel del PAR1 en la gran mayoría de casos [16].

La región pseudoautosómica menor (PAR2) del cromosoma Y mide sobre 320 Kb y no siempre participa en procesos de recombinación. Además, la actividad de recombinación de PAR2 no puede sustituir a la de PAR1 [17], no siendo, por tanto, ni necesaria ni suficiente para el éxito de la meiosis masculina.

La **región específica del cromosoma Y** no recombina con ninguno de los cromosomas, por lo cual sus loci se transmiten inalteradamente por línea paterna de generación en generación (en ausencia de mutaciones). Presenta algunos genes funcionales, de gran importancia para el desarrollo sexual del hombre. Entre ellos el SMCY que codifica el antígeno menor de histocompatibilidad H-Y [18]. El gen ZFY, codifica una proteína que presenta un motivo en dedo de zinc [19] cuya función no es conocida por el momento. El gen SRY ha sido aislado en el brazo corto del cromosoma Y, a 5 Kb en dirección proximal a PAR1 [20], presenta un solo exón y carece de homólogo en X.

La familia de genes YRRM se cree que podría intervenir en la regulación de la espermatogénesis. Como el gen anterior, este grupo carece de homólogos.

El RPS4Y, localizado entre SRY y ZFY, es un gen que codifica la proteína ribosómica S4. Posee un homólogo en X que escapa a la inactivación [21]. El AMGY es un gen que codifica la amelogenina (amelogenine-like, Yp11.1), también tiene un homólogo en X (AMGX) [22], que es responsable de la formación normal del esmalte dentario.

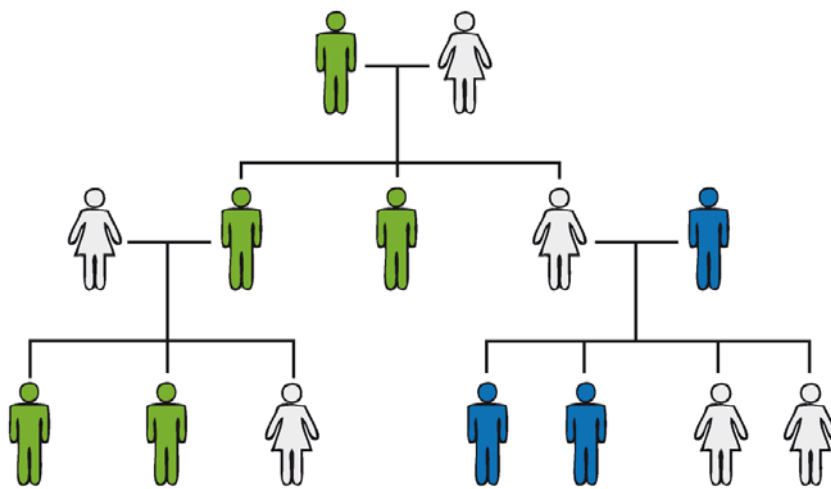


Figura I.3. Herencia del cromosoma Y. Por su haploidía tiene limitada la recombinación y su herencia es siempre de padre a hijo varón. En nuestra cultura, lo más usual es que la herencia del cromosoma Y está asociada a la herencia del apellido.

Existen distintos **marcadores genéticos en cromosoma Y**. Se diferencian loci bialélicos (SNPs y Alu) y multialélicos. Los resultados de los primeros permiten agrupar a los individuos en haplogrupos y los multialélicos los individualizan en haplotipo [23]. Los marcadores bialélicos poseen una menor tasa de mutación (1×10^{-8}) y son más numerosos. Los loci multialélicos son minisatélites (sólo se han descrito 2) y marcadores STRs (200) [24,25].

Los primeros **polimorfismos del cromosoma Y** se descubrieron en 1985 [26,27]. Los estudios iniciales se centraban en el análisis de la región denominada Y27H39 (conocida actualmente como DYS19), y demostraron su utilidad tanto en estudios antropológicos [28], como en análisis de paternidad en ausencia de progenitor masculino [29].

Si bien la aplicabilidad resultaba limitada por no disponer de marcadores adicionales, tal restricción pudo ser posteriormente superada. Los marcadores **Y-STRs** fueron aumentando drásticamente tras el PGH. De hecho en identificación genética el haplotipo mínimo recomendable era de seis STRs, aumentando en 2003 a los diez que se utilizan actualmente [23,24].

Dentro de los STRs de cromosoma Y hay que diferenciar marcadores de una sola copia y marcadores de varias copias (Y-single copy & Y-multy-copy markers). Existen dos regiones palindrómicas, duplicadas, razón por la que cuando se intenta amplificar algunos Y-STRs, el producto obtenido sea doble del esperado.

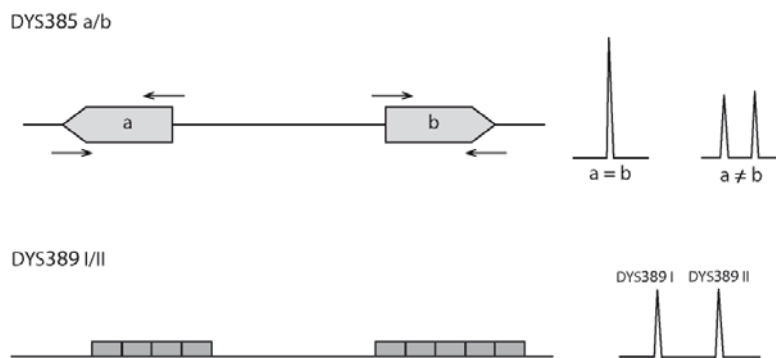


Figura I.4. En este estudio se incluyen dos STRs duplicados, aunque el origen es el mismo, existen varias copias de la misma zona debido a una inserción. La estrategia de amplificación en cada uno de ellos es distinta, debido a que en el caso del DYS389 la duplicación se inserta en el mismo sentido y en el caso del DYS385 la inserción se realizó en sentido contrario.

Un ejemplo de esto es lo que ocurre con el marcador DYS385 a/b, aunque una repetición de la otra estén separadas más de 40.775pb, y estén en sentido contrario, las zonas que flaquean el STR son iguales para ambos por eso el producto obtenido es el de dos zonas, pudiendo ser el mismo o distinto (ver figura I.4). Algo similar ocurre con el marcador DYS389I y II, es una duplicación, separada una de la otra 120pb, y en el mismo sentido. En este caso al realizar la amplificación el cebador (en adelante *primer*) reverso se une a la misma posición en ambos casos, y el *primer* forward se une a las zonas que flanquean cada una de las regiones de repetición. Por lo que cuando se habla de repeticiones en el marcador DYS389II realmente son la suma de las repeticiones de la primera y la segunda zona [24,30,31].

Dependiendo de la población, las frecuencias de los Y-STRs poseen una distribución variable [32]. Los distintos grupos poblacionales se estudiaron para

validar muchos de los marcadores en el campo forense [33-36][6]. Estas variaciones también se han aplicado en genética de poblaciones y antropología molecular [37,38]. La alta tasa de mutación de estos proporciona una información muy adecuada para diferenciar poblaciones cercanas desde el punto de vista filogenético. También se están descubriendo como una herramienta para poder asociar enfermedades hereditarias con incidencia en grupos familiares y/o poblacionales [39], sobretodo en aquellas con mayor incidencia en varones como puede ser el cáncer de próstata [40,41].

2.4 ORIGEN DE LOS POLIMORFISMOS DE ADN REPETIDO

Según el tamaño de la unidad de repetición, los polimorfismos de las secuencias repetitivas pueden deberse a diferentes mecanismos genéticos. En loci microsatélites las tasas de mutación en la línea germinal presentan variaciones considerables que pueden alcanzar niveles de hasta 8×10^{-3} . En las repeticiones de tetranucleótidos se sabe que se forman alelos de nueva longitud sin que haya intercambio de los marcadores flanqueantes, lo cual implica que estos no se originan por entrecruzamiento desigual. Parece que el mecanismo más probable para explicar la variación del tamaño es una forma de intercambio de información de secuencia que comienza con un apareamiento incorrecto por deslizamiento [42]. Este fenómeno ocurre cuando en el apareamiento de las dos hebras complementarias las secuencias repetidas de ambas hebras se reconocen pero no se aparean. Esto provoca un “deslizamiento” de una hebra sobre la otra que da lugar a un apareamiento incorrecto. Aunque este mecanismo puede ocurrir tanto en ADN replicante como no replicante, parece más probable que ocurra en el replicante denominándose deslizamiento replicativo.

La replicación con deslizamiento es responsable de deleciones y duplicaciones debido al apareamiento entre repeticiones no contiguas y se ha sugerido que constituye un mecanismo importante de evolución del genoma y de las secuencias de ADN. En loci satélites y minisatélites los polimorfismos de longitud pueden originarse como consecuencia de un intercambio desigual entre cromátidas hermanas. El mecanismo consiste en una forma de

recombinación en la que el entrecruzamiento afecta a secuencias no alélicas situadas en cromátidas hermanas. Tiene lugar preferentemente en posiciones en las que los bloques de secuencias repetidas en tándem son de tamaño moderado o grande. Si los cromosomas se rompen y se reúnen mientras las cromátidas están unidas por apareamientos incorrectos aparecerá una inserción o una deleción de una unidad de repetición completa. Una de las cromátidas quedará con una inserción y la otra con una deleción debido a que el intercambio es recíproco. Otro tipo de mecanismos genéticos que parecen implicados en la generación de polimorfismos de secuencias de ADN repetidas en tándem son las reacciones de conversión génica. La conversión génica consiste en una transferencia no recíproca de información entre un par de secuencias no alélicas de ADN (conversión génica interlocus) o entre secuencias alélicas (conversión génica interalélica). En este proceso una de las secuencias, denominada secuencia dadora, permanece inalterada mientras que la otra (secuencia aceptora) gana alguna secuencia procedente de la dadora. En una primera etapa se produciría la formación de un heterodúplex entre una hebra de ADN del gen dador y la hebra complementaria del gen aceptor.

Seguidamente, se produciría la conversión de un fragmento del gen aceptor mediante el mecanismo de reparación de apareamientos incorrectos, mecanismo mediante el que enzimas reparadoras que reconocen una zona mal apareada en un heterodúplex la reparan base a base [42].

Las mutaciones de la línea germinal pueden estudiarse detectando y caracterizando alelos minisatélite mutantes en gametos aislados. Para ello se hace una PCR de múltiples diluciones de alícuotas de ADN (cada una con unas 100 moléculas) extraído a partir de esperma humano. Los productos de PCR obtenidos a partir de cada alícuota permiten identificar cualquier nueva mutación de un tamaño lo suficientemente distinto del alelo progenitor como para poder definirla como un alelo nuevo [43].

Parece que la mayoría de las mutaciones que ocurren en los loci implican ganancia de unos pocos elementos en un extremo del bloque de secuencias

repetidas. Existe una desviación hacia la ganancia neta de elementos repetidos y las evidencias apuntan a la existencia de intercambios de secuencia no recíprocos entre alelos, lo cual sugiere una conversión génica interalélica.

2.5 TASAS DE MUTACIÓN EN EL ADN REPETIDO

Cuando se habla de tasas de mutación en STRs, ya sean en cromosomas autosómicos como en cromosomas sexuales, en general, se refiere a mutaciones, por inserción o deleción que aparecen en una nueva generación. Serán inserciones del tándem completo o de alguna base, pero por los métodos de detección, no se tienen en cuenta las mutaciones puntuales, por cambio de bases o polimorfismos.

Para calcular la tasa de mutación, se toman muestras de individuos con parentesco cercano, normalmente padre y madre, y se estudian las diferencias encontradas entre ambos individuos. Las tasas de mutación de cada STR se estiman mediante el número de mutaciones que aparecen entre el número de transmisiones alélicas (número de meiosis estudiadas). Las tasas de mutación en STRs son muy variables, desde valores bajos como 1×10^{-4} , en los loci TH01 o TPOX, a algunos más altos como 2×10^{-3} en el locus D18S51. Normalmente para actualizar estas tasas se hacen estudios interlaboratorios [44-48].

Las tasas de mutación en los marcadores más utilizados, están disponibles en distintas webs, como <http://www.cstl.nist.gov/biotech/strbase/mutation.htm> o <http://www.yhrd.org/>, siendo la fuente de la mayoría de las bases de datos la Asociación Americana de Bancos de Sangre (www.aabb.org) para marcadores autosómicos y de estudios interlaboratorios en caso de cromosomas sexuales.

Tabla I.1. Tasa de mutación de los STRs utilizados en el Kit Identifiler.

STRs	Meiosis Materna (%)	Meiosis Paterna (%)	Numero de cada uno de ellos	Numero Total de Mutaciones	Tasa de Mutaciones
D8S1179	96/409,869 (0.02)	779/489,968 (0.16)	364	1,239/899,837	0.14%
D21S11	464/435,388 (0.11)	772/526,708 (0.15)	580	1,816/962,096	0.19%
D7S820	59/440,562 (0.013)	745/644,743 (0.12)	285	1,089/1,085,305	0.10%
CSF1PO	95/304,307 (0.03)	982/643,118 (0.15)	410	1,487/947,425	0.16%
D3S1358	60/405,452 (0.015)	713/558,836 (0.13)	379	1,152/964,288	0.12%
TH01	31/327,172 (0.009)	41/452,382 (0.009)	28	100/779,554	0.01%
D13S317	192/482,136 (0.04)	881/621,146 (0.14)	485	1,558/1,103,282	0.14%
D16S539	129/467,774 (0.03)	540/494,465 (0.11)	372	1,041/962,239	0.11%
D2S1338	15/72,830 (0.021)	157/152,310 (0.10)	90	262/225,140	0.12%
D19S433	38/70,001 (0.05)	78/103,489 (0.075)	71	187/173,490	0.11%
VWA	184/564,398 (0.03)	1,482/873,547 (0.17)	814	2,480/1,437,945	0.17%
TPOX	18/400,061 (0.004)	54/457,420 (0.012)	28	100/857,481	0.01%
D18S51	186/296,244 (0.06)	1,094/494,098 (0.22)	466	1,746/790,342	0.22%
D5S818	111/451,736 (0.025)	763/655,603 (0.12)	385	1,259/1,107,339	0.11%
FGA	205/408,230 (0.05)	2,210/692,776 (0.32)	710	3,125/1,101,006	0.28%

3. EL ADN MITOCONDRIAL

Este ADN se encuentra en las mitocondrias, orgánulos ovalados albergados en las células, en número variable. La mitocondria posee dos membranas, una externa y otra interna, que engloban una matriz fluida. En este orgánulo se desarrollan los procesos que generan la energía celular.

Según la hipótesis endosimbionte, las mitocondrias provienen de una bacteria “independiente”. Ésta sería la explicación de que posean su propio material genético, el ADN mitocondrial (ADNmt) y su maquinaria para expresarlo.

Todo apunta a que el ADNmt es una reliquia de la supuesta bacteria simbiótica. Ya que aunque poseen su propio material genético, no son genéticamente autosuficientes, gran parte de sus proteínas funcionales y estructurales están codificadas por genes presentes en el núcleo celular.

Este tipo de ADN presenta ciertas características que lo hacen especialmente útil para estudios de identificación y evolución molecular, como por ejemplo el alto número de copias que posee cada individuo, la transmisión por vía materna, la presencia de numerosos polimorfismos, la completa caracterización de su genoma, su tasa rápida de evolución, etc.

3.1 ESTRUCTURA

El 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 16569 pares de bases (pb). Sus dos cadenas son complementarias y antiparalelas, asemejándose más a una ADN bacteriano que al ADN nuclear que está estrechamente asociado con proteínas y repartido en cromosomas.

Este ADN es un modelo de economía, porque sus genes están tan empaquetados que la mayor parte ($\approx 90\%$) de él parece tener alguna utilidad. La región codificante contiene 37 genes, de los cuales 22 son para ARN de transferencia, 2 para ARN ribosómicos y 13 para proteínas envueltas en la fosforilación oxidativa y en la producción de ATP [49].

La regulación de la molécula de ADNmt se sitúa en la **región control** con aproximadamente 1122 pares de bases de longitud y contiene los promotores de transcripción de ambas cadenas y el origen de replicación de la cadena pesada. Se encuentra entre la posición 16024 y la 16569 continuando desde la posición 1 hasta la 576.

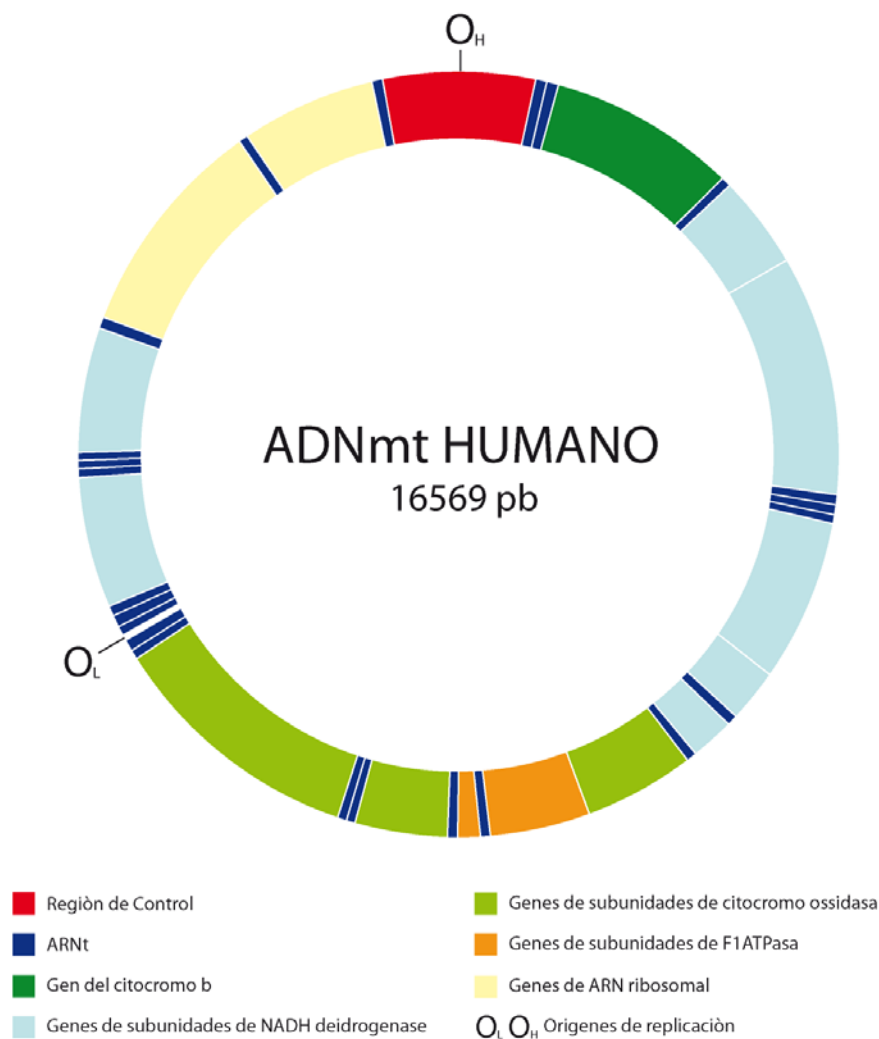


Figura I.5. Representación del ADNmt humano.

La región control tiene una alta variabilidad entre individuos, razón por la cual 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**, (16024-16365), la **región hipervariable 2, HV2**, (73 a la 340), y una tercera región hipervariable, menos utilizada, HV3, (440-560) (Lutz et al 2000). Existe una mayor variabilidad en la HV1 que en la HV2, por lo tanto ha sido el más usado en estudios de poblaciones humanas y del que se tiene una mayor información [50]. Otras regiones con un menor grado de variabilidad conocidas como **región variable 1 (VR1)** (16366-72) y **región variable 2 (VR2)** (341-576), poco utilizadas en Genética forense, pero cada vez más usadas en antropología molecular.

3.2 CARACTERÍSTICAS

Las características del ADNmt suponen una gran ventaja que hacen de él una fuente de información importante en multitud de áreas de estudio.

Un haplotipo mitocondrial es la secuencia de una determinada molécula de ADNmt, ya que el genoma mitocondrial es haploide. Esto es debido a que en humanos la herencia es estrictamente uniparental materna, **herencia materna**. Esto quiere decir, que salvo mutación, el ADNmt de hermanos, de hijos y madre y en general de aquellos individuos relacionados por vía materna es idéntico [51-53].

Las mitocondrias de los espermatozoides no contribuyen en la fertilización, debido en primer lugar a que en sus cabezas hay tan sólo unas pocas copias de ADNmt en comparación con los varios miles de copias de ADNmt del óvulo. Además de esto parece haber un mecanismo de reconocimiento específico que elimina las pocas mitocondrias paternas que pudieran introducirse en el óvulo. A consecuencia de todo esto el ADNmt no está sometido a procesos de recombinación [54] y su herencia es materna.

Este rasgo convierte al ADNmt en una herramienta muy útil para muchas disciplinas, puede ser utilizado con fines de identificación, para la reconstrucción filogenética de poblaciones humanas, las filogenias del ADNmt reflejarán directamente las genealogías maternas de los individuos estudiados.

El ADNmt posee menor información que el ADN nuclear, en cambio, existen muchas más copias por célula. Se estima un número entre 10 y 100 mitocondrias por célula y dentro de cada mitocondria entre 10 y 100 copias de ADNmt, por tanto hay un mínimo de 100 ADNmt por célula pudiendo alcanzar hasta 10000 copias [55,56]. 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.

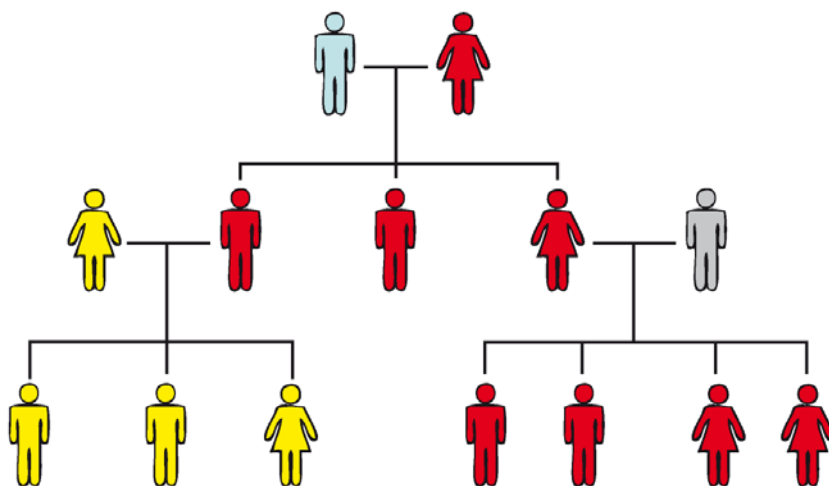


Figura I.6.
Herencia del ADNmt. Su herencia es siempre de la madre a todos sus hijos (tanto varones como hembras). Por su haploidía tiene limitada la recombinación.

Según los estudios de evolución molecular, debería haber un cambio evolutivo lento en el ADN de las mitocondrias. Cuando la función de un gen o proteína es importante, se espera un cambio evolutivo pausado en su estructura primaria. Debido a sus importantes funciones celulares la evolución del ADN de estos orgánulos debería estar muy restringida y sin embargo, el ADNmt tiene una **alta tasa de mutación**, de 5-10 veces mayor que el ADN nuclear. Esta

característica hace del ADNmt una herramienta muy útil para el análisis con fines de identificación y también para el estudio del proceso evolutivo [57].

La rápida evolución del ADNmt se debe a la aparición de mutaciones (tasa de mutación) y a que éstas se conserven (tasa de fijación). Las mutaciones surgen porque el ADN es sensible al daño oxidativo producido por los radicales libres, carece del efecto protector de las histonas y por la baja fidelidad de la polimerasa del ADNmt. Y se fijan por la carencia de mecanismos de reparación. Esto produce una alta hipervariabilidad entre la población humana.

Para utilizar el *reloj molecular* como método para datar eventos de interés es importante conocer la tasa de evolución del ADN. En el caso del ADNmt, la aplicación más relevante ha sido la estimación de la edad de los antecesores del ser humano actual.

La tasa de mutación en el ADNmt aun no está ajustada con certeza. Los primeros estudios realizados sugieren datos cercanos a una de cada treinta generaciones [51,58], pero a pesar de la herencia materna y de la ausencia de recombinación, lo cierto es que las mutaciones posibilitan que se puedan encontrar una o más diferencias en la secuencia de ADNmt en la rama materna de una familia. Se ha estimado que entre dos individuos de origen europeo no relacionados por vía materna, hay una media de 8 diferencias en la secuencia de su ADNmt [59].

La función principal de las mitocondrias es la producción de la mayor parte de la energía celular en forma de ATP mediante el sistema de fosforilación oxidativa (sistema OXPHOS). La alteración de esta síntesis puede dar lugar a una serie de trastornos que se conocen genéricamente como **enfermedades mitocondriales**. Este término se aplica fundamentalmente a los defectos en el sistema OXPHOS y particularmente a los defectos que son debidos a alteraciones en la secuencia del ADNmt, por ser los que mayoritariamente se han asociado directamente con estas enfermedades. El ADNmt codifica trece de los polipéptidos del sistema OXPHOS, mientras que el resto de las proteínas

están codificadas en el núcleo y se importan finalmente a la mitocondria ensamblándose con los anteriores.

En 1988 se encontraron las primeras mutaciones en el ADNmt asociadas a enfermedades. Poco a poco se han ido encontrando fenotipos asociados a ellas, creciendo de modo espectacular y ha generado lo que se podría llamar medicina mitocondrial [60,61].

El término de **heteroplasmia** define la presencia de tipos diferentes de ADNmt en la misma mitocondria, célula o individuo. Se denomina homoplasmia cuando se encuentra un único tipo de ADNmt.

La heteroplasmia se produce por errores en su la replicación o por el daño causado por radicales libres resultantes del metabolismo oxidativo. Cuando la heteroplasmia afecta a las moléculas de ADNmt del oocito puede transmitirse a la descendencia, que habitualmente manifiesta niveles variados de la misma debido a la segregación aleatoria de las mitocondrias durante la oogénesis.

Tabla I.2. Enfermedades ligadas al ADN mitocondrial, mutaciones y gen afectado.

ENFERMEDAD	MUTACION EN EL ADNmt	GEN AFECTADO	ENFERMEDAD	MUTACION EN EL ADNmt	GEN AFECTADO	
Mutaciones Primaria	G3460A	ND1	Miopatía Mitocondrial	T3250C	tRNA ^{Leu} (UUR)	
	G11778A	ND4		A3302G	tRNA ^{Leu} (UUR)	
	T14484C	ND6		C15990T	tRNA ^{Pro}	
Mutaciones Intermediad	G5244A	ND2	Sordera inducida por aminoglicosidos	A3243G	12S rRNA	
	G15257A	Citocromo b	Sordera sensoneural	T7445C	tRNA ^{Ser} (UCN)	
Mutaciones Secundarias	T3394C	ND1	Anemia sideroblástica	G12301A	tRNA ^{Leu} (CUN)	
	T4160C	ND1	Lipomatosis múltiple simétrica	A8244G	tRNA ^{Lys}	
	T4216C	ND1	Oftalmología progresica externa	A3243G	tRNA ^{Leu} (UUR)	
	A4917G	ND2		A5692G	tRNA ^{Asn}	
	G7444A	CO I		G5703A	tRNA ^{Asn}	
	T9101C	ATPasa 6	C3256T	tRNA ^{Leu} (UUR)		
	G9438A	CO III	LIMM	A15923G	tRNA ^{Thr}	
	G9808A	CO III	Muerte súbita	A15923G	A15923G	
	G13708A	ND5	Necrosis Bilateral del estriado	T9176C	ATPase 6	
	G13730A	ND5		T8851C	ATPase 6	
	G14459A	ND6	Multisistémica	A3251G	tRNA ^{Leu} (UUR)	
	G15812A	Citocromo b		A3253G	tRNA ^{Leu} (UUR)	
				C3256T	tRNA ^{Leu} (UUR)	
	MELAS	A3243G	tRNA ^{Leu} (UUR)	Corea y demencia	G5549A	tRNA ^{Trp}
		C3256T	tRNA ^{Leu} (UUR)	LHON y Distonia	G14459A	ND6
T3271C		tRNA ^{Leu} (UUR)	Diabetes y Miopatía	G14459A	ND6	
T3291C		tRNA ^{Leu} (UUR)	MERF	A8344G	tRNA ^{Lys}	
T9957C		CO III		T8356C	tRNA ^{Lys}	
Cardiomiopatía	A360G	tRNA ^{Leu} (UUR)	Leigh (MILS)	T8993G	ATPase 6	
	C3303T	tRNA ^{Leu} (UUR)	T8993C	ATPase 6		
	A4269G	tRNA ^{Ala}	Diabetes y Sordera	A3243G	tRNA ^{Leu} (UUR)	
	A4300G	tRNA ^{Ala}		T8993G	ATPase 6	
	A4317G	tRNA ^{Ala}	NARP	T8993G	ATPase 6	
	C4320T	tRNA ^{Ala}				
	G8363A	tRNA ^{Lys}				
T9997C	tRNA ^{Gly}					

La heteroplasmia del ADNmt suele estar relacionada con enfermedades mitocondriales [60,62-64] y el porcentaje de heteroplasmia suele guardar una relación directa con la severidad de la enfermedad en el individuo

(penetrancia). Inicialmente se consideró que la heteroplasma en regiones neutras del ADNmt como la región control era muy rara [65-70] pero hoy se piensa que pudiera ser mayor a lo estimado anteriormente [71-74]. Estas diferencias en la detección de la incidencia de heteroplasma pueden deberse a la escasa resolución de las técnicas de secuenciación, empleadas en los primeros estudios [70]. En la región control existe una elevada frecuencia de heteroplasma asociada a la longitud de los trectos de poli-Citosinas. En la HV1 la mutación 16189T/C, genera una heteroplasma de longitud en el tracto de poli-C entre las posiciones 16184 y 16193, que oscila entre 8 y 14 residuos [72,75,76]. En la HV2 existen varios casos, algunos asociados a ciertos haplogrupos europeos. Éste es el caso de la variabilidad asociada a la longitud del primer tracto de poli-Cs del motivo C7-T-C5 (posiciones 00303/00316).

En el haplogrupo J se han descrito dos variantes (C8-T-C6 y C7-T2-C6) con heteroplasma en la longitud del tracto citado. En algunos portadores del haplogrupo I europeo hay, también, heteroplasma asociada al tracto de poli-Citosinas en la misma región, a partir de la posición 00568. En la región control se ha descrito, asimismo, variabilidad en el número de repeticiones de CA a partir de la posición 00514 [72].

La heteroplasma en las regiones no codificantes es común en las células somáticas, lo que probablemente guarde cierta relación con el proceso de envejecimiento [71]. A su vez, la heteroplasma en las células germinales ha demostrado ser más común de lo que se afirmaba inicialmente [76-79].

Si surge una mutación en alguna molécula de ADNmt de un oocito, el individuo descendiente presentará inicialmente una mezcla de moléculas originales y mutantes. Para que el polimorfismo se establezca es necesario que la mutación pase a los descendientes, adquiera un estado homoplásmico a nivel individual y, finalmente, se fije en la población. La nueva mutación puede perderse en diversas fases del proceso. Sin embargo, hay evidencias de que la homoplasma puede llegar a alcanzarse en una o dos generaciones. La segregación aleatoria de las mitocondrias durante la oogénesis no permite

explicar la rapidez de este fenómeno por sí sola. Una posibilidad sugerida por varios autores es que, durante la oogénesis, se produzca una especie de cuello de botella en el que el número de moléculas de ADNmt se reduzca drásticamente [80,81]. En este caso, el ADNmt del oocito maduro y del embrión derivado del mismo descendería de un subconjunto del ADNmt total de la célula progenitora. Por tanto, un paso previo para todo polimorfismo es la heteroplasmia, debido al elevado número de copias de ADNmt.

3.3 POLIMORFISMOS EN EL ADNmt

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 [72,82].

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.

Por su herencia en bloque, se puede considerar el genoma mitocondrial como un locus en el que la combinación de mutaciones, inserciones y deleciones constituiría un alelo o haplotipo mitocondrial. Para una identificación correcta de todos los tipos mitocondriales debería secuenciarse el genoma mitocondrial completo. Existen actualmente kits específicos, como el mitoSEQr™ Resequencing de Applied Biosystem. Sin embargo, no es práctico cuando hay un elevado número de muestras, así que en muchos casos se escogen las zonas que se quieren estudiar, normalmente aquellas con más variabilidad.

Con la secuenciación de la totalidad o de ciertas regiones 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 [57,83]. En los primeros estudios, de baja resolución, se estudiaron numerosas muestras mediante una única enzima de restricción [84] o un número pequeño de éstas mediante varias enzimas [85]. La tendencia en estudios posteriores fue utilizar numerosas enzimas de restricción sobre un mayor número de muestras [86-89].

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.

4. APLICACIÓN DE LOS POLIMORFISMOS

El estudio de los loci polimórficos ha supuesto un gran avance en el campo de la identificación humana, en genética de poblaciones, en el desarrollo de los mapas de ligamiento humano y en nuevos procedimientos de diagnóstico clínico. Así, numerosos loci microsatélites han sido usados para el diagnóstico predictivo de muchas enfermedades genéticas, incluyendo distrofia miotónica, fibrosis quística, distrofia muscular de Duchenne/Becker y enfermedad de Huntington [90-93].

4.1 APLICACIÓN DE LOS STRs AUTOSÓMICOS

Dentro del área medico legal, las aplicaciones más importantes son:

Investigación de la paternidad: El análisis de los polimorfismos de ADN tiene como objetivo la confirmación de que un individuo es o no el padre biológico de otro. En España, son numerosas las sentencias del Tribunal Supremo que reconocen la utilidad de la investigación biológica de la paternidad con este método. Se puede realizar esta prueba si se dispone de todas las muestras necesarias o también en casos especiales de difícil solución, cuando se carece de muestra perteneciente al presunto padre [94]. Debido a la poca cantidad de muestra biológica necesaria también se puede hacer antes del nacimiento, con el análisis de líquido amniótico para determinación de paternidad en fetos [95], paternidades a realizar sobre restos óseos, en ausencia de madre, con familiares en ausencia de ambos progenitores, etc. En muy poco tiempo esta prueba se ha estandarizado, obteniendo gran información con mucha rapidez y menor coste relativo.

Criminalística: Cuando se necesita analizar vestigios mínimos, degradados u otras muestras (como espermatozoides, saliva, pelos, cabellos sin bulbo, fragmentos óseos, etc) los STRs son una herramienta muy adecuada. Son de gran ayuda, principalmente en delitos donde el presunto culpable niega los hechos, no suele

existir más indicio incriminatorio que el proporcionado por posibles restos de sangre o espermatozoides en prendas de vestir y cavidades corporales.

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, la posibilidad de diferenciar fácilmente muestras compuestas por mezclas de fluidos procedentes de distintos individuos [96]. En este momento están a la venta software informáticos específicos para este tipo de muestras [97]. Y en el caso particular de agresiones sexuales, permiten separar el ADN de células vaginales del ADN espermático [98].

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. Aún en restos muy antiguos es posible analizar los STRs partiendo, por ejemplo, de piezas dentarias o fragmentos óseos. La comparación con el ADN obtenido de muestras de sangre de algún ascendiente o descendiente indubitado (preferiblemente de los progenitores) suele ser suficiente para establecer una relación de parentesco.

En situaciones de grandes catástrofes, 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 [99], 11-M en Madrid [100] o accidentes como el de Barajas el 20 de Agosto del 2008.

4.2 POLIMORFISMOS EN CROMOSOMAS SEXUALES

La detección y validación de un gran número de secuencias polimórficas presentes en los cromosomas sexuales hizo posible la realización de estudios poblacionales más completos y su aplicación inmediata tanto en estudios de filiación deficientes como en casos forenses [101,102].

Las aplicaciones en **genética forense** se hicieron patentes rápidamente, sobre todo en los casos de criminalística donde aparecían agresiones sexuales, donde las muestras forenses muestran mezcla celular (víctima y agresor). En mezclas desbalanceadas, donde la presencia de células masculinas y femeninas es muy desproporcionada, será muy difícil poder diferenciar los perfiles del agresor y de la víctima. En el caso del uso del Y-STRs se podrá diferenciar perfectamente el perfil de un sospechoso, por mucho ADN de la víctima que se halle en nuestra muestra. En el caso de X-STRs este tipo de análisis tienen mayores limitaciones, de todas formas puede ser un complemento muy importante en muchos análisis tanto de filiación, como de criminalística [46].

Existen algunas limitaciones en sus aplicaciones con respecto a los STRs autosómicos. La herencia en bloque de estos marcadores exige un análisis estadístico global, donde el conjunto de frecuencias genéticas obtenidas constituye un haplotipo [101,103]. Aunque el tipado de un mayor número de marcadores polimórficos por muestra aumenta de forma notoria el poder discriminativo y el potencial informativo de estos microsatélites [104,105].

Estudios de evolución: los grupos de ligamiento ancestrales que reflejan el origen del cromosoma Y han sido largamente conservados a lo largo de la evolución, si bien, 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 [106].

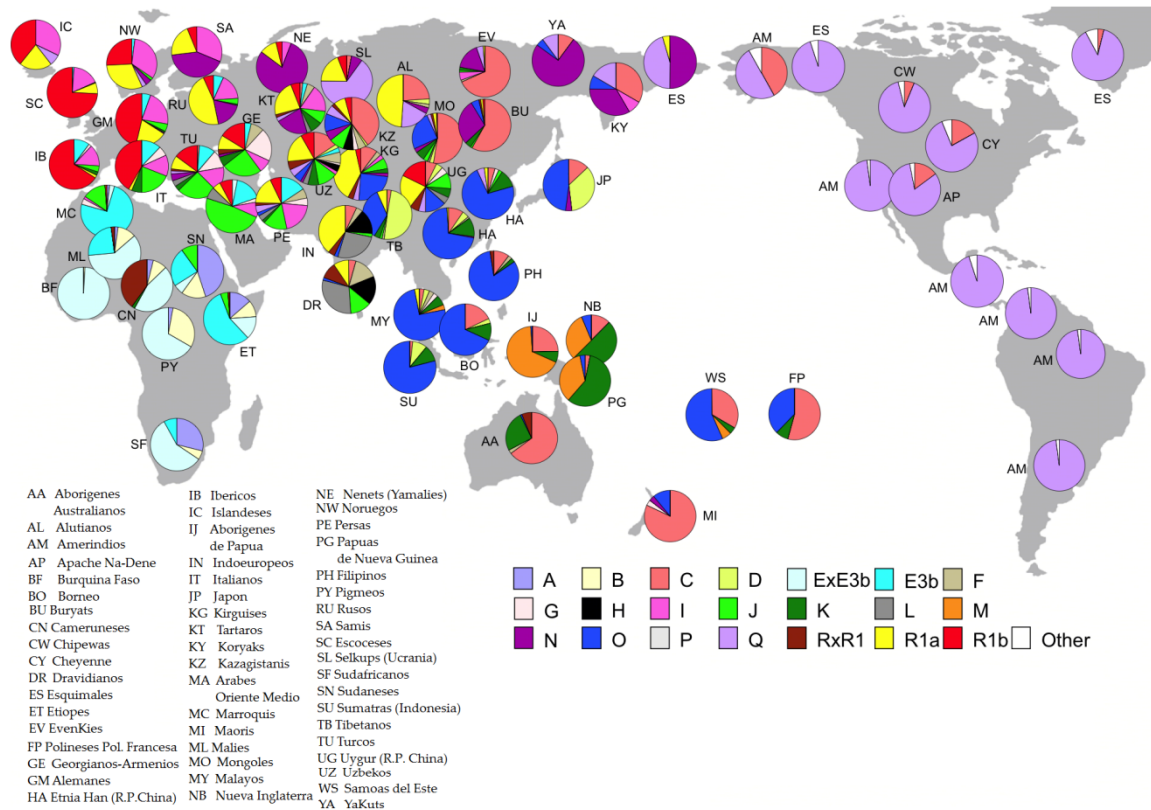


Figura I.7. Frecuencias de distribución de haplogrupos de cromosoma Y. Esquema traducido y modificado a partir de la imagen original representada por JD McDonald (2005). En los europeos el haplogrupo más frecuente es el R1b en los americanos predomina el Q.

La función que determina el sexo del cromosoma Y implica que es heredado de manera patrilineal y haploide. A causa de esta haploidía, la mayor parte de este cromosoma no recombina con ningún otro durante la meiosis. Estas propiedades tienen importantes consecuencias para la historia evolutiva de las poblaciones. El cromosoma Y pasa de padre a hijo sin cambios, excepto por la acumulación gradual de mutaciones; por ello, el retroceso en los linajes patrilineales hacia el pasado depende de la tasa de mutación de los marcadores del cromosoma Y analizados. El estudio de los polimorfismos del cromosoma Y con fines evolutivos pretende esclarecer en último término, algunos de los problemas todavía pendientes en lo que respecta al origen y la dispersión de los humanos modernos. Estos polimorfismos son de gran ayuda, ya que mientras los cromosomas autosómicos y el X tienen múltiples ancestros a causa de la recombinación, todos los cromosomas Y tienen un único ancestro paterno, aunque no es posible predecir a priori si el ancestro fue un humano reciente o un pre-humano [106]. Esto supone ventajas en el estudio de la evolución

humana, la ausencia de la recombinación simplifica la historia contenida en la molécula del ADN del cromosoma Y [107].

Se ha encontrado en los Y-STRs una herramienta perfecta para el estudio de las migraciones y su resultado en las poblaciones actuales. Un ejemplo de ello sería la estructura poblacional de la Península Ibérica, descrita a partir del cromosoma Y, pudiendo calcular los distintos aportes por migración de musulmanes y judíos a la población actual de la península [108].

En esta última década la **genealogía** es otra de las aplicaciones más frecuentes del estudio de los Y-STRs. La concordancia entre la herencia del apellido y la de las regiones no recombinantes del cromosoma Y, hace de ésta una herramienta idónea para realizar estudios de genealogía [109,110]. El estudio de este cromosoma permite el análisis de la herencia de un apellido para saber su origen, su difusión y las distintas migraciones que han hecho que un apellido se conserve en unas zonas o que se disgregue y se traduzca a muchísimos idiomas [111,112] e incluso seguir un apellido para poder conocer el origen de un personaje histórico [113].

En la **investigación en biomedicina**, por asociación a enfermedades se está generalizando el uso de estos marcadores. En muchas ocasiones cuando se realizan estudios de asociación de genes o polimorfismos para una enfermedad, como puede ser la diabetes, el cáncer de mama, ovario o próstata en una población determinada, se ven resultados bastante claros, en cambio en otras poblaciones estos mismos marcadores no producen los mismos resultados. Por tanto, uno de los problemas que se encuentra, es que los marcadores genéticos asociados a riesgos de enfermedad no son reproducibles en todas las poblaciones, por tanto hay que realizar réplicas de todos estos experimentos en todas las poblaciones. Pero se presentan muchos problemas, no sólo por el enorme número de muestras a estudiar, sino porque en poblaciones mixtas los resultados obtenidos pueden ser fruto de no haber podido clasificar las muestras estudiadas. Por eso 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 [114-116].

4.3 APLICACIONES DEL ANÁLISIS DE ADN_{mt}

Investigación de la hermandad y maternidad: El estudio del ADN_{mt} tiene unas aplicaciones concretas en Genética Forense y ha venido a complementar al análisis de polimorfismos del ADN nuclear. Su uso cobra especial interés 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 [117-120] o bien cuando se ha de realizar identificación de restos óseos contando con parientes maternos, donde habitualmente la cantidad de ADN nuclear esperada es baja o muy baja [121].

Muy frecuentemente los vestigios biológicos de interés en criminalística se encuentran en cantidades ínfimas, sufren deterioro o están contaminados de diversas formas, lo que dificulta enormemente su análisis.

La determinación de polimorfismos de ADN, en especial mediante la técnica de PCR, es un procedimiento óptimo cuando se necesita analizar manchas de sangre minúsculas, degradadas u otras muestras (como espermatozoides, saliva, pelos o cabellos sin bulbo, fragmentos óseos, etc).

En criminalística, a veces, 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 ADN_{mt} 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 inclusión basada únicamente en los polimorfismos del ADN_{mt} no es factible, ya que todos los familiares pertenecientes al mismo linaje materno comparten el mismo haplotipo y se encontrarían inexcusablemente implicados.

Utilizando sistemas autosómicos, la probabilidad de que dos individuos posean idéntico perfil es habitualmente muy baja, sin embargo, la determinación

aislada de polimorfismos de ADNmt incluye a todos los familiares por vía materna por compartir idéntico haplotipo. Pero, como consecuencia del alto grado de especificidad poblacional de algunos haplotipos de ADNmt, sería razonable hacer suposiciones sobre el origen poblacional de una muestra de ADN hallada en la escena del delito pudiendo acotar la búsqueda en el caso de portar un haplotipo característico de un grupo particular.

Evolución y genética de poblaciones: En el contexto de las poblaciones humanas cabe destacar sus aplicaciones en Antropología Física o Biológica. 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 [122]. Por ejemplo; la relación entre el hombre y los homínidos africanos [123-127], el origen y dispersión del hombre moderno [86,128-134] o la relación entre el hombre anatómicamente moderno y el Neandertal [133,135-139].

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

Como herramienta filogenética [140-144]: La variación en el ADNmt puede ser usada para construir árboles filogenéticos, o varios árboles alternativos dispuestos en una red [145-148], que describen la relación entre secuencias individuales. La estructura de este árbol contiene información demográfica que, junto con una tasa de mutación calibrada para la secuencia en estudio, puede ser utilizada para estimar la datación de eventos prehistóricos. Además, utilizando lo que es conocido por aproximación filogeográfica, ya anteriormente mencionada, la distribución de los linajes en un árbol o red puede ser usada para detectar movimientos de una región a otra [140,149-151]. 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.

Estudios de asociación a enfermedades: Como se cita en las aplicaciones de los STRs en cromosomas sexuales, la asociación de genes a enfermedades está muy ligada al estudio de las poblaciones. Por eso el ADNmt es una herramienta muy importante al igual que los STRs en cromosomas sexuales, debido a sus similitudes.

5. IDENTIFICACIÓN GENÉTICA EN MEDICINA LEGAL

Gisbert Calabuig, 1983, afirmaba que *identificar a una persona, establecer su individualidad, es determinar aquellos rasgos o conjunto de cualidades que la distinguen de todos los demás y hacen que sea ella misma.*

El estudio de caracteres hereditarios permite hacer diagnósticos de identificación. Estos caracteres deben cumplir ciertos requisitos: en primer lugar ser altamente variables, para poder diferenciarlos entre sí, poseer un modelo de herencia conocida y fijada, y por último, tienen que ser inmutables a lo largo de la vida del individuo e idénticos en los distintos tejidos de su organismo.

La identificación humana en Medicina Legal es un proceso de comparación entre los resultados obtenidos en el análisis de los indicios y los obtenidos a partir de las muestras de origen conocido (muestras indubitadas). Las posibilidades de que exista otra persona que comparta los resultados de los análisis van disminuyendo cuando se aumentan los estudios sobre las muestras que se comparan. Los resultados obtenidos con el estudio del ADN permiten identificar a un individuo con una probabilidad muy próxima al 100%. A principio de la década de los 80 comenzó a realizarse el estudio de los polimorfismos de ADN lo que revolucionó las técnicas de identificación. Con un número suficiente de regiones de ADN es posible demostrar que la coincidencia entre dos personas es ínfima.

La probabilidad de que dos individuos no emparentados tomados al azar posean el mismo perfil de ADN después de estudiar 5 loci hipervariables, puede llegar a ser de uno entre un billón. Ahora que la mayoría de los estudios se realizan con más de 13 loci hacen que se convierta en un método de identificación prácticamente absoluta (excluyendo a los gemelos monocigóticos).

6. ESTUDIOS POBLACIONALES Y BASES DE DATOS APLICADOS A LA IDENTIFICACIÓN GENÉTICA

Para seleccionar un STR de uso forense se requiere que cumpla una serie de requisitos: una alta heterocigosidad, parámetro estadístico que está relacionado con el polimorfismo de un determinado marcador; un valor alto de heterocigosidad es indicativo de un alto grado de polimorfismo. También una tasa de mutación baja, ya que marcadores con una tasa de mutación alta podrían presentar variaciones de una generación a otra y, en el caso de un estudio biológico de paternidad originar una falsa exclusión. Debe ser fácilmente amplificable, de pequeño tamaño y fácilmente coamplificable junto a otros STR en una misma reacción de PCR. Existen kits comerciales que proporcionan la posibilidad de amplificación conjunta de un número determinado de STRs: PowerPlex e Y-Powerplex, de Promega o Identifiler e Yfiler de Applied Biosystem, que amplifican 15 loci tipo STR autosómicos ó 16 en el caso de los Y-STRs.

Por tanto, cualquier STR del genoma no puede ser utilizado para la identificación genética. Un requisito fundamental para poder utilizar nuevos polimorfismos de ADN con fines identificativos y obtener resultados altamente fiables es el estudio y caracterización de dichos polimorfismos en la población que posteriormente se utilizará como referencia. Esto es lo que en Genética Forense se denominan **Estudios Poblacionales**.

En primer lugar, es necesario un grupo de individuos (o muestra) de la población donde se vaya a utilizar el grupo de STRs. Dependiendo del tamaño de la población de estudio así deberá de ser el tamaño de muestra recogida, para que los resultados sean significativos. Es de suma importancia que los individuos de la muestra que van a analizarse no tengan ninguna relación de

parentesco entre sí, ya que la consanguinidad puede alterar las frecuencias génicas.

El estudio poblacional tendrá como finalidad determinar las frecuencias alélicas y genotípicas para cada uno de los marcadores que se estén estudiando. Es importante confirmar que el locus o los loci estudiados se encuentran en equilibrio de Hardy-Weinberg y que estos STRs se encuentran en independencia con respecto al grupo analizado. Por último, es necesaria la comparación de los resultados obtenidos en nuestra población con los obtenidos para otras poblaciones, y así observar las diferencias y similitudes con poblaciones más cercanas y lejanas filogenéticamente hablando. [152-154].

Una vez obtenidos los perfiles genéticos de la población de referencia el laboratorio comienza a realizar análisis de la población, ya sea con fines de criminalística o de investigación. Las nuevas técnicas de amplificación así como las plataformas de genotipado y secuenciación generan una cantidad de información extraordinaria que va almacenando el laboratorio. Los perfiles obtenidos por un laboratorio forense deben ser clasificados y almacenados en bases de datos.

Las **bases de datos**, en el presente, son una herramienta informática donde se almacena de modo ordenado y coherente la información obtenida, para poder ser rescatada de modo automático de acuerdo a parámetros previamente establecidos. El manejo apropiado de los datos acumulados que se van generando en diversas investigaciones produce una de las más importantes vías de conocimiento, algunas de ellas inicialmente ni siquiera relacionadas entre sí. Por tanto, acumular datos es necesario, y nadie puede acotar la libre generación de los mismos, a su vez son útiles a la sociedad en general y a las personas en particular.

Pero es imprescindible que las bases de datos de ADN estén perfectamente clasificadas y reguladas [155] y que ofrezcan las máximas garantías a los ciudadanos, y por ello conviene conocer perfectamente lo que son y lo que contienen.

Por todas estas razones ha existido una gran preocupación con respecto a las bases de datos genéticos, en estos últimos años, que se ha subsanado, en muchos países, con una regulación adecuada. Esta regulación ha tenido en cuenta el tipo de datos que posee cada base, su uso, el personal que puede acceder a ella, etc.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	
1	Sample Info	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA															
2	BU-211	10	14	29	31	10	13	12	12	16	18	6	7	8	8	12	13	19	22	13	14	14	16	8	9	12	17	9	13	19	24
3	BU-213	11	13	29	31.2	12	12	11	12	14	15	6	6	9	9	10	11	23	23	13	15	16	17	8	12	12	15	11	14	25	26
4	BU-219	13	13	32.2	33.2	11	12	12	12	16	16	7	9.3	8	9	12	13	21	26	12	14	16	18	8	9	12	13	7	11	19	15
5	BU-225	13	14	29	29	10	10	12	12	17	18	6	9	8	12	11	12	17	24	12	14	14	19	8	11	18	18	11	12	19	21

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
referencia	DYS 456	DYS 389 I	DYS 390	CYS 389 II	DYS 458	DYS 19	DYS 385	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	Y GATA H	DYS 437	DYS 438	DYS 434		
EY001	16	13	25	29	16	14	12	14	14	11	12	23	13	11	15	12	19	
EY002	14	14	23	29	16	13	13	16	13	10	12	21	13	11	15	9	19	
EY003	15	12	24	28	17	14	11	11	13	11	12	23	13	12	15	12	19	
EY004	16	13	24	29	17	14	11	14	14	11	13	23	13	12	15	12	19	
EY005	15	12	23	28	17	15	11	14	13	11	12	23	13	12	15	12	18	
EY006	13	12	27	28	16	15	13	16	13	10	11	21	11	10	15	10	19	
EY007	16	13	24	29	18	14	11	14	13	11	13	23	13	10	15	12	18	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	HV1	1	2	3	4	5	6	7	8	9	10	2	HV1	2	3	4	5	
2	MEST-001	16111T	16189C	16223T	16290T								MEST-213	16111T	16156A	16183C	16189C	16290T
3	MEST-003	16183C	16189C	16217C	16270C								MEST-219	16111T	16168C	16223T	16290T	16290T
4	MEST-006	16111T	16223T	16290T	16319A								MEST-232	16183C	16189C	16223T	16325C	16325C
5	MEST-010	16111T	16140C	16192T	16209C	16223T	16290T	16319A	16362C				MEST-233	SIN POLIMORFISMOS				
6	MEST-012	16111T	16223T	16290T	16299G	16319A	16362C						MEST-241	16111T	16129A	16136C	16183C	16183C

Figura I.8. Bases de datos originadas en este estudio. En la parte superior base de datos de cromosomas autosómicos; en la media STRs en cromosoma Y; en la inferior polimorfismos en el ADNmt.

Las bases de datos sólo pueden contener resultados, informes alfanuméricos (si se habla de ADN extraído o material biológico se estará hablando de bancos o archivos biológicos y estos deben tener otra regulación específica). Los perfiles de ADN sí son resultados alfanuméricos que tienen vinculación con una persona. Son bases virtuales, o sea, datos que existen simplemente en archivos informáticos o impresos en papeles. El acceso a los datos debe 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.

La **clasificación de las bases de datos en genética forense** depende básicamente de su finalidad.

- Generales: comprenderían a toda una población o país, y en este momento no hay ninguna en marcha en el mundo. La más amplia corresponde a la población de Islandia, es con consentimiento de los donantes dedicado a fines médicos de

investigación de características de interés genético biomédico, al tratarse de una población peculiar.

- Profesionales: incluyen a profesionales de riesgo, siendo el principal y mayor de los ejemplos el reservorio de material biológico de las fuerzas armadas de los Estados Unidos. Realmente es un archivo biológico de muestras de sangre que han sido obtenidas con consentimiento y sólo se analizan en caso de necesidad. Este material se usará exclusivamente para identificar a los restos del donante en caso de que haya fallecido en trágicas circunstancias o haya sido dado por desaparecido (la posibilidad de acceder a un material directo de la persona facilita enormemente el análisis comparativo de ADN).

- Judiciales o Forenses: pueden ser de tipo criminal y de tipo civil. Se analizan con detalle en el apartado siguiente, pero se resumen sus características básicas a continuación.

- Bases de datos forenses criminales: de modo general, almacenan datos procedentes de personas que han sido procesadas o condenadas, en algunos casos pueden considerarse perfiles de víctimas conocidas, con objeto de facilitar la resolución de delitos. Su característica principal es que algunas de las muestras y datos considerados se obtienen sin el consentimiento de las personas implicadas. LEY ORGÁNICA 10/2007, de 8 de octubre [156], reguladora de la base de datos policial sobre identificadores obtenidos a partir del ADN.

-Bases de datos forenses civiles: su único fin es la identificación de personas (niños y adultos) desaparecidos, lo cual se hace comparando el ADN de las personas no identificadas (normalmente restos óseos) con el de los familiares. Por sus características humanitarias, es requerimiento ético que los familiares colaboren voluntariamente y tras firmar un consentimiento informado [157,158].

Por lo tanto, las bases de datos de utilidad en el área forense o judicial son bases de datos de criminales de identificación genética (BDCIG), que contienen sólo

datos alfanuméricos de identificación y cuya finalidad es ayudar a la Justicia en la resolución de casos de criminalística o en identificaciones civiles.

Existen diferentes bases de datos en funcionamiento en este momento, y entre todas ellas, la de mayor aplicación es el **sistema CODIS** (Combined DNA Index System) desarrollado en Estados Unidos por el FBI, y de aplicación positiva en la práctica. Es un gran instrumento de trabajo debido a que permite la compatibilidad y el trabajo independiente de diferentes laboratorios; posibilita el acúmulo e intercambio de datos a diferentes niveles: Local (L-DIS), Estatal (S-DIS) y Nacional (N-DIS) cuya unión conforma el sistema CODIS; contiene diferentes apartados o índices, incluyendo los criminales y civiles de personas desaparecidas; su instalación es gratuita así como el asesoramiento por parte de los Estados Unidos a través del FBI a los laboratorios oficiales de los diferentes países, siempre que haya una petición oficial y se den una serie de circunstancias favorables.

6.1 PD y PE

El estudio de los polimorfismos y de sus frecuencias en la población ofrece las herramientas necesarias para poder identificar a una persona, teniendo su perfil o el de sus familiares. Una vez hecho el estudio de una población se puede hacer el cálculo de multitud de índices que caracterizan la validez del sistema en esa población determinada. Estos índices se denominan "a priori", al ser estadísticos que se conocen antes de la llegada de la muestra o evidencia al laboratorio.

Siempre se compara el perfil de una muestra de referencia, muestra de origen conocido o indubitada, con otros perfiles, coincidentes, no coincidentes o relacionados. Cuando se tiene acceso a las muestras se harán otra serie de análisis que darán el resultado definitivo. Pero si se desea tener resultados concluyentes, estos parámetros tienen que ser óptimos, por ejemplo, si se quiere obtener una alta razón de verosimilitud (LR) hay que partir de un sistema que tenga un gran poder de discriminación y exclusión.

6.1.1 Poder de Discriminación (PD)

El poder de discriminación es la probabilidad de que dos individuos no relacionados puedan ser diferenciados genéticamente mediante el análisis de uno u varios marcadores, igualmente, se puede analizar un vestigio, si bien se debe diferenciar de otro tomado al azar. Depende del número de loci analizados, de las variables alélicas encontradas y frecuencias de cada locus, en cada población.

$$PD = 1 - \sum_{i=a}^n G_i^2$$

Donde G_i es la frecuencia de haplotipos i en la población.

El PD es una medida relativa de la eficacia del sistema o sistemas analizados. Un PD cercano a uno sería, una situación deseable para establecer una identificación.

6.1.2 Poder de Exclusión a Priori (PE)

El poder de exclusión es la probabilidad de que un sistema específico muestre evidencias que conduzcan a la exclusión de un sospechoso (Boyd, 1954). Éste es un estadístico que predice como se producirá la identificación en el análisis a posteriori.

$$PE = h^2 (1 - 2 * h * H^2)$$

Donde H es el número de individuos homocigóticos y h el número de heterocigóticos.

El valor obtenido es un estadístico porcentual, que depende del grado de polimorfismo de un marcador y del equilibrio de las variables encontradas en el marcador.

7. GENÉTICA FORENSE Y ANTROPOLOGÍA MOLECULAR

Existen grandes analogías entre la genética forense y la antropología molecular. Como se cita anteriormente, ambas analizan la misma herramienta; caracteres heredables que se transmiten de generación en generación, con una herencia y una tasa de mutación conocida.

Los marcadores escogidos por ambas ciencias, son comunes en su mayoría, aunque en genética forense es más habitual el uso de STRs autosómicos y STRs en cromosoma Y. El uso de ADNmt y de SNPs, está reducido a grandes laboratorios o centros de investigación. En antropología molecular, ocurre justo lo contrario, es más frecuente el uso del ADNmt y SNPs que el de STRs. Esto se debe a las diferencias de los análisis filogenéticos que se van a estudiar. En el caso de estudios de herencia inmediata siempre los STRs serán mucho más informativos, por tanto mucho más interesantes en el caso de la identificación. Todo lo contrario ocurre si se está estudiando poblaciones antiguas o la evolución del ser humano, donde interesan locus sin recombinación o con recombinación reducida y con tasas de mutación baja.

Ambas ciencias se han complementado desde sus comienzos, pero el trabajo con ADN antiguo (aDNA) ha permitido que ambas se solapen, y en muchas ocasiones utilicen centros de trabajo comunes.

Los grandes avances en las técnicas forenses han permitido que se puedan obtener perfiles genéticos de muestras que hace unos años eran impensables, obtener el perfil genético de un solo pelo, de la saliva de un sello, de huellas sobre superficies, de fragmentos de huesos en muy mal estado, etc. El procesamiento de este tipo de muestras es comparable al análisis del aDNA [159], el ADN molde está normalmente en muy pequeña cantidad, y generalmente muy degradado, e incluso puede haber modificaciones químicas que dificultan su amplificación. También es un problema común la

contaminación de las muestras por ADN moderno, haciendo necesario en ambos casos laboratorios con equipamientos específicos para evitar este problema, así como controles rigurosos de los resultados emitidos.

En general los resultados obtenidos en muestras con ADN en baja concentración "*low copy number*", suelen tener los mismos problemas: pérdida de marcadores de alto peso molecular, aumento de la homocigosis por pérdida alélica, aparición de extra picos debidos a la degradación, etc. Estos errores se ponen aún más de manifiesto si aumenta el número de ciclos en la amplificación. La interpretación de estos datos, hace necesaria la repetición de los análisis y los contraanálisis en otros laboratorios [160]. Esto ha hecho continuo el trabajo común y la complementación de laboratorios de antropología molecular y genética forense. Proyectos como la identificación genética de los restos del Zar Nicolás II [66,161-163] o la identificación genética de Colón y sus familiares, indican la necesidad del trabajo común de ambas ciencias.

8. ADN Y EL ESTUDIO DE LAS POBLACIONES HUMANAS

Cuando se realizó la secuenciación del ADN mitocondrial (ADNmt) por primera vez [49] 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. Pero el ADNmt no es la única molécula que tiene un gran potencial filogenético. El cromosoma Y, portador de los genes determinantes del sexo masculino, también se hereda uniparentalmente y no

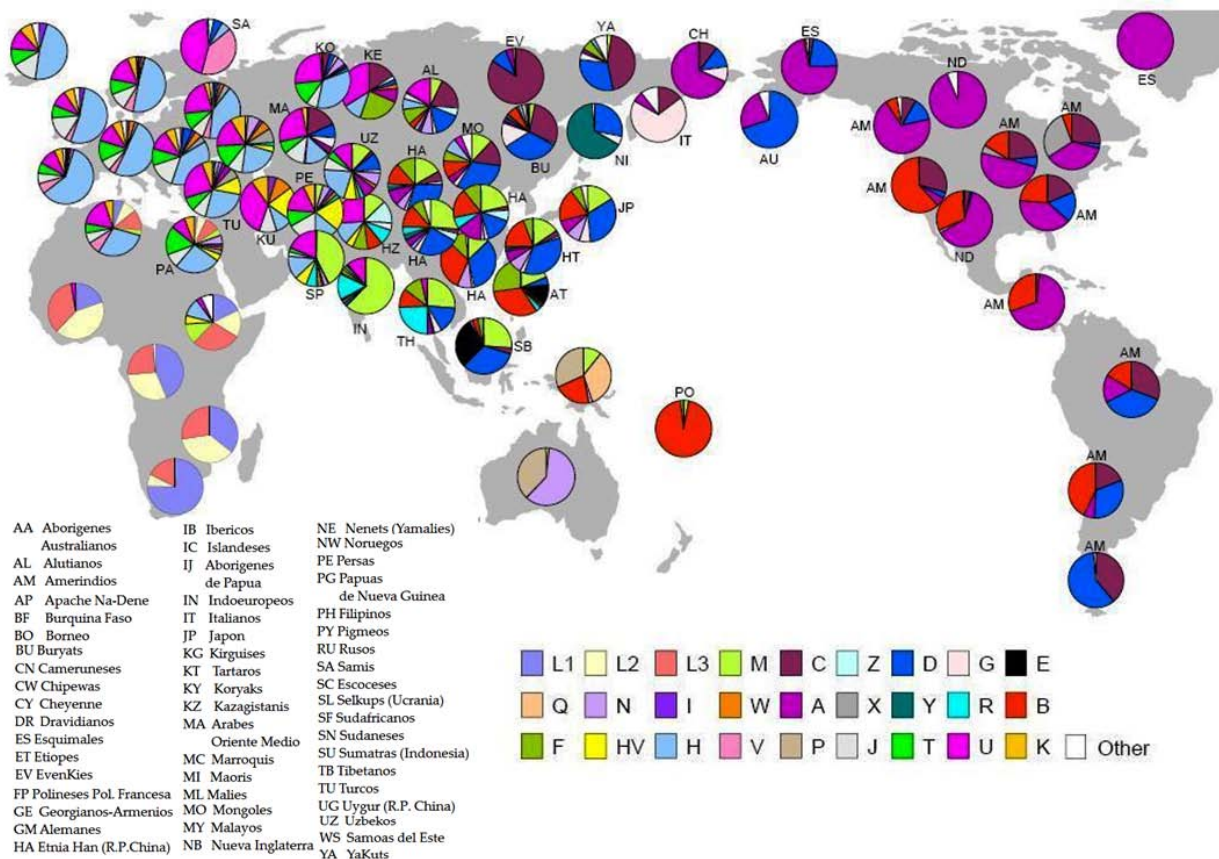


Figura I.9. Frecuencias de distribución de haplogrupos de ADNmt. Imagen traducida y modificada a partir de la original representada por JD McDonald (2004). Como se puede observar, en los europeos hay mayor diversidad de haplogrupos, siendo el más frecuente el H. Sin embargo, en los americanos, y más acentuado en Centroamérica predomina el A.

padece fenómenos de recombinación en prácticamente toda su extensión [164]. Esta falta de recombinación hace que el cromosoma se comporte como un gran fragmento único de ADN, portador de un haplotipo que se transmite intacto de padres a hijos, a no ser que actúen fenómenos de mutación.

El ADNmt presenta como ya se ha citado, una tasa de sustitución de bases diez veces superior a la del ADN nuclear, lo que significa que existe un gran número de secuencias diferentes de ADNmt en la población, por tanto, una gran diversidad. Al mismo tiempo, esta fuerte tasa de mutación también es causa de inconvenientes: se pueden producir mutaciones puntuales varias veces en el mismo punto o aparecer independientemente en diferentes linajes, haciendo imposible determinar el estado ancestral. Por el contrario, algunas mutaciones puntuales en el cromosoma Y se pueden considerar como acontecimientos únicos que son específicos de un linaje, ya que cuando un grupo de cromosomas Y portan una mutación puntual dada es muy probable que tengan el mismo origen.

8.1 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. Los polimorfismos genéticos (PG) son características genéticas que poseen todos los individuos de una población, que de igual modo se puede comparar con el resto de las poblaciones y entre los individuos. Las secuencias del ADN que tienen interés para ese uso son aquellas que están presentes en las poblaciones en forma de variantes individuales debidas a dos o más formas alternativas (alelos): la existencia de estas diferencias se denomina polimorfismo.

El grado de diversidad de los individuos respecto a un fragmento del genoma y su distinta distribución en los diversos grupos humanos está altamente relacionado con la utilidad de un PG como marcador antropológico.

Los PG deben tener una serie de **características para poder ser marcadores antropológicos**:

- Un origen único, por tanto, que sea monofilético (sólo es posible en casos donde la frecuencia de los eventos moleculares que hayan generado los alelos sea tan pequeña que haga irracional suponer que se haya originado en más de una ocasión). Estos PG, llamados únicos, son los instrumentos ideales para las reconstrucciones filogenéticas. Encontrar el mismo polimorfismo en dos o más poblaciones distintas permite inferir que éste sea idéntico por descendencia, que existía ya en una población ancestral común y no que se haya originado por mutaciones independientes. Los PG recurrentes deben ser utilizados con una cierta cautela por su posible polifiletismo (estos pueden haber aparecido de la misma forma más veces y en más lugares independientemente).
- Estos deben ser selectivamente neutros. Sólo así las informaciones que aportan podrán ser utilizadas para comprender la historia evolutiva (relaciones de antecesor-descendiente).
- La última característica es su grado de variabilidad entre poblaciones o capacidad informativa. El marcador será informativo cuando su variabilidad interpoblacional (politipia) sea superior a la intrapoblacional (polimorfismo).

Los PG del ADN se pueden subdividir respecto a su situación en el genoma, función y a su grado de variabilidad [165].

Respecto a su posición, los PG se pueden dividir en: mitocondrial y nuclear, que a su vez se divide en autosómico y heterocromosómico o par sexual (de cromosoma X y del cromosoma Y). Si se refiere a la sede funcional pueden ser: PG de regiones intergénicas o alejados de los genes estructurales entre los cuales están interpuestos; como es el caso de muchos STRs.

Al analizar la variabilidad, se puede destacar PG de elevada variabilidad intrapoblacional (con un grado de heterocigosidad H elevado) y otros de alta variabilidad interpoblacional.

Los PG de alta variabilidad intrapoblacional se pueden diferenciar en:

- Secuencias de alta variabilidad haplotípica, entre las cuales están el cluster (conjunto de genes que confiere una característica) de los genes estructurales de los antígenos del HLA.
- Sitios polimórficos puntuales constituidos por una secuencia repetida en tándem "n" veces. (Ej. STRs, Alu, RFLP).

Los STRs son marcadores particularmente ilustrativos. Por tanto, la información obtenida de estos marcadores no resulta siempre simple y unívoca a causa de posibles fenómenos de homoplasmia, que pueden ser verificados en el curso de las generaciones. Aunque es razonable afirmar que poblaciones con repertorios micro- y minisatélites muy diversos entre sí sean distintas, no se puede concluir lo contrario cuando se encuentra dos poblaciones con igual o similar frecuencia en estos, no tiene porque implicar relaciones evolutivas estrechas, puesto que la semejanza genética podría ser el resultado de fenómenos de convergencia evolutiva.

Una de las herramientas más útiles para realizar estudios entre poblaciones es construir haplotipos, mediante la combinación de diferentes formas alélicas de PG diversos (bien sean polimorfismos únicos o recurrentes) y analizar las frecuencias de estos haplotipos en las poblaciones. Si bien, existe un gran riesgo de homoplasmia para un simple locus, la probabilidad cuando se analiza más loci disminuye cuanto mayor sea el número de estos.

Los polimorfismos genéticos más utilizados para hacer haplotipos y haplogrupos son los que poseen una recombinación limitada, ADNmt y Cromosoma Y (Y-SNPs y Y-STRs).

Cuando se comenzaron a buscar PG en cromosoma Y, mediante lugares de restricción, no se evidenciaron variaciones apreciables, ya que el número de

polimorfismos encontrados era extremadamente bajo respecto al de otras partes del genoma [166,167]. El cromosoma Y está constituido mayoritariamente por secuencias repetidas y por no estar sujeto a vínculos funcionales puede acumular más mutaciones [168]. Posteriormente, el descubrimiento de nuevas tecnologías de análisis de ADN ha aumentado de forma progresiva el espectro de polimorfismos individuales en cromosoma Y, de modo análogo a lo que ha sucedido para los polimorfismos autosómicos y mitocondriales. De todos estos estudios se ha concluido que la tasa de mutación del cromosoma Y es bastante más baja respecto a la de otros cromosomas nucleares y del ADNmt. Esto puede ser debido al diferente número de cromosomas autosómicos y de cromosomas Y en una población, a la estructura matrimonial de las poblaciones humanas y a los probables efectos selectivos debidos a *hitch-hiking* (un marcador neutro puede ser beneficiado, en la evolución, por la presencia de genes importantes que se transmiten ligados a él).

Hoy se sabe que el cromosoma Y presenta diversos tipos de polimorfismos, cada uno de los cuales están caracterizados por una tasa de mutación propia y por ello adaptados a estudios evolutivos de diversa escala temporal: reordenaciones cromosómicas (deleciones, inserciones, inversiones y duplicaciones), polimorfismos asociados a la presencia de repeticiones en tándem de breves secuencias (mini- y microsatélites) y simples sustituciones nucleotídicas.

8.2 MEDIDAS DE DIVERSIDAD

La **distancia genética** define el grado en que dos poblaciones difieren en sus frecuencias alélicas. 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 [169,170]. El cálculo de las distancias genéticas permite entender las relaciones evolutivas entre poblaciones [171].

El cálculo de esta distancia reproduce una estimación relativa del tiempo transcurrido desde que las poblaciones se diferenciaron. Cuando la distancia es

pequeña se hace referencia a subestructuras de una población donde puede existir flujo genético entre ellas, o podría indicar un completo aislamiento de las mismas pero que se han separado hace poco tiempo, por lo que siempre se necesita de otros datos que ayuden a interpretar los resultados. Dos poblaciones aisladas genéticamente se van diferenciando por procesos de mutación y deriva, cambiando así sus frecuencias alélicas; a mayor tiempo de separación, mayor diferencia entre las frecuencias alélicas [172].

Los **índices de diversidad** 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 [169], Reynolds [173], Slatkin [174] y Cavalli-Sforza [122], la diversidad nucleotídica y el número medio de diferencias por parejas.

La **distribución de las diferencias por parejas** es otra forma habitual de representar la diversidad, también llamada “mismatch distribution” o distribución de las “mean pairwise differences”. Este tipo de representación es adecuado para aquellos datos en los que pueden computarse diferencias discretas entre los diferentes alelos o haplotipos, en ADNmt e Y-SNPs. Las diferencias pueden ser sustituciones de bases, RFLPs o número de repeticiones de microsatélites. En el caso de los haplotipos las secuencias de una muestra determinada se comparan dos a dos y se calcula el número de diferencias. Por lo tanto, si existen n secuencias en la muestra será necesaria la realización de $n(n-1)/2$ de estas comparaciones.

Además de describir la diversidad de una muestra, estas distribuciones permiten realizar inferencias acerca de la historia demográfica de las poblaciones. Se ha demostrado que los episodios de crecimiento demográfico de declive o los cuellos de botella sufridos por una población dejan huellas características en la distribución de las diferencias por parejas.

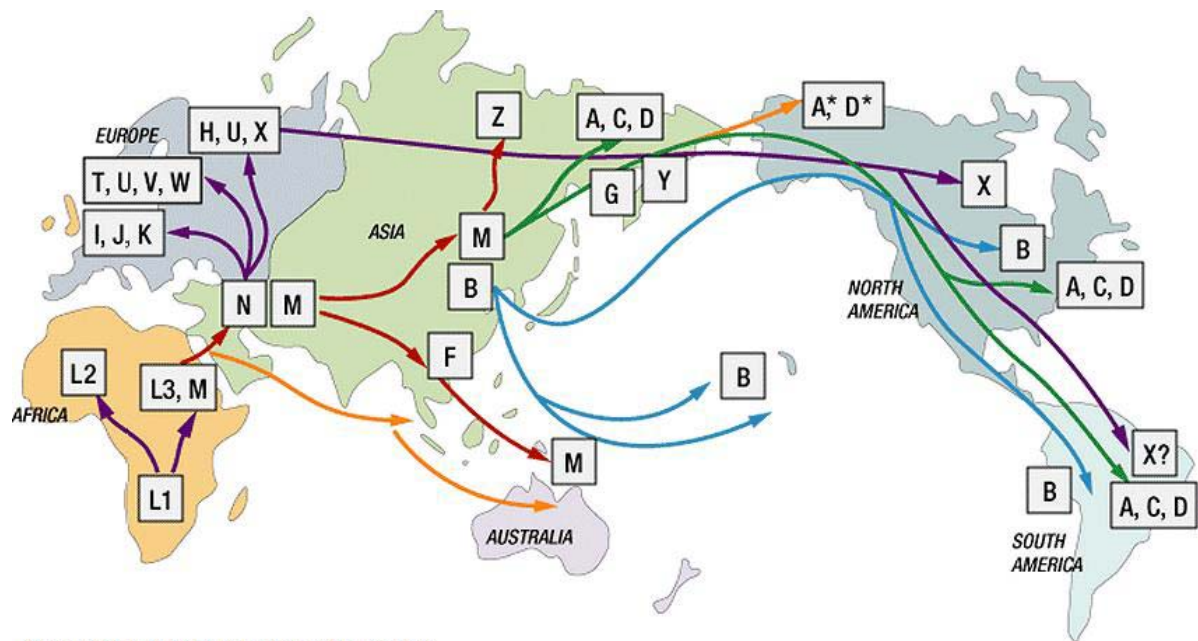
8.3 HAPLOGRUPOS

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 [175]. Los haplogrupos más comúnmente estudiados son los del cromosoma Y y del ADNmt. Ambos pueden ser usados para definir poblaciones genéticas.

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. El incremento de nuevos marcadores antropológicos ha llevado a una disección cada vez más compleja de los diferentes haplogrupos mundiales, reflejada en su nomenclatura.

Los haplogrupos mitocondriales

Son los haplogrupos mejor caracterizados, y más conocidos. El ADN de la mitocondria posee un cierto número de posiciones que definirán el haplogrupo de un individuo. Son cada vez más las posiciones caracterizadas para poder hacer el diagnóstico de un haplogrupo, con lo que cada vez aparecen más subdivisiones.



Expansión de las poblaciones Humanas

En África	120.000-150.000	Europa	35.000-50.000
Salida de África	55.000-75.000	América	15.000-35.000
Asia	40.000-70.000	Na-Dene/ Esquimos/	
Australia	40.000-60.000	Aleutianos	8.000-10.000

Figura I.10. Expansión de las poblaciones humanas siguiendo la distribución de haplogrupos de ADNmt. Mapa traducido y modificado a partir de la imagen original representada por FamilyTreeDNA (2006).

Wallace y Torroni [175] propusieron una nomenclatura que identifica los clusters principales designándolos con una letra mayúscula (Ej.: A, B, C, D, H, I, L, U, V). Los clusters se dividen en sub-clusters que se designan por su letra correspondiente seguida de un número (Ej.: L2). Las subdivisiones al nivel siguiente se realizan alternando letras minúsculas y números (Ej.: U5a1b).

Basándose en la teoría del reloj molecular [176-178] se estableció el haplogrupo mitocondrial más antiguo (entre 150.000 hasta 200.000 años), completado teoría de Eva mitocondrial, que era la primera mujer *Homo sapiens* y había nacido en África.

Este haplogrupo africano ancestro es el haplogrupo L, fue evolucionando, L1, L5, etc., desde el L3 del que partirían los haplogrupos M y N de los que se desarrollarían el resto de los haplogrupos. El haplogrupo N, formaría los A, S, W, X, Y y R, de éste último evolucionaría hacia muchos de los haplogrupos

Europeos (H, J, T, etc.). Del megahaplogrupo M, se originarían los C, D, E, G, Q y Z.

El haplogrupo mayoritario en la población española es el haplogrupo H [108,179]. Éste tiene origen probable en el Medio Oriente, hace unos 30.000 años, migrando a Europa hace 20.000 a 25.000 años [180]. La mayor diversidad se encuentra en el Cáucaso, parte de Europa Oriental y Medio Oriente, el origen más probable estaría en Asia Menor o zonas adyacentes [181].

El haplogrupo más frecuente en los grupos étnicos mesoamericanos es el A [182,183], nacido en Asia y datado entre 30.000 y 50.000 años atrás. Es muy frecuente en norteamericanos. La mayor población con este haplogrupo está en el este de Asia, y posee una gran diversidad en el este de Siberia. Éste fue, casi con toda seguridad, el haplogrupo fundador de las poblaciones del nuevo mundo.

Los haplogrupos en cromosoma Y

Los distintos polimorfismos presentes en el cromosoma Y producen un haplotipo que se agrupan en haplogrupos. Los haplogrupos en cromosoma Y están definidos por distintos tipos de polimorfismos: SNPs, STRs, etc.

La nomenclatura de los haplogrupos de cromosoma Y es muy similar a la utilizada en el ADNmt. Una vez calculadas las frecuencias de los distintos polimorfismos en el cromosoma Y, se fueron estableciendo haplogrupos y se definió el haplogrupo ancestro, Adán. Partiendo de este haplotipo calculado, los haplogrupos más antiguos son el A y B que sólo se encuentran en el África subsahariana, o poblaciones que por migración proceden de allí (migraciones debidas al comercio de esclavos) y de estos derivan el resto de los haplogrupos.

Junto al B, según la filogenia calculada, se encuentra el llamado haplogrupo CT, del que surgirían los haplogrupos D y E, y por otro lado el C y el F.

El haplogrupo D se creó en Asia. Sus representantes actuales se encuentran dispersos en varias zonas de Asia, en especial en el Tíbet, en el archipiélago del Japón y al sudeste de la India.

El haplogrupo E se divide a su vez en tres subgrupos denominados E1, E2 y E3. E1 y E2 se encuentran en África. E3 se divide a su vez en E3a y E3b, éste último tiene frecuencias significativas en Europa, en la mitad este y oeste de Asia y grupos minoritarios de blancos dentro de África. Esto se atribuye a las antiguas migraciones del este de África a través del norte de África en el periodo Neolítico.

El haplogrupo C, como el D, es de origen Asiático y se sitúa en una amplia región desde Siberia hasta Oceanía.

Todo apunta a que el origen del haplogrupo F sea africano. Éste es el haplogrupo del que derivan la mayor parte de los haplogrupos, un 90% de los haplotipos en la actualidad se encuentran englobados en él. Debido a la gran antigüedad de este haplogrupo, se han acumulado numerosas mutaciones y es en el que se ha producido una mayor dispersión.

En un principio, cuando se habían estudiado sólo los Y-STRs, el cálculo de haplogrupos era complicado, pero gracias al estudio continuado de SNPs y STRs se vio que existía una concordancia entre los perfiles de Y-STRs y los haplogrupos en Y [184,185]. Gracias a nuevas herramientas on-line asociadas a una gran base de datos, los perfiles de Y-STRs pueden ser transformados en haplogrupos y se pueden hacer predicciones según todos los datos obtenidos por estudios poblacionales; un ejemplo de éstas serían el 21-haplogroup program y el 23-Haplogroup Beta Program, disponibles en <http://www.hprg.com/hapest5/>.

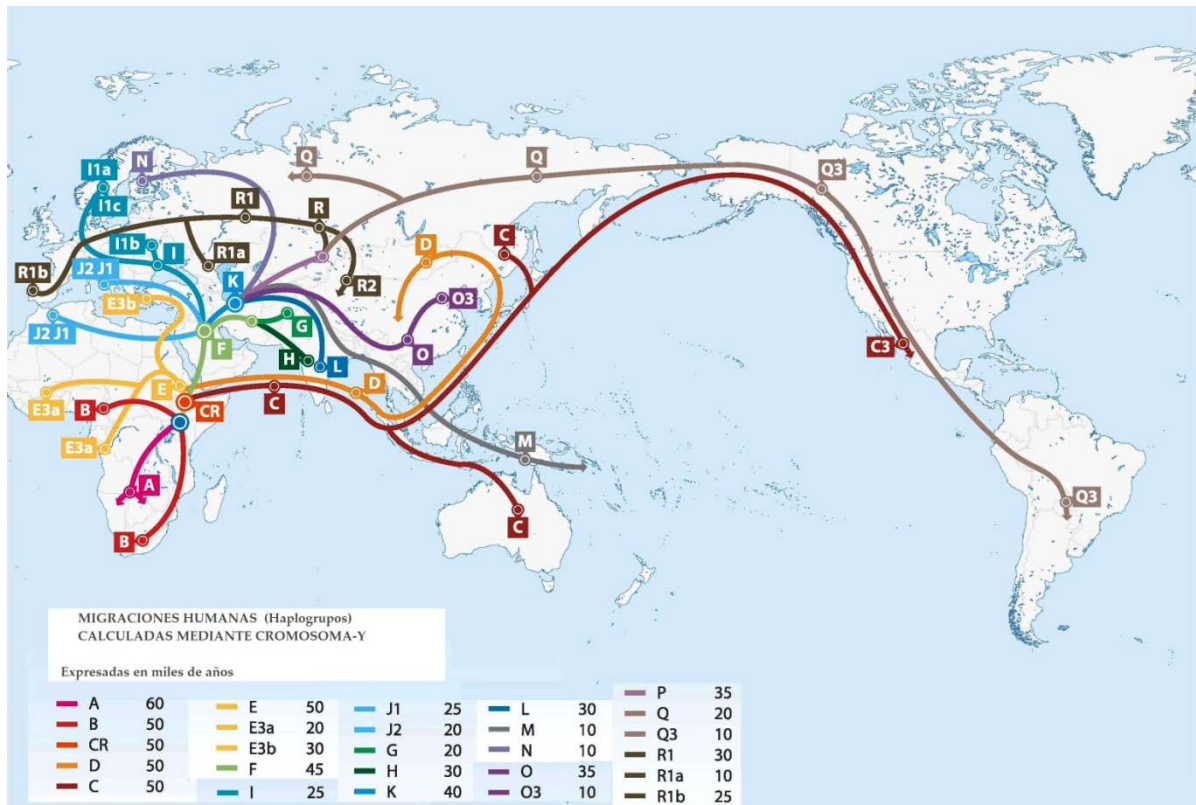


Figura I.11. Expansión de las poblaciones humanas rastreando la distribución de haplogrupos de cromosoma Y. Imagen original representada por FamilyTreeDNA (2006) traducida y modificada para este trabajo.

Destacar que los haplogrupos mayoritarios en Mesoamérica y España son los Haplogrupos Q y R respectivamente:

- Haplogrupo Q: El origen de este haplogrupo parece ser siberiano ya que contiene el ancestro patrilineal de muchos Siberianos, asiáticos e indígenas americanos. Una migración de Asia a Alaska por el Estrecho de Bering llevó este haplogrupo a tierras americanas, donde ahora se encuentra en una proporción muy alta. Los subgrupos en los que se divide son Q1, Q2, Q3 (típico de los indígenas americanos), Q4, Q5, y Q6.
- Haplogrupo R: Su origen se establece en algún lugar del norte de Asia. Se subdivide en R1 y R2, y R1 está formado por R1a y R1b. La mayoría de los miembros del haplogrupo R, pertenecen al haplogrupo R1, muy común en Europa y oeste de Eurasia. El R1 se concentra en el oeste de Europa con una frecuencia del 70%, se encuentra especialmente en Irlanda, España (R1b) y Países Bajos.

8.4 ANÁLISIS DE FILOGENIAS

Toda la información obtenida de los marcadores genéticos en las poblaciones estudiadas se encuentra en las matrices de distancia genética, pero cuando se tienen varios grupos poblacionales, éstas son muy difíciles de interpretar sin análisis adicionales. Las distancias obtenidas se tratan de nuevo mediante análisis de agrupamiento para facilitar la interpretación mediante representaciones gráficas.

Las representaciones más habituales son mediante árboles de distancia (fenéticos y cladísticos), dendogramas o mapas de la matriz de distancias entre poblaciones.

Para calcular los árboles o dendogramas se necesita en primer lugar un software, como el paquete Philip, que trate los datos de distancia genética, que utiliza matrices y algoritmos para agrupar los datos. Los métodos de agrupación suelen ser el UPGMA (unweighted pair group method using arithmetic averages) [186] y el NJ (neighbor-joining) [169]. Primero agrupan parejas más cercanas, posteriormente van añadiendo las siguientes más cercanas a éstas y repitiendo este proceso hasta que todas las poblaciones están incluidas.

El problema que presentan los dendogramas, es que cuando incorporan dos poblaciones, fija este punto como una nueva distancia a las siguientes (estático) y esto fija una raíz. Aunque muchas veces facilita muchísimo la interpretación de los datos puede que el resultado sea erróneo, ya que en la interpretación de las medidas de distancias hay que tener, como ya se ha mencionado antes, otras consideraciones geográficas, culturales, etc.

Existen además de estos, otros métodos para el análisis de las filogenias, como puede ser las herramientas estadísticas que ofrecen los análisis multivariantes como el análisis de correspondencias y el escalamiento multidimensional. Con las gráficas obtenidas con estos métodos se logra que la representación sea más

abierta a una interpretación, y así poder tener en cuenta otros puntos de vista antropológicos, no cerrándose sólo a los datos moleculares.

8.4.1 Análisis Multivariados

Estos métodos permiten el análisis de un conjunto de dos o más variables que pueden estar interrelacionadas si se considera las variables de una forma simultánea. Estas técnicas permiten realizar interpretaciones más complejas que las que surgen mediante la utilización de métodos univariados.

Los análisis multivariados pueden ser lineales como el análisis de componentes principales, análisis de funciones discriminantes, análisis de agrupamientos o clusters, regresión múltiple, análisis multivariado de la varianza, análisis de correspondencias y coordenadas principales. También pueden ser métodos no lineales como análisis factorial, correlación canónica, modelos logarítmicos lineales, escalamiento multidimensional y la regresión logística múltiple. Estos métodos citados son los usados con más frecuencia.

En general los análisis multivariados aprovechan las relaciones entre las variables para buscar patrones o estructuras en los datos. En este sentido, podría encontrarse un origen de patrones cuando se realizan mediciones sobre grupos de objetos similares. Pueden existir casos donde no se conoce a priori si los grupos están ya formados, cuántos son, o que objetos pertenecen a cada grupo; es allí donde habría que determinar qué tipo de análisis es más apropiado.

El **análisis de correspondencia** (AC) intenta explicar una variable hipotética (factor), por medio de un modelo lineal en el que el factor (o varios factores) está en función de un conjunto grande de variables observables (es equivalente al análisis de componentes principales). Es una técnica descriptiva para representar tablas de contingencia. En general, una tabla de contingencia es un conjunto de números positivos dispuestos en una matriz, donde el número de cada casilla representa la frecuencia observada. El AC tiene como objetivo encontrar una estructura más simple reduciendo la dimensionalidad de las

variables sin perder información. Para simplificar el análisis de los datos se reduce el número de variables a un pequeño número de índices o factores.

El AC es un procedimiento de ordenación apropiado para datos de frecuencias (tablas de contingencia). La inercia es la suma ponderada de distancias al centroide calculado, cuanto menor sea, mejor es el ajuste (siempre que el número de variables no varíe). El término inercia se utiliza por su analogía en física. Se define como λ^2 total para una tabla de r filas y c columnas, dividida entre la suma total.

El Análisis de Correspondencia puede condensar la información de un gran número de alelos y loci en pocas variables sintéticas. Con este método las frecuencias alélicas de las poblaciones en todos los loci se usan como variables y el cluster de cada población se representa gráficamente [187].

El **escalamiento multidimensional** (MDS) es una técnica multivariante de interdependencia que trata de representar en un espacio geométrico de pocas dimensiones las proximidades existentes entre un conjunto de objetos. Este análisis opera sobre una matriz de medidas de similitud o disimilitud. Está diseñado para encontrar, en cada dimensión especificada, la mejor representación geométrica de los datos en el espacio. El nombre de "escalamiento multidimensional" realmente no es una técnica, sino un conjunto de ellas que persiguen como objetivo la representación de datos a través de una configuración de puntos cuando se conoce una determinada información sobre similitudes (o proximidades) entre objetos. Desde las primeras técnicas, simplex o Torgeson (que fue el primero en trabajar con ellas) hasta los últimos análisis más complejos que obtienen una representación euclídea, intentan constituir un gráfico que respete lo más posible las relaciones de proximidad entre los objetos [188]. Siempre la representación pretende ajustarse lo más posible a las proximidades de partida y mediante un estadístico conveniente se mide la bondad de ajuste entre las distancias de la configuración y las disparidades. Existen diferentes definiciones de este estadístico, pero la mayoría surge de la definición del llamado índice de esfuerzo (en inglés: STRESS). En

este momento estos análisis son los que más se están desarrollando y constituyen en los estudios de genética molecular y a la antropología molecular la mejor herramienta para representar las distancias genéticas calculadas. [108]

8.4.2 Networks

En un principio los análisis de agrupación básicos para poder hacer árboles, bastaban para estudiar el origen o los antecesores de una población, pero a medida que se fueron acumulando multitud de datos, aumentando el número de polimorfismos estudiados tanto en STRs como en ADNmt y el número de muestras la reconstrucción filogenética era cada vez más ardua, menos efectiva. Por ello fue necesario el desarrollo de nuevas herramientas específicas para el análisis de poblaciones [145,147,189]. El Método de las redes medias (Median Networks), permite explorar el espacio de los árboles más parsimoniosos generados a partir de un conjunto de datos y resumirlos, formando una única red. Finalmente se obtiene una serie de puntos, haplotipos (que variarán su tamaño dependiendo del número de individuos), unidos por redes. El primer punto será el ancestro común y a partir de éste se obtendrá el resto. El software para poder realizarlo, Network 4.510, se distribuye de forma gratuita. Disponible en <http://www.fluxus-engineering.com>.

9. ANTROPOLOGÍA MOLECULAR EN EL NUEVO MUNDO

En la última década se han realizado una enorme cantidad de estudios sobre las Américas dentro de los distintos campos de la Antropología. Muchos laboratorios de antropología, de todo el mundo, se han interesado por determinar la procedencia del primer pueblo americano.

Estaba bastante claro que los primeros indígenas americanos no proceden de Europa como se pensó en un principio, sino que procedían de Asia o Eurasia, pero hay dudas a la hora de determinar el momento en que se formó la pasarela helada que permitió que los pueblos asiáticos entraran en el continente. Tampoco se puede asegurar que grupos asiáticos serían los más próximos a los nuevos colonos del continente americano, y si eran uno o varios grupos.

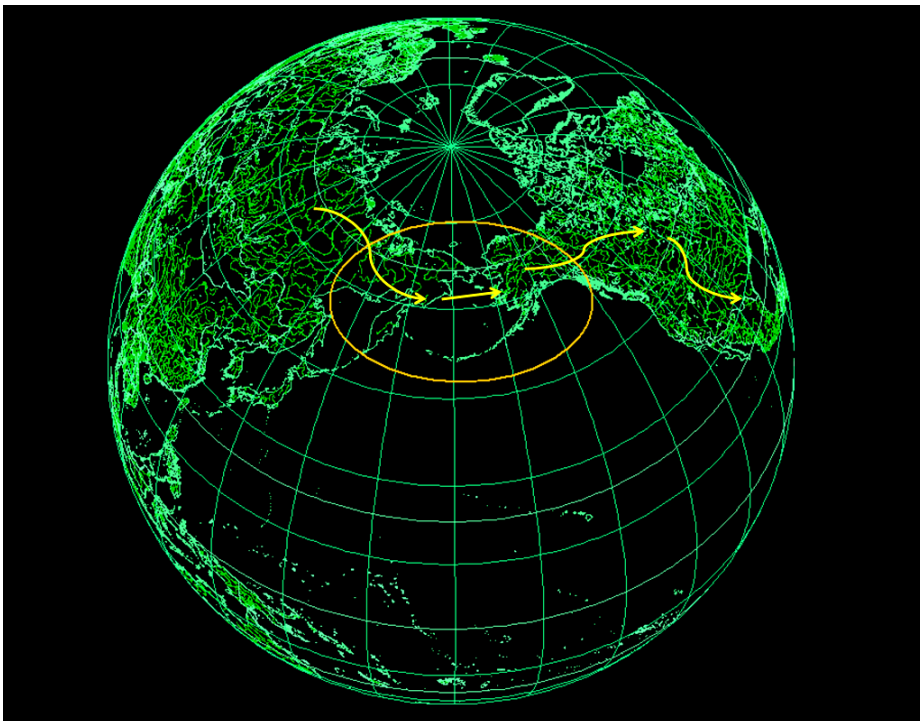


Figura I.12. La entrada de los estos primeros habitantes del nuevo mundo cruzarían desde el extremo oriental de Siberia, hasta el más occidental de Alaska, por el estrecho de Bering. Poco a poco, por migración, fueron colonizando la zona norte (Canadá y EEUU) posteriormente zona central del continente americano y por último el sur, alcanzando en no mucho tiempo la tierra de fuego.

9.1 HISTORIA DE AMÉRICA DEFINIDA POR MARCADORES GENÉTICOS

Los primeros estudios realizados por Wallace en 1990 [52], tras ver los haplogrupos obtenidos en las poblaciones siberianas y en las nativas americanas, confirman por genética molecular que las migraciones que originaron a los primeros amerindios ocurrieron entre el 20.000 a.c. y el 14.000 a.c. Ésta sería una primera migración que produciría la entrada de los primeros pobladores. Estos estudios, seguidos de otros [190-195] basados en ADNmt, confirman que estos primeros habitantes del nuevo mundo cruzarían desde el extremo oriental de Siberia, hasta el más occidental de Alaska, por el estrecho de Bering. Poco a poco por migración fueron colonizando la zona norte (actual Canadá y EEUU) y posteriormente zona central del continente Americano y por último el Sur alcanzando en no mucho tiempo la tierra de fuego. Esta teoría se apoyaba en el espíritu cazador de estas poblaciones nómadas que perseguían la caza de grandes animales. Las primeras migraciones originarían los pueblos que muchos años después formarían las grandes culturas e imperios del clásico y post clásico americano, pueblos de centro América, dando como fruto los andinos, mayas y caribes.

Posteriormente una segunda migración, que casi con toda seguridad proviene desde la misma zona de Siberia que la primera, amplía con más repobladores del continente asiático la población americana. Estos nuevos indígenas no colonizan de nuevo todo el continente, quedándose su mayoría en el norte de América y algunos llegaron a la zona central.

Finalmente, una última migración de las poblaciones cercanas al estrecho de Bering poblaría las Islas Aleutianas y el norte de América, dando como fruto los Aleutianos, Esquimales y los Indios Na-Dené.

Estudios basados en el cromosoma Y confirmaron los datos obtenidos con el ADNmt [196][116] poniendo de manifiesto los dos grandes descensos del tamaño poblacional americano, que han dificultado su estudio. Un primer

descenso estaría cercano a la segunda migración y otro con la llegada de los europeos al continente americano.

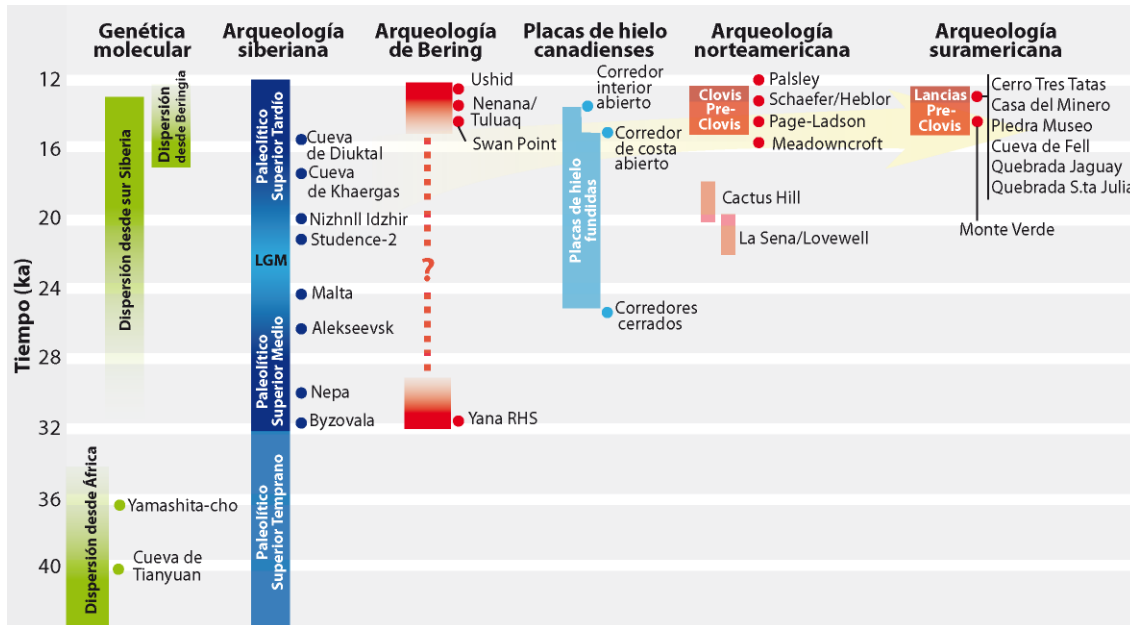


Figura I 13. Datos arqueológicos, geológicos y genéticos revelan el avance de las primeras poblaciones humanas en el continente Americano. Los marcadores genéticos estudiados fueron SNPs y STRs de cromosoma Y y ADNmt. Imagen traducida y modificada a partir de la original representada por Goebel, Science (2008).

Recientemente, combinando los estudios arqueológicos con estudios basados en cromosoma Y, ADNmt y también mediante el estudio de aDNA en muestras antiguas [197] se confirma que una única dispersión, desde una sola población de Siberia que migró hasta el estrecho de Bering (posiblemente hace 22.000 años) y hace 16.500 años los primeros colonizadores se asentaron en el norte de América estableciéndose hace 15.000 años. Tanto los datos del ADNmt como los del cromosoma Y emergen de la misma región central de Asia, por tanto hay un único pool genético ancestro, que comparten todos los pueblos americanos. Este estudio realizado por Goebel afirma que los estudios filogenéticos no pueden discriminar entre una migración simple y múltiples migraciones, apoyado por recientes estudios que han descrito distintos haplogrupos en diferentes poblaciones siberianas.

9.2 MESTIZAJE

El Mestizaje es el fruto del intercambio biológico y cultural de etnias con distinto origen, que debido a su mezcla, origina etnias nuevas. En varios lugares como Filipinas, Sudáfrica o Estados Unidos hay pueblos que han atravesado un proceso de encuentro entre varias etnias. Pero este término describe perfectamente proceso transcurrido en Iberoamérica dando como resultado el estado racial y cultural actual.

Desde los encuentros de los primeros pueblos, el mestizaje se hace presente en la historia del hombre. Gracias a los nuevos medios de transporte, de comunicación, nuevas tecnologías, etc., este proceso es cada vez más rápido. La globalización en los procesos de intercambio internacional y una organización mundial más compleja e interrelacionada hace del hombre moderno un hombre sin prejuicios y esto facilita el intercambio cultural y biológico de las etnias.

El **mestizaje en Iberoamérica** se origina cuando los europeos llegan al continente americano, y simultáneamente los esclavos africanos que vinieron con ellos. Esto produjo la aparición de los mestizos (indígena y europeo), mulatos (africano y europeo), zambo (africano e indígena) y castizo (mestizo y europeo). Estos nuevos grupos étnicos tenían claramente unas grandes diferencias con sus predecesores, no sólo desde el punto de vista biológico, sino también desde el mundo de vista social y cultural. De hecho, el mestizaje ha sido uno de los temas fundamentales en el continente americano, y en especial en América Latina.

Las sociedades mestizas en Latinoamérica son la fusión de tres grupos fundamentales, el amerindio, el europeo y el africano, grupos poblacionales parentales. La proporción de cada una de estos grupos va a variar mucho, tanto en aporte, como en origen, según la zona del continente americano.

Teniendo en cuenta las diferencias entre esos tres grandes grupos se puede encontrar dentro de lo que se denominaba *raza* amerindia muchos grupos poblacionales (Oto-Mangue, Maya, Yuta-Nawan, Hokan, Na-dené, etc.). Del

mismo modo dentro de los que se llamaban caucásicos se encuentra diferencias entre celtas, íberos, germanos, flamencos, sajones, etc. Y por último dentro de las poblaciones Subsaharianas se pueden encontrar también distintos grupos y culturas.

La llegada al continente por parte de los europeos así como por parte de los africanos fue distinta dependiendo de la zona y dependiendo de si eran tierras más o menos ricas y más o menos pobladas. Esto hace del continente americano un crisol de culturas originarias que se refleja en la riqueza de los grupos mestizos.

Para saber cómo se ha formado y cuál es el origen y la estructura actual de una comunidad mestiza dispone de distintas herramientas que permiten ver como se ha formado esta nueva comunidad a partir de las poblaciones parentales: datos históricos, culturales o la antropología clásica. Una nueva herramienta es la antropología molecular que permite observar qué tanto por ciento de caracteres genéticos [198,199] comparten las distintas poblaciones y como es actualmente la estructura de esta nueva población mestiza [200].

Tabla I.3. Aproximación al mestizaje en América latina (1954). El número de individuos en estas últimas seis décadas ha crecido, pero los porcentajes actuales son fruto de la estructura de la población en el pasado. Datos tomados de (Ro-senblat A. *La población indí-gena y el mestizaje en América*. Edi-to-rial No-va, 1954.).

	Total		Latinos		Indodescendientes		Afrodescendientes		Mulatos y Negros	
	Miles	%	Miles	%	Miles	%	Miles	%	Miles	%
Paises indomestizos (Guatemala, Ecuador Perú y Bolivia)	58022	100	29836	51,4	47734	82,3	3942	6,8	3920	6,8
Guatemala	11382	100	5237	46	10816	95	22	0,2	0	0
Paises mestizos (Mexico, El Salvador, Honduras, Nicaragua, Paraguay)	122134	100	106045	86,8	103754	85	1037	0,8	795	0,7
P. Afromestizos (Panamá, Colombia, Venezuela)	69131	100	63942	92,5	34065	49,3	21629	31,3	21187	30,6
P. Afrocriollos (Cuba, Puerto Rico, R. Domi- nicana, Brasil)	193893	100	178984	92,3	682	0,4	92233	47,6	92041	47,5

El número de mestizos dentro de la población latinoamericana, es cada vez más alto, y el de indígenas americanos es más bajo o incluso han desaparecido de muchas regiones. Los porcentajes de población mestiza en muchos países es muy alto: Honduras (85.6%), Nicaragua (78.3%), Paraguay (74.5%), México (70.0%) o El Salvador (91%). Pero en otros, gracias a sus características orográficas, a que la mayoría de su población es rural, etc., se pueden encontrar muchos grupos indígenas y la población mestiza oscila cercana al 50%, como es el caso de Guatemala, Ecuador y Bolivia.

En Guatemala, las poblaciones parentales mayoritarias son la maya y la ibérica. En la época colonial, solamente eran españoles e indígenas los que se encontraban en el reino de Guatemala. Hasta 1815 Guatemala siguió siendo dependiente de España siendo, por tanto, la que mayor flujo poblacional europeo aportara a la comunidad mestiza. Aunque también se encuentran de forma muy poco numerosa italianos, alemanes e ingleses. Como en otros países la población de origen africano en Guatemala sería la siguiente contribuidora, pero en mucha menor proporción comparándola con las dos grandes parentales, indígena y española. Como en el resto de iberoamericana, se introdujeron a partir del siglo XVI, pero en el caso de Guatemala, por su orografía y sus habitantes, el número de esclavos fue algo menor ya que no había grandes latifundios ni zonas costeras deshabitadas, donde en general en el resto de los países la influencia africana se ve mucho más acentuada.

La población mestiza lleva muchos años luchando por su identidad y estudiando la riqueza de las poblaciones mestizas como el fruto de la unión de las distintas culturas que la forman. En estas últimas décadas también las poblaciones mayas comienzan a luchar por su identidad, buscando sus costumbres, sus tradiciones, conservando sus lenguas, etc. Lo que sigue siendo un hecho hoy en día en Guatemala es las diferencias geográficas, lingüísticas, económicas, sociales, culturales, de acceso a la educación, etc., que existen entre los mestizos o ladinos y la población indígena o maya. Esto hace que estas dos identidades permanezcan en un equilibrio constante.

En la comunidad de Guatemala, la población mestiza es actualmente una población equilibrada [201] por lo que es difícil asignar en este tipo de población la procedencia de sus componentes. Pero se puede comparar ésta con las poblaciones parentales para intentar asignar cuanto comparten cada población.

Se conocen herramientas para **asignar individuos dentro de una población mestiza** [200,202-208].

Existen dos métodos para asignar individuos a poblaciones:

- Por su distancia genética: Se calcula la matriz de distancia entre individuos y se representa (al igual que se observa en el análisis de filogenias). Son más fáciles de aplicar, aunque tienen desventajas, ya que los distintos métodos difieren mucho unos de otros, pero sus resultados son fáciles de interpretar y dan un cálculo directo de la cercanía de un individuo o población de individuos a su población parental en tanto por ciento [199,209,210].
- Modelos probabilísticos: Trabajan bajo los supuestos de que las frecuencias alélicas se encuentran en equilibrio Hardy-Weinberg y que no existe desequilibrio de ligamiento. Pueden ser métodos basados en frecuencias o métodos Bayesianos.

Los métodos Bayesianos, actualmente más utilizados, se basan en determinar si unas partes del genoma (clusters) son heredadas con mayor tasa desde una población parental. Este método asigna individuos a poblaciones estimando las frecuencias alélicas de cada locus.

Pritchard y col. establecieron este nuevo método para identificar diferentes poblaciones [200]; y tras esto se estudia la ascendencia de los individuos. Se parte de dos modelos para calcular la ascendencia de los individuos, un modelo no-combinado, donde se asume que los individuos son tomados de forma pura de una de las k poblaciones y otro, el modelo combinado, donde se permite mezcla de los ancestros; es decir, una fracción q_k del genoma de un individuo viene de la subpoblación K ($\sum_k q_k = 1$). Estos modelos están disponibles en el

programa Structure v2.3.1 disponible en: <http://pritch.bsd.uchicago.edu/>. El requerimiento de este programa es que las frecuencias en la población debe encontrarse en equilibrio de Hardy-Weinberg y no puede haber ligamiento entre los loci, para ello se requiere un muestreo adecuado [208].

10. LA POBLACIÓN DE GUATEMALA

En la zona norte de Centroamérica se encuentra Guatemala, un país con características peculiares y una cultura autóctona muy fuerte influenciada principalmente por su pasado Maya.

Guatemala cuenta con un territorio montañoso, con gran cantidad de bosques. Limita al oeste y norte con México, al este con Belice y el Golfo de Honduras, al sureste con Honduras y El Salvador, y al sur con el océano Pacífico. El país tiene 108.889 km². Su capital es Nueva Guatemala, llamada Ciudad de Guatemala o *Guatemala City* a nivel internacional. Su población indígena compone un tercio de la población del país, y la población mestiza supera la mitad de la población. Su idioma oficial es el español, asimismo cuenta con idiomas mayas característicos de cada grupo indígena.

Étnicamente es una nación mayoritariamente amerindia maya. El 40,5% guatemalteca es indígena de origen maya pertenecientes a diferentes ramas (K'iche', Kaqchiqueles, Mames, etc.) aunque hay algunos estudios que afirman superar el 60%. Aproximadamente un 50% son mestizos. También se encuentran en un pequeño porcentaje descendientes de los colonizadores españoles y de otras naciones europeas, que aún hoy, se denominan criollos (≈9%). De esta manera se ubica, según UNICEF, entre las naciones latinoamericanas con poblaciones de origen mayoritariamente indígena tales como Bolivia, o Ecuador.

10.1 GEOGRAFÍA DE GUATEMALA

Guatemala es el país más poblado de América Central, con una población estimada en aproximadamente 13 millones de habitantes. La sociedad guatemalteca está, desde la conquista española, dividida en dos grupos: por una parte los indios de origen maya, que constituyen la mitad de la población de Guatemala, de una extrema pobreza y viven principalmente en las altas

tierras poco fértiles; por otra parte, los "ladinos", mestizos y descendientes de los españoles, esencialmente urbanos.

La geografía física de este país ha facilitado la existencia de los distintos grupos poblacionales. Guatemala es uno de los países más volcánicos del mundo. Aproximadamente dos terceras partes del territorio de Guatemala está formado por montañas, muchas de ellas de origen volcánico. Volcanes como el de Santa María con 3.772 metros sobre el nivel del mar o el pico más alto de América Central, el Tajumulco con 4.220, contrastan con las dos grandes mesetas que separa la Sierra Madre. La cadena montañosa de Sierra Madre cruza del este a oeste todo el territorio Guatemalteco. Estas zonas altas del centro así como el altiplano Guatemalteco se oponen a las planicies costeras de vertiente Pacífico y las tierras del Petén, meseta calcárea que recubre el norte del país, donde se encuentra una selva tropical muy densa.

10.2 DISTRIBUCIÓN DE LA POBLACIÓN

Políticamente Guatemala está dividida en 22 departamentos administrativos. La ciudad más grande del país es Ciudad Guatemala y su población (los guatemaltecos) sobrepasa los 2.300.000 habitantes. Las otras ciudades más importantes son: Quezaltenango (96.000 habitantes), Puerto Barrios (40.000 habitantes), Mazatenango y Antigua. La población ladina se encuentra mayoritariamente en estos grandes núcleos poblacionales.

La población Kaqchiquel se encuentra en la región centro occidental de Guatemala. Se distribuyen en una zona central en Chimaltenango, alrededor de la Ciudad de Guatemala y de San Juan Sacatepéquez, al sur de Antigua y en San Martín Jilotepeque.

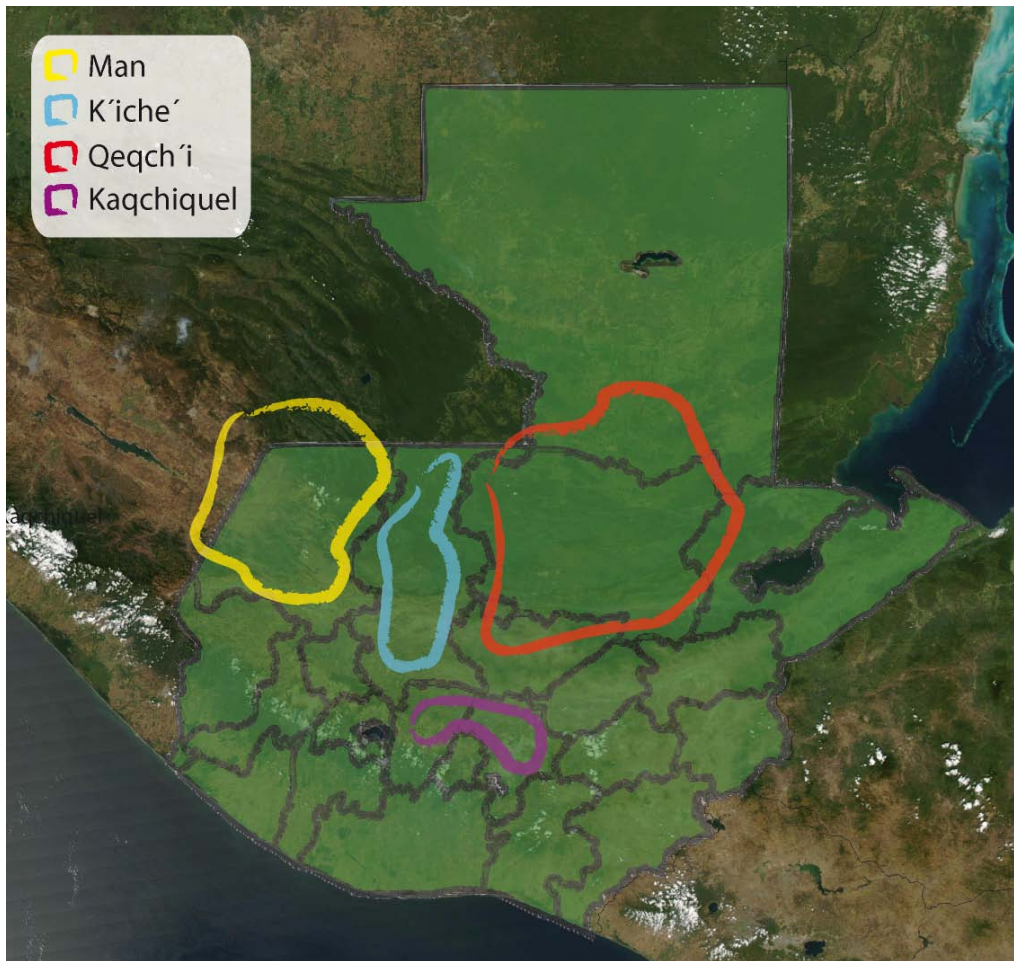


Figura I.14. Mapa de Guatemala y distribución de los cuatro grupos poblacionales estudiados.

El grupo K'iche', se encuentra situado en el altiplano guatemalteco. La mayoría de ellos se encuentran en el departamento que lleva su mismo nombre. Santa Cruz del Quiché (traducción de K'iche' por los españoles), es la población más numerosa de esta etnia, se encuentra cercana a las ruinas de antigua capital maya, Gumarcaaj. El departamento de Quiché fue nombrado así en alusión a este pueblo que en su inicio ocupó el territorio denominado "Quix Ché" que significa "árboles con espinas" pues en la región abundaron los magueyes, nopales, ortigas, etc. En tiempos recientes, los K'iche's de han dispersado sobre un área más amplia del territorio guatemalteco.

Los Q'eqchi', es uno de los grupos poblacionales que ocupa mayor extensión geográfica, se encuentran dispersos en zonas poco habitadas. Se sitúa principalmente en los departamentos de Alta Verapaz, Baja Verapaz, Petén, e Izabal.

En el noreste de Guatemala se encuentran los Mames aunque algunos de ellos ocupan el sureste de México.

10.3 HISTORIA, CULTURA Y LENGUA

Los indígenas son los descendientes originarios del continente americano. Se les suele agrupar en “familias” por lo general dependiendo del origen de su lengua, por su ubicación geográfica, rasgos culturales y por una historia común. Éste es el caso de los pueblos andinos, mayas y caribes.

Los pueblos indígenas de América han sufrido un gran descenso demográfico, siendo reducidos solamente a áreas rurales, debido fundamentalmente a la discriminación y racismo por ladinos y criollos. Históricamente se han descrito matanzas, la existencia de un trabajo servil, etc., pero hasta la actualidad estos grupos están sumergidos en condiciones de pobreza.

La lengua indígena, como dialecto, era un prejuicio que considera las lenguas mayenses, por ejemplo, formas de comunicarse menos complejas o desarrolladas que el español, el inglés o el francés, aunque las lenguas indígenas americanas son tan complejas y ricas como cualquier otra lengua.

En este momento hay más de 28 lenguas distintas de origen maya que se han generado, como se decía anteriormente, por el aislamiento que los pueblos indígenas han sufrido en las zonas rurales y de montaña [211-213].

Los descendientes de los mayas están, en este momento, distribuidos por distintos territorios del sur de Méjico, Guatemala, Belice, Honduras y El Salvador.

Inicialmente se pensaba que el pueblo Maya era el pueblo descendiente de los Olmec, pero en este momento la mayoría de las investigaciones afirman que son civilizaciones distintas que convivieron y entre las que hubo intercambio de conocimientos [214]. Lo que sí es cierto que el idioma Olmec está íntimamente relacionado con los lenguajes Proto-mayas [215].

Los primeros restos de la cultura maya aparecen en el periodo preclásico, alrededor del 1800 a.c. en el mirador de Basin en el norte de Guatemala y se calcula que el primer asentamiento en esta zona se establecería en el 6000 a.c. Ya a mediados de la época preclásica, en el 1000- 400 a.c., se pueden encontrar signos de la aparición de una élite social, de un gobierno, y la superioridad de los mayas frente a sus rivales.

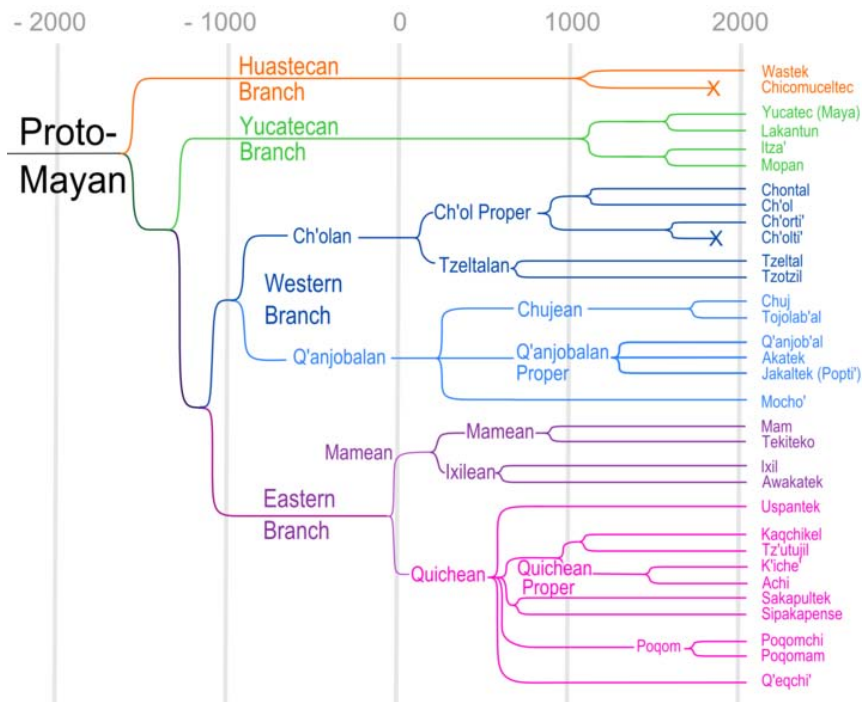


Figura I.15. Genealogía de la lengua maya. Las primeras lenguas en formarse, son las del norte de Méjico: Huasteca y posteriormente las Yucatán. A continuación aparecerían las ramas del este. De la del este surgirán las lenguas que actualmente usan poblaciones del centro y sur de Méjico. Y finalmente la oriental, donde aparecen las lenguas que actualmente hablan los pueblos mayas guatemaltecos. El pueblo Q'eqchi', el Kaqchikel y el K'iche' comparten una misma rama de la que se habría separado previamente los Mam.

Durante el periodo clásico, cuando surgió el gran desarrollo de la cultura maya, aparecen las grandes construcciones, el estudio de las ciencias (matemáticas, astronomía, etc.). Pero al final de este periodo debido a los cambios climáticos, a desastres naturales, conflictos en su gobierno, sobreexplotación de su entorno, etc., hubo un gran declive de esta cultura y un gran descenso de la población [214].

En el periodo post clásico, 900-1.500 d.c., comienza a disgregarse el imperio maya apareciendo los estados independientes, donde destacan el K'iche' y Kaqchikel. Estos nuevos estados estuvieron en continuo conflicto con las civilizaciones Mejicanas de los Toltec y los Aztecas, que establecieron un dominio de toda Mesoamérica, lo que provocó la reducción de estos estados hasta la llegada de los Españoles [214].

10.4 PUEBLOS MAYA Y LOS POLIMORFISMOS DEL ADN

Uno de los pueblos mesoamericanos más interesantes desde el punto de vista antropológico es el pueblo maya. La aparición de restos de arquitectura, de escritura, del conocimiento de las matemáticas y la astrología [211-213] así como su supremacía en toda la llanura mesoamericana durante tantos siglos hace de esta cultura una de las más interesantes del continente americano.

Este interés se ha reproducido en distintos estudios, aunque no son muchos los grupos poblacionales mesoamericanos que han sido estudiados hasta la fecha con marcadores genéticos. Además la heterogeneidad de los marcadores estudiados hace difícil la comparación entre los distintos estudios [216-220]. La mayoría de los estudios se han centrado en norte y centro América [194]. Estos estudios se centran en la comparación con poblaciones europeas, africanas y asiáticas con marcadores HLA [216,219,221-223], también estudios con ADNmt [224-226] y cromosoma Y [227,228] o con inserciones Alu [218,229,230] y en todos ellos destaca la comparación de sus poblaciones, no con poblaciones vecinas, sino con poblaciones de diversos lugares del mundo [194].

Destacar un estudio muy interesante sobre las poblaciones mayas y las relaciones entre sus pueblos, de Lisa Ibarra- Ribera publicado en 2008. En este estudio, realizado con 15 marcadores de tipo STR, se comparan cuatro poblaciones maya, K'iche' y Kaqchiquel de Guatemala y dos Mejicanas, Campeche y Yucatán. Estas cuatro poblaciones son comparadas con 22 poblaciones. Este estudio revela la cercanía de los pueblos mesoamericanos entre sí y con sus poblaciones mestizas (utilizando estos marcadores autosómicos STR) y sus diferencias con poblaciones más alejadas, como las europeas africanas o las asiáticas. También se ve como los grupos étnicos cercanos entre sí (K'iche', Kaqchiquel, Campeche, Yucatan y Choles) divergen menos de los grupos mestizos mesoamericanos cercanos que de otros grupos étnicos mesoamericanos, aunque hayan sido durante siglos poblaciones vecinas, e incluso puedan tener orígenes comunes, existe una mayor divergencia. Es importante destacar la divergencia con otros grupos como los

Otomani o los Huastecas Mejicanos. Se puede observar igualmente como en estos grupos étnicos disminuye la variabilidad e incluso a veces no se respeta el equilibrio Hardy Weinberg (HWE) en algunos loci. Es interesante tener como antecedente esta descripción de los pueblos mayas antes de revisar los resultados de nuestro estudio en los siguientes capítulos. Ibarra-Ribera y sus colaboradores ponen de manifiesto la necesidad de seguir estudiando estos pueblos por el interés que despiertan los resultados obtenidos, por ver si se reproducen estos resultados en otros pueblos mayas, y más aun, con estos y otros marcadores genéticos.

11. BIBLIOGRAFÍA

- [1] A.R. Wyman, R. White, A highly polymorphic locus in human DNA, *Proc.Natl.Acad.Sci.U.S.A.* 77 (1980) 6754-6758.
- [2] K.B. Mullis, Target amplification for DNA analysis by the polymease chain reaction, *Ann.Biol.Clin.(Paris)*. 48 (1990) 579-582.
- [3] 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.
- [4] M.F. Singer, Highly Repeated Sequences in Mammalian Genomes, in: G.H. Bourne and J.F. Danielli (Ed.), *International Review of Cytology*, Academic Press, 1982, pp. 67-112.
- [5] Y. Nakamura, M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, et al., Variable number of tandem repeat (VNTR) markers for human gene mapping, *Science*. 235 (1987) 1616-1622.
- [6] R.I. Richards, G.R. Sutherland, Heritable unstable DNA sequences, *Nat.Genet.* 1 (1992) 7-9.
- [7] A. Edwards, H.A. Hammond, L. Jin, C.T. Caskey, R. Chakraborty, Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups, *Genomics*. 12 (1992) 241-253.
- [8] J.S. Beckmann, J.L. Weber, Survey of human and rat microsatellites, *Genomics*. 12 (1992) 627-631.
- [9] 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.
- [10] M. Litt, J.A. Luty, A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene, *Am.J.Hum.Genet.* 44 (1989) 397-401.
- [11] J. Weber, May P., Abundant class of human DNA polymorphism which can be typed using then polymerase chain reaction. *Am.J.Hum.Genet.* 44 (1989) 338-396.
- [12] B. Brinkmann, Population genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci, *European Journal of Human Genetics*. 4 (1996) 175-182.

- [13] J. Czarny, T. Grzybowski, M.V. Derenko, B.A. Malyarchuk, D.M. Śliwka, Genetic variation of 15 STR loci (D3S1358, vWA, FGA, TH01, TPOX, CSF1PO, D5S818, D13S317, D7S820, D16S539, D2S1338, D8S1179, D21S11, D18S51, and D19S433) in populations of north and central Poland, *Forensic Sci.Int.* 147 (2005) 97-100.
- [14] N. Ellis, P.N. Goodfellow, The mammalian pseudoautosomal region, *Trends in Genetics.* 5 (1989) 406-410.
- [15] N.A. Ellis, P.J. Goodfellow, B. Pym, M. Smith, M. Palmer, A.-. Frischauf, et al., The pseudoautosomal boundary in man is defined by an Alu repeat sequence inserted on the Y chromosome, *Nature.* 337 (1989) 81-84.
- [16] T.J. Hassold, S.L. Sherman, D. Pettay, D.C. Page, P.A. Jacobs, XY chromosome nondisjunction in man is associated with diminished recombination in the pseudoautosomal region, *Am.J.Hum.Genet.* 49 (1991) 253-260.
- [17] D. Freije, C. Helms, M.S. Watson, H. Donis-Keller, Identification of a second pseudoautosomal region near the Xq and Yq telomeres, *Science.* 258 (1992) 1784-1787.
- [18] S.S. Wachtel, G.C. Koo, E.A. Boyse, Evolutionary conservation of H-Y ('male') antigen, *Nature.* 254 (1975) 270-272.
- [19] D.C. Page, R. Mosher, E.M. Simpson, E.M.C. Fisher, G. Mardon, J. Pollack, et al., The sex-determining region of the human Y chromosome encodes a finger protein, *Cell.* 51 (1987) 1091-1104.
- [20] A.H. Sinclair, P. Berta, M.S. Palmer, J.R. Hawkins, B.L. Griffiths, M.J. Smith, et al., A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif, *Nature.* 346 (1990) 240-244.
- [21] E.M.C. Fisher, P. Beer-Romero, L.G. Brown, A. Ridley, J.A. McNeil, J.B. Lawrence, et al., Homologous ribosomal protein genes on the human X and Y chromosomes: Escape from X inactivation and possible implications for Turner syndrome, *Cell.* 63 (1990) 1205-1218.
- [22] Y. Nakagome, S. Nagafuchi, S. Seki, Y. Nakahori, T. Tamura, M. Yamada, et al., A repeating unit of the DYZ1 family on the human Y chromosome consists of segments with partial male-specificity, *Cytogenet.Cell Genet.* 56 (1991) 74-77.
- [23] P. De Knijff, Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome, *Am.J.Hum.Genet.* 67 (2000) 1055-1061.
- [24] J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, *J.Forensic Sci.* 48 (2003) 1054-1064.

- [25] 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.
- [26] M. Casanova, P. Leroy, C. Boucekkine, A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance, *Science.* 230 (1985) 1403-1406.
- [27] G. Lucotte, N.Y. Ngo, p49f, A highly polymorphic probe, that detects Taq1 RFLPs on the human Y chromosome. *Nucleic Acids Res.* 13 (1985) 8285.
- [28] L. Roewer, M. Nagy, P. Schmidt, J.T. Epplen, G. Herzog-Schröder, Microsatellite and HLA class II oligonucleotide typing in a population of Yanomami Indians. *EXS.* 67 (1993) 221-230.
- [29] F.R. Santos, S.D.J. Pena, J.T. Epplen, Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique, *Hum.Genet.* 90 (1993) 655-656.
- [30] R. Kittler, A. Erler, S. Brauer, M. Stoneking, M. Kayser, Apparent intrachromosomal exchange on the human Y chromosome explained by population history, *European Journal of Human Genetics.* 11 (2003) 304-314.
- [31] A.J. Redd, A.B. Agellon, V.A. Kearney, V.A. Contreras, T. Karafet, H. Park, et al., Forensic value of 14 novel STRs on the human Y chromosome, *Forensic Sci.Int.* 130 (2002) 97-111.
- [32] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: An evolutionary marker comes of age, *Nature Reviews Genetics.* 4 (2003) 598-612.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] R. Lessig, S. Willuweit, M. Krawczak, F.-. Wu, C.-. Pu, W. Kim, et al., Asian online Y-STR haplotype reference database, *Leg.Med.* 5 (2003) S160-S163.

- [37] 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.
- [38] 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.
- [39] 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.
- [40] C. Krausz, L. Quintana-Murci, G. Forti, Y chromosome polymorphisms in medicine, *Ann.Med.* 36 (2004) 573-583.
- [41] S. Lindström, H.O. Adami, J. Adolfsson, F. Wiklund, Y chromosome haplotypes and prostate cancer in Sweden, *Clinical Cancer Research.* 14 (2008) 6712-6716.
- [42] T. Strachan, A.P. Read, Instability of the human genome: Mutation and DNA repair, *Human Molecular Genetics* 2nd Edition. (1999) 209-217.
- [43] D.G. Monckton, R. Neumann, T. Guram, N. Fretwell, K. Tamaki, A. MacLeod, et al., Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism, *Nat.Genet.* 8 (1994) 162-170.
- [44] M. Goedbloed, M. Vermeulen, R.N. Fang, M. Lembring, A. Wollstein, K. Ballantyne, et al., Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpF ISTR® Yfiler® PCR amplification kit, *Int.J.Legal Med.* (2009) 1-12.
- [45] L. Gusmão, C. Alves, P. Sánchez-Diz, M.T. Zarrabeitia, M.A. Abovich, I. Aragón, et al., Results of the GEP-ISFG collaborative study on an X-STR Decaplex, *Forensic Science International: Genetics Supplement Series.* 1 (2008) 677-679.
- [46] L. Gusmão, P. Sánchez-Diz, C. Alves, I. Gomes, M.T. Zarrabeitia, M. Abovich, et al., A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: Data on 15 Iberian and Latin American populations, *Int.J.Legal Med.* 123 (2009) 227-234.
- [47] P. Sánchez-Diz, C. Alves, E. Carvalho, M. Carvalho, R. Espinheira, O. García, et al., Population and segregation data on 17 Y-STRs: Results of a GEP-ISFG collaborative study, *Int.J.Legal Med.* 122 (2008) 529-533.
- [48] American Association of Blood Banks, ANNUAL REPORT SUMMARY FOR TESTING IN 2003, (2004).
- [49] S. Anderson, A.T. Bankier, B.G. Barrell, Sequence and organization of the human mitochondrial genome, *Nature.* 290 (1981) 457-465.

- [50] O. Handt, S. Meyer, A. Von Haeseler, Compilation of human mtDNA control region sequences, *Nucleic Acids Res.* 26 (1998) 126-129.
- [51] 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.
- [52] D.C. Wallace, A. Torroni, American Indian prehistory as written in the mitochondrial DNA: a review. *Human Biology.* 64 (1992) 403-416.
- [53] D.C. Wallace, Mitochondrial genetics: A paradigm for aging and degenerative diseases? *Science.* 256 (1992) 628-632.
- [54] A. Torroni, L. D'Urbano, C. Rengo, R. Scozzari, M. Sbracia, C. Manna, et al., Intracytoplasmic injection of spermatozoa does not appear to alter the mode of mitochondrial DNA inheritance [1], *Human Reproduction.* 13 (1998) 1747-1749.
- [55] D.A. Clayton, Structure and function of the mitochondrial genome, *J.Inherit.Metab.Dis.* 15 (1992) 439-447.
- [56] R.C. Shuster, A.J. Rubenstein, D.C. Wallace, Mitochondrial DNA in anucleate human blood cells, *Biochem.Biophys.Res.Commun.* 155 (1988) 1360-1365.
- [57] 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.
- [58] L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A.C. Wilson, African populations and the evolution of human mitochondrial DNA, *Science.* 253 (1991) 1503-1507.
- [59] B. Budowle, M.R. Wilson, J.A. DiZinno, C. Stauffer, M.A. Fasano, M.M. Holland, et al., Mitochondrial DNA regions HVI and HVII population data, *Forensic Sci.Int.* 103 (1999) 23-35.
- [60] D.C. Wallace, M.D. Brown, M.T. Lott, Mitochondrial DNA variation in human evolution and disease, *Gene.* 238 (1999) 211-230.
- [61] D.C. Wallace, M.D. Brown, S. Melov, B. Graham, M. Lott, Mitochondrial biology, degenerative diseases and aging, *Biofactors.* 7 (1998) 187-190.
- [62] P.F. Chinnery, N. Howell, R.M. Andrews, D.M. Turnbull, Mitochondrial DNA analysis: Polymorphisms and pathogenicity, *J.Med.Genet.* 36 (1999) 505-510.
- [63] P.F. Chinnery, D.M. Turnbull, Mitochondrial DNA and disease, *Lancet.* 354 (1999) 17-21.

- [64] P.F. Chinnery, D.R. Thorburn, D.C. Samuels, S.L. White, H.-M. Dahl, D.M. Turnbull, et al., The inheritance of mitochondrial DNA heteroplasmy: Random drift, selection or both? *Trends in Genetics*. 16 (2000) 500-505.
- [65] T.J. Parsons, J.A. Irwin, Questioning evidence for recombination in human mitochondrial DNA, (2000).
- [66] P.L. Ivanov, M.J. Wadhams, R.K. Roby, M.M. Holland, V.M. Weedn, T.J. Parsons, Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II, *Nat.Genet*. 12 (1996) 339-420.
- [67] E.E. Jazin, L. Cavelier, I. Eriksson, L. Orelund, U. Gyllensten, Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA, *Proc.Natl.Acad.Sci.U.S.A.* 93 (1996) 12382-12387.
- [68] R.J. Monnat Jr., D.T. Reay, Nucleotide sequence identity of mitochondrial DNA from different human tissues, *Gene*. 43 (1986) 205-211.
- [69] D. Comas, S. Paabo, J. Bertranpetit, Heteroplasmy in the control region of human mitochondrial DNA, *Genome Res*. 5 (1995) 89-90.
- [70] K.E. Bendall, V.A. Macaulay, J.R. Baker, B.C. Sykes, Heteroplasmic point mutations in the human mtDNA control region, *Am.J.Hum.Genet*. 59 (1996) 1276-1287.
- [71] C.D. Calloway, R.L. Reynolds, G.L. Herrin Jr., W.W. Anderson, The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age, *Am.J.Hum.Genet*. 66 (2000) 1384-1397.
- [72] 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.
- [73] L.A. Tully, T.J. Parsons, R.J. Steighner, M.M. Holland, M.A. Marino, V.L. Prenger, A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region, *Am.J.Hum.Genet*. 67 (2000) 432-443.
- [74] L.A. Tully, B.C. Levin, Human mitochondrial genetics, *Biotechnology and Genetic Engineering Reviews*. 17 (2000) 147-177.
- [75] K.E. Bendall, B.C. Sykes, Length heteroplasmy in the first hypervariable segment of the human mtDNA control region, *Am.J.Hum.Genet*. 57 (1995) 248-256.
- [76] D.R. Marchington, G.M. Hartshorne, D. Barlow, J. Poulton, Homopolymeric tract heteroplasmy in mtDNA from tissues and single oocytes: Support for a genetic bottleneck, *Am.J.Hum.Genet*. 60 (1997) 408-416.

[77] X. Chen, R. Prasser, S. Simonetti, J. Sadlock, G. Jagiello, E.A. Schon, Rearranged mitochondrial genomes are present in human oocytes, *Am J Hum. Genet.* 57 (1995) 238-247.

[78] M. Lagerström-Fermér, C. Olsson, L. Forsgren, A.-. Syvänen, Heteroplasmy of the human mtDNA control region remains constant during life, *Am.J.Hum.Genet.* 68 (2001) 1299-1301.

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

[80] W.W. Hauswirth, P.J. Laipis, Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. *Proc.Natl.Acad.Sci.U.S.A.* 79 (1982) 4686-4690.

[81] M.V. Ashley, P.J. Laipis, W.W. Hauswirth, Rapid segregation of heteroplasmic bovine mitochondria, *Nucleic Acids Res.* 17 (1989) 7325-7331.

[82] R.L. Cann, A.C. Wilson, Length mutations in human mitochondrial DNA, *Genetics.* 104 (1983) 699-711.

[83] 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.

[84] 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.

[85] 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.

[86] R.L. Cann, A.C. Stoneking, A.C. Wilson, Mitochondrial DNA and human evolution, *Nature.* 356 (1987) 389-390.

[87] 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.

[88] 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.

[89] 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.

[90] Mully J., Gedeon A., White S., Haan E., Richards R., Predictive diagnosis of myotonic dystrophy with flanking microsatellite markers, *J.Med.Genet.* 28 (1991) 448-452.

- [91] N. Morral, J. Bertranpetit, X. Estivill, The origin of the major cystic fibrosis mutation (delta F508) in European populations, *Nat.Genet.* 7 (2004) 169-175.
- [92] C. Oudet, R. Heilig, A. Hanauer, J.-. Mandel, Nonradioactive assay for new microsatellite polymorphisms at the 5' end of the dystrophin gene, and estimation of intragenic recombination, *Am.J.Hum.Genet.* 49 (1991) 311-319.
- [93] B. Weber, O. Riess, G. Wolff, S. Andrew, C. Collins, R. Graham, et al., Delineation of a 50 kilobase DNA segment containing the recombination site in a sporadic case of Huntington's disease. *Nat.Genet.* 2 (1992) 216-222.
- [94] S.J. Odelberg, D.B. Demers, E.H. Westing, A.A. Hossaini, Establishing paternity using minisatellite DNA probes when the putative father is unavailable for testing, *J.Forensic Sci.* 33 (1988) 921-928.
- [95] I. Ishiyama, T. Yoshii, M. Honma, M. Mukaida, T. Yamaguchi, DNA fingerprints: The importance in forensic medicine. II. The significance of examining DNA polymorphisms of placental tissues for the purpose of paternity determination during the early stages within the first trimester, *Zeitschrift fur Rechtsmedizin.* 99 (1988) 241-248.
- [96] I.W. Evett, C. Buffery, G. Willott, A guide to interpreting single locus profiles of DNA mixtures in forensic cases, *Journal of the Forensic Science Society.* 31 (1991) 41-47.
- [97] Applied Biosystems, GeneMapper® ID-X Software Mixture Analysis Tool, v1.1 (2008).
- [98] P. Gill, D.J. Werrett, Exclusion of a man charged with murder by DNA fingerprinting, *Forensic Sci.Int.* 35 (1987) 145-148.
- [99] 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.
- [100] 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.
- [101] M. Kayser, A. Caglia, 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.
- [102] 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.

- [103] C. Pestoni, M.L. Cal, M.V. Lareu, M.S. Rodríguez-Calvo, A. Carracedo, Y chromosome STR haplotypes: Genetic and sequencing data of the Galician population (NW Spain), *Int.J.Legal Med.* 112 (1998) 15-21.
- [104] Q. Ayub, A. Mohyuddin, R. Qamar, K. Mazhar, T. Zerjal, S.Q. Mehdi, et al., Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res.* 28 (2000).
- [105] S.L. White, V.R. Collins, R. Wolfe, M.A. Cleary, S. Shanske, S. DiMauro, et al., Genetic counseling and prenatal diagnosis for the mitochondrial DNA mutations at nucleotide 8993, *Am.J.Hum.Genet.* 65 (1999) 474-482.
- [106] M.A. Jobling, C. Tyler-Smith, Fathers and sons: The Y chromosome and human evolution, *Trends in Genetics.* 6 (1995) 799-803.
- [107] M.A. Jobling, C. Tyler-Smith, New users for new haplotypes the human Y chromosome, disease and selection, *Trends in Genetics.* 16 (2000) 356-362.
- [108] 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, *The American Journal of Human Genetics.* 83 (2008) 725-736.
- [109] M.A. Jobling, In the name of the father: Surnames and genetics, *Trends in Genetics.* 17 (2001) 353-357.
- [110] T.E. King, M.A. Jobling, What's in a name? Y chromosomes, surnames and the genetic genealogy revolution, *Trends in Genetics.* 25 (2009) 351-360.
- [111] B. McEvoy, D.G. Bradley, Y-chromosomes and the extent of patrilineal ancestry in Irish surnames, *Hum.Genet.* 119 (2006) 212-219.
- [112] U.-. Immel, M. Krawczak, J. Udolph, A. Richter, H. Rodig, M. Kleiber, et al., Y-chromosomal STR haplotype analysis reveals surname-associated strata in the East-German population, *European Journal of Human Genetics.* 14 (2006) 577-582.
- [113] E. Martínez-Espin, Análisis y comparación de 16 loci STR en cromosoma Y de varios grupos poblacionales (apellidados Colón y no apellidados Colón) en la cuenca noroeste mediterranea, (2008).
- [114] 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 Epidemiology Biomarkers and Prevention.* 13 (2004) 681-687.
- [115] A.A. Romero, M.G. Jamison, A.C. Weidner, Are single nucleotide polymorphisms associated with pelvic organ prolapse? *Journal of Pelvic Medicine and Surgery.* 14 (2008) 37-43.

- [116] W. Kim, T.-. Yoo, S.-. Kim, D.-. Shin, C. Tyler-Smith, H.-. Jin, et al., Lack of association between Y-chromosomal haplogroups and prostate cancer in the Korean population, *PLoS ONE*. 2 (2007).
- [117] S. Paabo, Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification, *Proc.Natl.Acad.Sci.U.S.A.* 86 (1989) 1939-1943.
- [118] S. Pääbo, Amplifying DNA from archeological remains: A meeting report, *Genome Res.* 1 (1991) 107-110.
- [119] S. Paabo, D.M. Irwin, A.C. Wilson, DNA damage promotes jumping between templates during enzymatic amplification, *J.Biol.Chem.* 265 (1990) 4718-4721.
- [120] S. Paabo, D.M. Irwin, A.C. Wilson, DNA damage promotes jumping between templates during enzymatic amplification, *J.Biol.Chem.* 265 (1990) 4718-4721.
- [121] 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), *Science and Justice - Journal of the Forensic Science Society.* 44 (2004) 51-53.
- [122] L.L. Cavalli-Sforza, M.W. Feldman, The application of molecular genetic approaches to the study of human diversity, *Nat.Genet.* 33 (2003) 267.
- [123] S. Horai, K. Hayasaka, R. Kondo, K. Tsugane, N. Takahata, Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs, *Proc.Natl.Acad.Sci.U.S.A.* 92 (1995) 532-536.
- [124] S. Horai, Evolution and the origins of man: clues from complete sequences of hominoid mitochondrial DNA. *Southeast Asian J.Trop.Med.Public Health.* 26 Suppl 1 (1995) 146-154.
- [125] P. Gagneux, C. Wills, U. Gerloff, D. Tautz, P.A. Morin, C. Boesch, et al., Mitochondrial sequences show diverse evolutionary histories of African hominoids, *Proc.Natl.Acad.Sci.U.S.A.* 96 (1999) 5077-5082.
- [126] R. Noda, C.G. Kim, O. Takenaka, R.E. Ferrell, T. Tanoue, I. Hayasaka, et al., Mitochondrial 16S rRNA sequence diversity of hominoids, *J.Hered.* 92 (2001) 490-496.
- [127] L. Vigilant, B.J. Bradley, Genetic variation in gorillas, *Am.J.Primatol.* 64 (2004) 161-172.
- [128] K.B. Bulayeva, T.A. Pavlova, R.M. Kurbanov, S. Leal, O.A. Bulayev, Genetic and epidemiological studies in daghestan highland isolates, *Genetika.* 39 (2003) 413-422.

- [129] K. Bulayeva, L.B. Jorde, C. Ostler, S. Watkins, O. Bulayev, H. Harpending, Genetics and Population History of Caucasus Populations, *Human Biology*. 75 (2003) 837-853.
- [130] M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, Mitochondrial genome variation and the origin of modern humans, *Nature*. 408 (2000) 708-713.
- [131] L.B. Jorde, M.J. Bamshad, W.S. Watkins, R. Zenger, A.E. Fraley, P.A. Krakowiak, et al., Origins and affinities of modern humans: A comparison of mitochondrial and nuclear genetic data, *Am.J.Hum.Genet.* 57 (1995) 523-538.
- [132] N. Maca-Meyer, A.M. González, J.M. Larruga, C. Flores, V.M. Cabrera, Major genomic mitochondrial lineages delineate early human expansions. *BMC genetics* [electronic resource]. 2 (2001) 13.
- [133] M. Krings, A. Stone, R.W. Schmitz, H. Krainitzki, M. Stoneking, S. Pääbo, Neandertal DNA sequences and the origin of modern humans, *Cell*. 90 (1997) 19-30.
- [134] L. Quintana-Murci, R. Veitia, S. Santachiara-Benerecetti, K. McElreavey, M. Fellous, T. Bourgeron, Mitochondrial DNA, Y chromosome and human population history, *Medecine/Sciences*. 15 (1999) 974-982.
- [135] G. Gutiérrez, D. Sánchez, A. Marín, A reanalysis of the ancient mitochondrial DNA sequences recovered from neandertal bones, *Mol.Biol.Evol.* 19 (2002) 1359-1366.
- [136] D. Caramelli, C. Lalueza-Fox, C. Vernesi, M. Lari, A. Casoli, F. Mallegni, et al., Evidence for a genetic discontinuity between neandertals and 24,000-year-old anatomically modern europeans, *Proc.Natl.Acad.Sci.U.S.A.* 100 (2003) 6593-6597.
- [137] D. Serre, A. Langaney, M. Chech, M. Teschler-Nicola, M. Paunovic, P. Menecier, et al., No evidence of Neandertal mtDNA contribution to early modern humans, *PLoS Biology*. 2 (2004).
- [138] M. Currat, L. Excoffier, Modern humans did not admix with Neanderthals during their range expansion into Europe, *PLoS Biology*. 2 (2004).
- [139] A. Cooper, A.J. Drummond, E. Willerslev, Ancient DNA: Would the real neandertal please stand up? *Current Biology*. 14 (2004) R431-R433.
- [140] M.B. Richards, V.A. Macaulay, H.J. Bandelt, B.C. Sykes, Phylogeography of mitochondrial DNA in Western Europe, *Ann.Hum.Genet.* 61 (1998) 251-254.
- [141] 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.

- [142] S. Finnilä, I.E. Hassinen, K. Majamaa, Phylogenetic analysis of mitochondrial DNA in patients with an occipital stroke: Evaluation of mutations by using sequence data on the entire coding region, *Mutation Research - Mutation Research Genomics*. 458 (2001) 31-39.
- [143] 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.
- [144] Q.-. Kong, Y.-. Yao, C. Sun, H.-. Bandelt, C.-. Zhu, Y.-. Zhang, Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences, *Am.J.Hum.Genet.* 73 (2003) 671-676.
- [145] H.-. Bandelt, P. Forster, B.C. Sykes, M.B. Richards, Mitochondrial portraits of human populations using median networks, *Genetics*. 141 (1995) 743-753.
- [146] H.-. Bandelt, V. Chepoi, Decomposition and l1-Embedding of Weakly Median Graphs, *European Journal of Combinatorics*. 21 (2000) 701-714.
- [147] H.-. Bandelt, A. Henkman, F. Nicolai, Powers of distance-hereditary graphs, *Discrete Mathematics*. 145 (1995) 37-60.
- [148] H.-. 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.
- [149] N. Maca-Meyer, A.M. González, J. Pestano, C. Flores, J.M. Larruga, V.M. Cabrera, Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography, *BMC Genetics*. 4 (2003).
- [150] V. Macaulay, M. Richards, E. Hickey, E. Vega, F. Cruciani, V. Guida, et al., The emerging tree of west Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs, *Am.J.Hum.Genet.* 64 (1999) 232-249.
- [151] M. Richards, V. Macaulay, H.-. Bandelt, Analyzing genetic data in a model-based framework: Inferences about European prehistory, *Examining the Farming/Language Dispersal Hypothesis*. (2002) 459-466.
- [152] C. Entrala, J.A. Lorente, M. Lorente, J.C. Alvarez, B. Budowle, E. Villanueva, Spanish population data on the loci D13S317, D7S820, and D16S539 generated using silver staining (SilverSTR III(TM) Multiplex), *J.Forensic Sci.* 44 (1999) 1032-1034.
- [153] J.C. Alvarez, Polimorfismos de ADN mitocondrial en la población española, (2003).

[154] C. Entrala, Análisis del polimorfismo de diez de loci tipo microsatélite en la población andaluza para su aplicación en identificación genética humana, (2000).

[155] J.A. Lorente, J.C. Alvarez, C. Entrala, E. Martinez-Espin, F. Fernandez-Rosado, M. Lorente, et al., Bases de datos de ADN: Su uso en la investigación criminal y en la identificación civil. Anotaciones prácticas para su desarrollo, *Forénsica*. 1 (2002) 31-44.

[156] Boletín Oficial del Estado, LEY ORGÁNICA 10/2007, de 8 de octubre, reguladora de la base de datos policial sobre identificadores obtenidos a partir del ADN. 17634 (2007) 40969-40972.

[157] J.A. Lorente, C. Entrala, J.C. Alvarez, M. Lorente, E. Villanueva, F. Carrasco, et al., Missing persons identification: Genetics at work for society [2], *Science*. 290 (2000) 2257-2258.

[158] J.A. Lorente, C. Entrala, J.C. Alvarez, M. Lorente, B. Arce, B. Heinrich, et al., Social benefits of non-criminal genetic databases: Missing persons and human remains identification, *Int.J.Legal Med*. 116 (2002) 187-190.

[159] C. Capelli, F. Tschentscher, V.L. Pascali, "Ancient" protocols for the crime scene?: Similarities and differences between forensic genetics and ancient DNA analysis, *Forensic Sci.Int*. 131 (2003) 59-64.

[160] P. Gill, J. Whitaker, C. Flaxman, N. Brown, J. Buckleton, An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Sci.Int*. 112 (2000) 17-40.

[161] P.L. Ivanov, Identification of the remains of the romanov family: contribution of molecular genetics, *Vestnik Rossijkoj Akademii Nauk*. 66 (1996) 310-316.

[162] P. Gill, P.L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, et al., Identification of the remains of the Romanov family by DNA analysis, *Nat.Genet*. 6 (1994) 130-135.

[163] P.L. Ivanov, Identification of the remains of the romanov family: contribution of molecular genetics, *Vestnik Rossijkoj Akademii Nauk*. 66 (1996) 310-316.

[164] U. Wolf, W. Schempp, G. Scherer, Molecular biology of the human Y chromosome. *Reviews of Physiology Biochemistry and Pharmacology*. 121 (1992) 147-213.

[165] O. Rickards, C. Martinez-Labarga, G. Scano, G.F. De Stefano, G. Biondi, M. Pacaci, et al., Genetic history of the population of Sicily, *Human Biology*. 70 (1998) 699-714.

[166] K.Y. Ngo, G. Lucotte, Strategies for detecting restriction polymorphisms of Y chromosome sequences, *Ann.Genet*. 29 (1986) 88-92.

- [167] S. Jakubiczka, J. Arnemann, H.J. Cooke, M. Krawczak, J. Schmidtke, A search for restriction fragment length polymorphism on the human Y chromosome, *Hum.Genet.* 84 (1989) 86-88.
- [168] K.D. Makova, W.-. Li, Strong male-driven evolution of DNA sequences in humans and apes, *Nature.* 416 (2002) 624-626.
- [169] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* 4 (1987) 406-425.
- [170] M. Nei, Genetic Distance, in: Sydney Brenner, Jeffrey H. Miller (Eds.), *Encyclopedia of Genetics*, Academic Press, New York, 2001, pp. 828-832.
- [171] Y. Nagamine, M. Higuchi, Genetic distance and classification of domestic animals using genetic markers, *J.Anim.Breed.Genet.* 118 (2001) 101-109.
- [172] J. Felsenstein, 'Likelihood calculations on coalescents', *Inferring Phylogenies.* (2004) 470-487.
- [173] 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.
- [174] M. Slatkin, Linkage disequilibrium in growing and stable populations, *Genetics.* 137 (1994) 331-336.
- [175] A. Torroni, T.G. Schurr, 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.
- [176] E. Zuckerkandl, L. Pauling, Molecules as documents of evolutionary history, *J.Theor.Biol.* 8 (1965) 357-366.
- [177] A.C. Wilson, V.M. Sarich, A molecular time scale for human evolution. *Proc.Natl.Acad.Sci.U.S.A.* 63 (1969) 1088-1093.
- [178] M. Kimura, Evolutionary rate at the molecular level, *Nature.* 217 (1968) 624-626.
- [179] 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.
- [180] 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.

- [181] P. Soares, L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Röhl, et al., Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock, *Am.J.Hum.Genet.* 84 (2009) 740-759.
- [182] A. Torroni, R.I. Sukernik, T.G. Schurr, Y.B. Starikovskaya, M.F. Cabell, M.H. Crawford, et al., mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans, *Am.J.Hum.Genet.* 53 (1993) 591-608.
- [183] 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.
- [184] T.W. Athey, Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach, *Journal of Genetic Genealogy.* 1 (2005) 1-7.
- [185] T.W. Athey, Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach, *Journal of Genetic Genealogy.* 2 (2006) 34-39.
- [186] P.H.A. Sneath, R.R. Sokal, *Numerical Taxonomy.* (1973).
- [187] M.H. Li, K. Sternbauer, P.T. Haahr, J. Kantanen, Genetic components in contemporary Faroe Islands cattle as revealed by microsatellite analysis, *J.Anim.Breed.Genet.* 122 (2005) 309-317.
- [188] T.F. Cox, G. Ferry, Discriminant analysis using non-metric multidimensional scaling, *Pattern Recognit.* 26 (1993) 145-153.
- [189] D. Penny, M. Steel, P.J. Waddell, M.D. Hendy, Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*, *Mol.Biol.Evol.* 12 (1995) 863-882.
- [190] F. Rothhammer, C. Silva, S.M. Callegari-Jacques, E. Llop, F.M. Salzano, Gradients of HLA diversity in South American Indians, *Ann.Hum.Biol.* 24 (1997) 197-208.
- [191] F.M. Salzano, Molecular variability in Amerindians: Widespread but uneven information, *An.Acad.Bras.Cienc.* 74 (2002) 223-263.
- [192] T.G. Schurr, The peopling of the New World: Perspectives from molecular anthropology, *Annual Review of Anthropology.* 33 (2004) 551-583.
- [193] T.G. Schurr, S.T. Sherry, Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: Evolutionary and demographic evidence, *Am.J.Hum.Biol.* 16 (2004) 420-439.
- [194] F.M. Salzano, Molecular variability in Amerindians: Widespread but uneven information, *An.Acad.Bras.Cienc.* 74 (2002) 223-263.

- [195] M.H. Crawford, The origins of Native Americans, Evidence from Anthropological Genetics. (1998).
- [196] C.J. Mulligan, K. Hunley, S. Cole, J.C. Long, Population genetics, history, and health patterns in Native Americans, Annual Review of Genomics and Human Genetics. 5 (2004) 295-315.
- [197] T. Goebel, M.R. Waters, D.H. O'Rourke, The Late Pleistocene dispersal of modern humans in the Americas, Science. 319 (2008) 1497-1502.
- [198] R. Chakraborty, Estimation of race admixture. A new method, Am.J.Phys.Anthropol. 42 (1975) 507-511.
- [199] J.C. Long, The genetic structure of admixed populations, Genetics. 127 (1991) 417-428.
- [200] J.K. Pritchard, M. Stephens, N.A. Rosenberg, P. Donnelly, Association mapping in structured populations, Am.J.Hum.Genet. 67 (2000) 170-181.
- [201] E. Martinez-Espin, L.J. Martínez-Gonzalez, F. Fernandez-Rosado, C. Entrala, J.C. Alvarez, J.A. Lorente, et al., Guatemala mestizo population data on 15 STR loci (Identifiler® kit), J.Forensic Sci. 51 (2006) 1216-1218.
- [202] D. Paetkau, W. Calvert, I. Stirling, C. Strobeck, Microsatellite analysis of population structure in Canadian polar bears, Mol.Ecol. 4 (1995) 347-354.
- [203] D. Paetkau, R. Slade, M. Burden, A. Estoup, Genetic assignment methods for the direct, real-time estimation of migration rate: A simulation-based exploration of accuracy and power, Mol.Ecol. 13 (2004) 55-65.
- [204] D. Paetkau, The optimal number of markers in genetic capture-mark-recapture studies, J.Wildl.Manage. 68 (2004) 449-452.
- [205] D. Paetkau, C. Strobeck, The molecular basis and evolutionary history of a microsatellite null allele in bears, Mol.Ecol. 4 (1995) 519-520.
- [206] B. Rannala, J.L. Mountain, Detecting immigration by using multilocus genotypes, Proc.Natl.Acad.Sci.U.S.A. 94 (1997) 9197-9201.
- [207] J.-. Cornuet, S. Piry, G. Luikart, A. Estoup, M. Solignac, New methods employing multilocus genotypes to select or exclude populations as origins of individuals, Genetics. 153 (1999) 1989-2000.
- [208] 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.

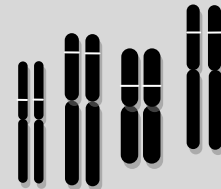
- [209] R. Chakraborty, L. Jin, Determination of relatedness between individuals using DNA fingerprinting. *Human Biology*. 65 (1993) 875-895.
- [210] R. Chakraborty, L. Jin, A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. *EXS*. 67 (1993) 153-175.
- [211] M.D. Coe, *The Maya*. (1999).
- [212] M.D. Coe, *The Maya (Ancient Peoples and Places)*. 6th Ed. (1999).
- [213] M.D. Coe, *Breaking the Maya Code*. (1999) 13.
- [214] R.J. Sharer, A.K. Balkansky, J.H. Burton, G.M. Feinman, K.V. Flannery, D.C. Grove, et al., On the logic of archaeological inference: Early Formative pottery and the evolution of Mesoamerican societies, *Latin American Antiquity*. 17 (2006) 90-103.
- [215] M.D. Coe, R. Koontz, *Mexico: From the Olmecs to the Aztecs*, 6th ed., London: Thames & Hudson, London, UK, 2002.
- [216] E. Gómez-Casado, J. Martínez-Laso, J. Moscoso, J. Zamora, M. Martín-Villa, M. Pérez-Blas, et al., Origin of Mayans according to HLA genes and the uniqueness of Amerindians, *Tissue Antigens*. 61 (2003) 425-436.
- [217] C. Sánchez, C. Barrot, M. Ortega, A. González-Martín, A. Gorostiza, J. Corbella, et al., Genetic diversity of 15 STRs in Choles from northeast of Chiapas (Mexico), *J. Forensic Sci*. 50 (2005) 1499-1501.
- [218] R.J. Herrera, D.P. Rojas, M.C. Terreros, Polymorphic Alu insertions among Mayan populations, *J. Hum. Genet*. 52 (2007) 129-142.
- [219] A. Arnaiz-Villena, G. Vargas-Alarcón, J. Granados, E. Gómez-Casado, J. Longas, M. Gonzalez-Hevilla, et al., HLA genes in Mexican Mazatecans, the peopling of the Americas and the uniqueness of Amerindians, *Tissue Antigens*. 56 (2000) 405-416.
- [220] C. Barrot, C. Sánchez, M. Ortega, A. González-Martín, C. Brand-Casadevall, A. Gorostiza, et al., Characterisation of three Amerindian populations from Hidalgo State (Mexico) by 15 STR-PCR polymorphisms, *Int. J. Legal Med*. 119 (2005) 111-115.
- [221] F.L. Black, Reasons for failure of genetic classifications of South Amerind populations. *Human Biology*. 63 (1991) 763-774.
- [222] M.A. Fernández-Viña, A.M. Lázaro, C.Y. Marcos, C. Nulf, E. Raimondi, E.J. Haas, et al., Dissimilar evolution of B-locus versus A-locus and class II loci of the HLA region in South American Indian tribes, *Tissue Antigens*. 50 (1997) 233-250.

- [223] G. Vargas-Alarcón, G. Hernández-Pacheco, J. Moscoso, N. Pérez-Hernández, L.E. Murguía, A. Moreno, et al., HLA genes in Mexican Teeneks: HLA genetic relationship with other worldwide populations, *Mol.Immunol.* 43 (2006) 790-799.
- [224] S.L. Bonatto, F.M. Salzano, Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World, *Am.J.Hum.Genet.* 61 (1997) 1413-1423.
- [225] S.L. Bonatto, F.M. Salzano, A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data, *Proc.Natl.Acad.Sci.U.S.A.* 94 (1997) 1866-1871.
- [226] W.A. Silva Jr., S.L. Bonatto, A.J. Holanda, A.K. Ribeiro-dos-Santos, B.M. Paixão, G.H. Goldman, et al., Mitochondrial genome diversity of native americans supports a single early entry of founder populations into America, *Am.J.Hum.Genet.* 71 (2002) 187-192.
- [227] N.O. Bianchi, G. Bailliet, C.M. Bravi, R.F. Carnese, F. Rothhammer, V. Martínez-Marignac, et al., Origin of amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers, *Am.J.Phys.Anthropol.* 102 (1997) 79-89.
- [228] F.R. Santos, L. Rodriguez-Delfin, S.D.J. Pena, J. Moore, K.M. Weiss, North and South Amerindians may have the same major founder Y chromosome haplotype [1], *Am.J.Hum.Genet.* 58 (1996) 1369-1370.
- [229] G.E. Novick, C.C. Novick, J. Yunis, E. Yunis, P. Antunez De Mayolo, W.D. Scheer, et al., Polymorphic Alu insertions and the Asian origin of native American populations, *Human Biology.* 70 (1998) 23-39.
- [230] L.H. Mateus Pereira, A. Socorro, I. Fernandez, M. Masleh, D. Vidal, N.O. Bianchi, et al., Phylogenetic information in polymorphic L1 and Alu insertions from East Asians and Native American populations, *Am.J.Phys.Anthropol.* (2005).

Capítulo II: Genotipado de STRs en Cromosomas Autosómicos

Chapter II: Autosomal

STRs Genotyping.



ABSTRACT

Most studies of Central America genetic markers show us hybrid populations, called Mestizo populations, as a result of admixture processes and also there are some studies about Native Americans genetic structure. The genetic structure of Mayan populations and the admixture processes have been studied previously in detail in nearby countries like Mexico, but in fact, Guatemala still today has different population groups whose structure expresses that they are isolated influenced by geographic, linguistic and cultural factors.

The objective of this study was to define the genetic structure within the Guatemalan population, Mestizo and Mayan groups which are located neighbouring central America populations. Moreover, we study the Mestizos group representing $\approx 60\%$ of the total of the Guatemala population in the present day. Admixture among Amerindians and Spanish (the five parental populations) currently constitutes the gene pool of Guatemalan Mestizos; so we study the contribution of Mayas and Spanish samples to the Mestizo gene pool.

In this work, we analyze the genetic data of fifteen autosomal short tandem repeat (STR) polymorphisms in 400 unrelated donors living in Guatemala, 200 Mestizos, 200 representing four Mayan population samples and 200 from unrelated donors from Spain.

The loci to analyze are D3S1358, TH01, D21S11, D18S51, D2S1338, D5S818, D13S317, D7S820, D16S539, CSF1PO, D19S433, vWA, D8S1179, TPO X, FGA, and these are analyzed in Mestizos (n=200), Kaqchiquel (n = 50), k'iche' (n = 50), Mam (n = 50), Q'eqchi' (n = 50), and in the Spanish population (n= 200).

Observed allelic frequency distribution and locus diversity values in these population groups are analyzed. The allele frequencies and genetic diversity among the five samples groups are compared, and also compared with Guatemala Mestizos.

The analysis of locus-by-locus pairwise and F_{st} between populations test discloses us the distance between the groups.

To characterize this country we also included STRs data of other European, Asiatic, Oceanian, African and also, American countries to compare with our populations studied. Genetic distance, multidimensional scaling and corresponding analysis have been performed.

In addition, this work contains the summary of the most common forensic data: Calculation race-based allele frequencies, percent heterozygote and homozygote, matching probability, power of discrimination, typical paternity index, power of exclusion and polymorphism information content (PIC) of the population studied.

Statistical analyses shows us the Mestizo composition, and intermediate position between Mayan and Spanish populations. These results show us the Mayan distances with other populations, which are expected considering the historical, geographic, linguistic and cultural factors.

1. INTRODUCCIÓN

Guatemala posee una población de 13.276.517 habitantes [1]. La mayor proporción son Mestizos resultado de la mezcla entre población amerindia y española. Más del 40% de la población es indígena, pueblos amerindios de procedencia Maya, cuyos grupos mayoritarios son los K'iche', Kaqchiquel, Mam, y Q'eqchi'. Existen otros grupos maya, pero no superan el 8% y otros indígenas no mayas que no superan el 0.25%. Se pueden encontrar también europeos, principalmente españoles, o descendientes de estos, criollos, en pequeño número, sin embargo es difícil diferenciar este grupo de los mestizos [2].

Son muchos los pueblos mayas que se pueden encontrar en los altiplanos guatemaltecos como los Achi, Akatek, Chuj, Ixil, Jakaltek, Kaqchiquel, K'iche', Mam, Poqomam, Poqomchi', Q'anjob'al, Q'eqchi', Tz'utujil and Uspantek. Pero son los Kaqchiquel, K'iche', Mam, y Q'eqchi' los más importantes, tanto en número de componentes, como en el área que ocupan [3].

Desde el punto de vista forense, se presenta una población difícil de caracterizar, con dos grupos poblacionales muy diferenciados, por un lado los mestizos y por otro los grupos mayas. La dificultad que suelen ofrecer los estudios de los grupos étnicos, que en muchas ocasiones, por causas de los fenómenos evolutivos (formación por efecto fundador, cuellos de botella y endogamia), aunque posean un origen común, pueden encontrarse grandes diferencias entre ellas.

Junto a la población de Guatemala, se han caracterizado 200 individuos españoles, que ayudarán a comparar muchos de los parámetros, forenses y antropológicos de la población de Guatemala.

En estos seis grupos poblacionales se estudiarán los siguientes quince marcadores microsatélite: D3S1358, TH01, D21S11, D18S51, D2S1338, D5S818,

D13S317, D7S820, D16S539, CSF1PO, D19S433, vWA, D8S1179, TPO X, FGA, y junto a estos la amelogenina (Tabla II.1). Todos los marcadores estudiados son tetranucleótidos, que se encuentran en distintos cromosomas autosómicos y son, en este momento, los más utilizados desde el punto de vista forense (estando trece de ellos incluidos en el sistema CODIS).

Tabla II.1: Locus STRs utilizados con mayor frecuencia, localización, unidad de repetición, acceso GenBank y rango alélico.

Locus	Localización cromosómica	Unidad de repetición (ISFH)	Acceso GenBank	Rango alélico
CSF1PO	5q33.3-34	TAGA	X14720	6-16
FGA	4q28	CTTT	M64982	15-51.2
THO1	11p15.5	TCAT	D00269	3-14
TPOX	2p23-pter	GAAT	M68651	6-13
VWA	12p12-pter	[TCTG][TCTA]	M25858	10-24
D3S1358	3p	[TCTG][TCTA]	No disponible	9-20
D5S818	5q21-31	AGAT	G08446	7-16
D7S820	7q11.21-22	GATA	G08616	6-15
D8S1179	8	[TCTA][TCTG]	G08710	8-19
D13S317	13q22-31	TATC	G09017	5-15
D16S539	16q24-qter	GATA	G07925	5-15
D18S51	18q21.3	AGAA	L18333	7-27
D21S11	21q21	Complex [TCTA][TCTG]	AP000433	24-38
D12S391	12	[AGAT][AGAC][AGAT]*	G08921	15-26
DIS1656	1q	TAGA	No disponible	15-26
D18S535	18	[GATA][GACA][GATA][GAT][GATA]	No disponible	9-16

En primer lugar, en estos seis grupos poblacionales se comprobará que los marcadores estudiados se comportan de la forma esperada, la homocigosidad de la muestra, si todos los loci se encuentran en equilibrio **HW** y que tanto unos locus como otros no se encuentran ligados.

Aunque el uso más frecuente de los STRs es la genética forense [4]; el uso de las bases de datos para hacer estudios de estimación de distancia genética poblacional, la estimación del mestizaje y la estructura poblacional, han sido ya validados en repetidas ocasiones por multitud de autores [5-8].

Desde el punto de vista forense, es importante conocer el poder de discriminación de cada uno de los loci, y ver también en su conjunto cómo se

comportan en cada población, para así poder comprobar la validez del conjunto de marcadores que ofrece el Kit Identifiler®, en esta población en particular.

Con especial atención hay que valorar los resultados obtenidos en los grupos más aislados, ya que se pueden encontrar fenómenos producidos por la consanguinidad y los efectos de la deriva genética [9]. Efectos como la pérdida de variabilidad alélica, la disminución de poder de discriminación, así como, el aumento de la homocigosis, pueden suponer un grave problema (como se verá más adelante) a la hora de realizar la identificación genética de un individuo.

Estos parámetros, frecuencias y poder de discriminación de cada locus, proporcionan una visión, a priori, de cómo serán posteriormente los cálculos más usados desde el punto de vista forense, como son el índice de paternidad (maternidad, hermandad, etc.) y la razón de verosimilitud (LR).

Por otro lado, desde un punto de vista antropológico y poblacional, se observará la distancia genética que separa a cada una de estas poblaciones. Determinando cuales son más cercanas, y si estos datos corresponden con los esperados desde el punto de vista geográfico, lingüístico,... también se comprobará el flujo genético que ha podido aparecer entre estas poblaciones y cuanto comparten entre ellas. Como ya se ha mencionado antes, la población mestiza en Guatemala se debe principalmente, al aporte de los indígenas mayas y de los españoles, por ello se encuentran individuos en estas poblaciones fruto de la mezcla de los marcadores genéticos [10]. Durante mucho tiempo ha existido un flujo migratorio continuo que no se sabe ciertamente en cuanto a afectado a la estructura genética de estas poblaciones.

Comparar los datos obtenidos en el análisis de estos grupos con otras poblaciones estudiadas para los mismos loci es muy interesante desde todos los puntos de vista.

Esto se debe a que los resultados obtenidos ayudarán a resolver la historia de nuestros grupos problema y también a interpretar resultados obtenidos en identificación. De esta manera se podrán interpolar los datos de distancia

obtenidos entre las poblaciones para nuevos posibles casos donde previamente no se haya realizado un estudio poblacional. Así mismo, es interesante para poder realizar mapas de asociación, estudios de evolución humana e incluso la predicción de riesgos de enfermedades asociadas a una población [11-13].

2. MATERIAL Y MÉTODOS

2.1 LA MUESTRA

Un total de 600 individuos han sido caracterizados en este estudio, 400 muestras fueron tomadas en Guatemala y 200 en España.

De los 400 individuos tomados en Guatemala 200 muestras procedían de mestizos, y otras 200 de grupos mayas: Kaqchiquel 50, k'iche' 50, Mam 50 y Q'eqchi' 50.

En la toma de la muestra se tomaron indistintamente hombres y mujeres, en todos los casos no relacionados entre sí. En el mismo momento de la toma de muestras se realizó una entrevista con el responsable de la toma (científico miembro del Ministerio Público de Guatemala) donde el mismo donante afirmaba su etnicidad, así como la de sus antepasados. Todos los donantes fueron informados del propósito de este estudio, aceptaron su participación en él y firmaron, según la guía ética de la declaración de Helsinki, un consentimiento informado.

A todos los donantes se les hizo una punción dactilar y se tomaron alrededor de 25µl de sangre que se depositaron sobre una tarjeta FTA (Whatman®).

Todas las muestras se dejaron secar, almacenaron, ordenaron, numeraron y precintaron en la sede investigaciones técnicas del Ministerio Público de Guatemala. Mediante un sistema de mensajería (DHL) llegaron al Laboratorio de Identificación Genética de la Universidad de Granada.

2.2 ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA

La toma de la muestra así como su posterior análisis están estrictamente registrados por rigurosos controles, tanto intralaboratorio como interlaboratorio. Todos los procesos de análisis llevados a cabo en el Laboratorio de Identificación Genética de la Universidad de Granada cumplen

los requisitos estándar de los laboratorios de genética forense. Todo perfil genético obtenido en éste sigue las recomendaciones de la ISFG (Sociedad Internacional de Genética Forense).

2.2.1 Extracción

La muestra llegó al laboratorio en papel FTA[®]. Este soporte facilita la extracción y purificación del ADN ya que lisa las membranas celulares, desnaturaliza las proteínas e inmoviliza los ácidos nucleicos al ponerse en contacto con este soporte. Es uno de los métodos más utilizados en el ámbito forense, por la facilidad tanto del transporte, como de su posterior extracción.

La purificación del ADN de la muestra se realizó siguiendo los protocolos optimizados de Whatman (Whatman, Clifton, NJ, USA).

2.2.2 Amplificación

Para poder detectar las zonas de estudio, se debe aumentar su número de copias, para lo que se realiza la reacción de la polimerasa en cadena, PCR, ideada en 1989 por Kary B. Mullis [14].

Para realizar esta técnica se tomó como ADN molde un punch de papel FTA (pequeño trozo de papel FTA, con forma circular y 1.2 mm de diámetro).

De éste se amplificaron las zonas que interesaban, un total de quince loci tipo STR y un locus indicador de sexo (amelogenina), incluidos en un kit multiplex: Identifiler[®] (Applied Biosystems, Foster City, CA), preparado para ser detectado por electroforesis capilar con emisión de fluorescencia [15]. La amplificación de todos los loci se realizó en un termociclador de la casa Biometra (T3 Thermocycler).

El **kit identifiler** permite el análisis simultáneo de 16 loci (Tabla II.2). Uno de los primeros 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.

Los fluorocromos que se usan para cada marcador son los siguientes:

- _ 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 II.2: Componentes, composición y volumen del Kit Identifiler®.

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 9947 ^a	ADN humano 0.1ng/µl	4µl

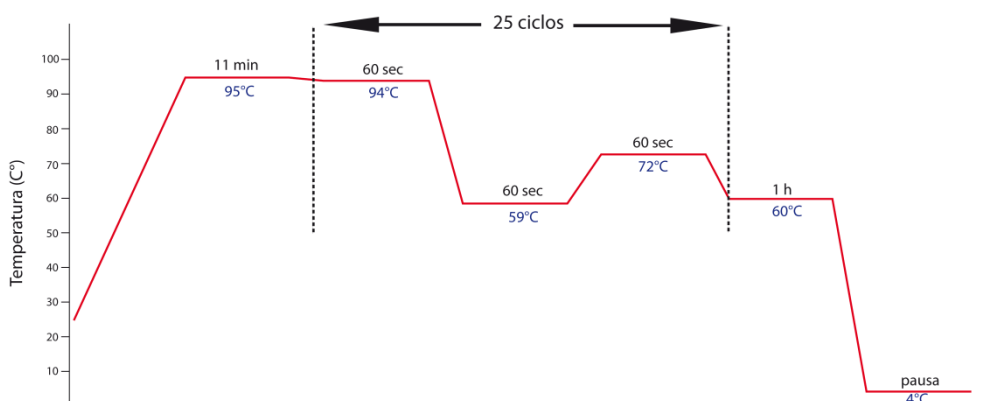


Figura II.1: Gráfico de las condiciones de amplificación.

En las amplificaciones (Figura II.1), además de las muestras, se amplificaron sistemáticamente dos controles: uno positivo, incluido en el kit, para comprobar que la reacción de amplificación ha transcurrido correctamente, y otro negativo, para descartar una posible contaminación por ADN extraño [16].

2.2.3 Electroforesis Capilar

Las muestras amplificadas con el kit fueron visualizadas por electroforesis capilar en un analizador genético ABI Prism 310 de Perkin Elmer [17].

La electroforesis capilar (EC), es un método para la separación de fragmentos por tamaño y carga, pero en este caso el soporte o medio de separación es un

polímero incluido en un capilar de sílica de unos 50 μ m y de 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. La gran ventaja en cuanto a rapidez de la técnica se debe a que la preparación del gel y la carga de muestras se hacen de manera automática. Aumenta la capacidad para diferenciar alelos que se distinguen en un par de bases. Además, los resultados obtenidos son analizados por un software evitándose así problemas de interpretación y permitiendo que estos queden almacenados para posibles futuros análisis.

Los analizadores genéticos constan de dos partes: una en la que se lleva a cabo todo el proceso de inyección de la muestra, electroforesis y detección de la fluorescencia y otra que es el sistema informático acoplado al analizador.

El capilar donde se lleva a cabo la electroforesis está recubierto por poliamida opaca excepto en una zona denominada “ventana del capilar” que atravesará el láser, donde se registrará la señal de los fluorocromos. La llegada de estos marcará el tamaño del fragmento (Figura II.2).

Tabla II.3: Condiciones analizador genético ABI Prism 310 y software asociado.

Preparación Muestras	Electroforesis:	Software:
24 μ l formamida desionizada	capilar 47cm X 50 μ m	GENESCAN
1 μ l estándar de tamaño. (500pb marcado en naranja-LIZ)	Polímero para el análisis de fragmentos, POP-4 (Performance Optimized Polymer 4, Perkin Elmer)	Asignación de Tamaño: “Local Southern Method”.
1 μ l producto amplificado	Tiempo de inyección.....3 sg.	GENOTYPER
** muestra asignación de alelos: 1 μ l ladder identifiler + 24 μ l formamida+ 1 μ l estándar de tamaño	Voltaje de inyección.....15 kv. Voltaje de electroforesis.....15 kv. Temperatura de electroforesis....60°C. Tiempo de electroforesis.....30 min.	Asignación de Alelos: Automated Genotyping with Identifiler Alelic Ladder

El software de recogida de datos, GeneScan [18], en primer lugar, define ciertas áreas en una cámara CCD (acoplada a la ventana del capilar) para recoger exclusivamente las emisiones procedentes del grupo de fluorocromos que se utilicen (Tabla II.3).

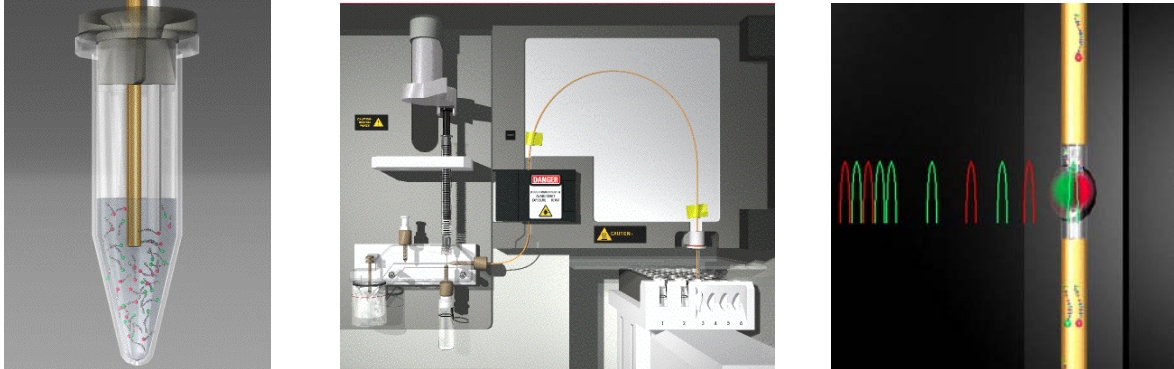


Figura II.2: El ADN se adhiere a las paredes del capilar. II.3: la diferencia de potencial entre los electrodos produce la migración entre ambos. II.4: El ADN marcado, produce fluorescencia al ser excitado por el láser al llegar a la ventana del capilar.

Estas áreas se conocen con el nombre de filtros virtuales y su funcionamiento es similar al de los filtros físicos que se utilizan para separar las diferentes longitudes de onda. El solapamiento entre los diferentes rangos que aún pudiera aparecer minimizado por una matriz matemática. En segundo lugar, este software trasforma la coordenada tiempo de llegada en pares de bases, gracias a un estándar de tamaño interno que se ha introducido en las muestras.

Un segundo software, GenoTyper [19], compara cada muestra con una muestra “allelic ladder” que contiene la mayoría de las variantes alélicas para asignar el número de repeticiones en cada STR [19].

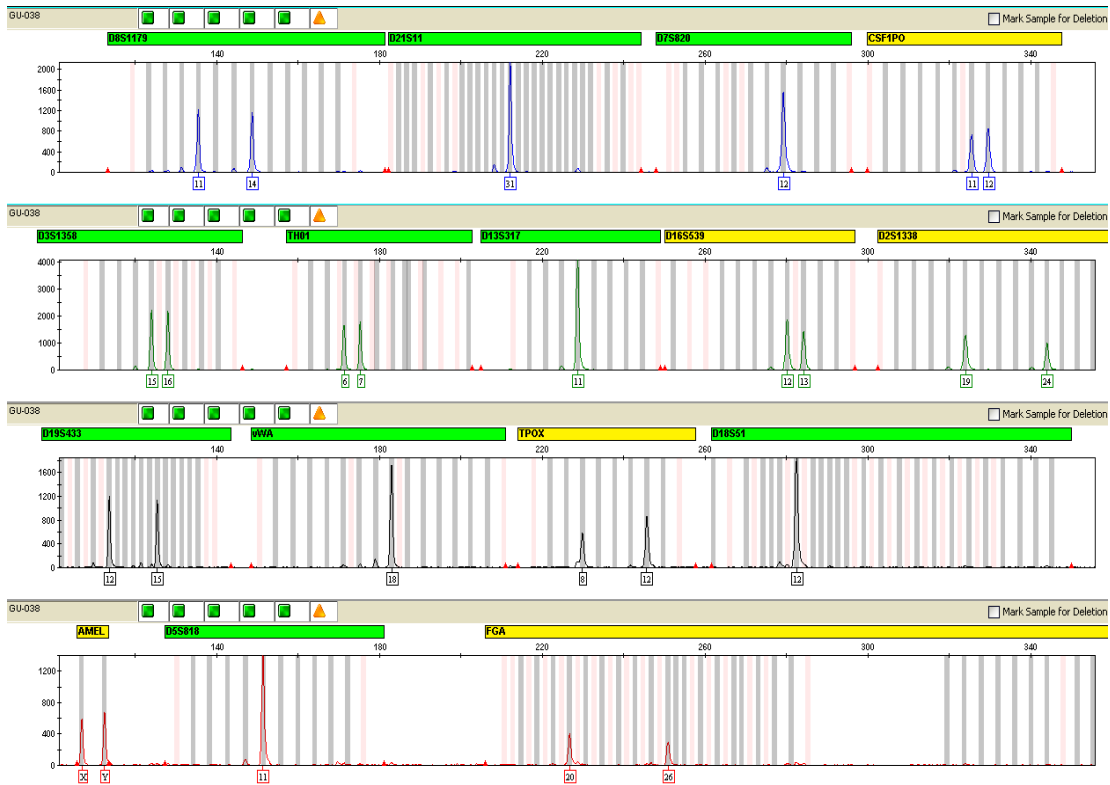


Figura II.5: Perfil genético obtenido de una de las muestras de este estudio. Incluidos los 15 marcadores STRs y la amelogenina.

Como se ve en la figura II.5, finalmente de cada una de las muestras se obtiene el perfil genético de un individuo, donde en cada uno de los locus estudiados se observa el número de repeticiones. Esta serie de quince parejas de números forman el perfil que será introducido en la base de datos, para su posterior análisis.

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, se hizo por simple recuento. Se estudiaron cada uno de los genotipos para cada marcador y se vio el número de individuos que compartían la misma repetición en cada marcador estudiado.

En segundo lugar, se recalculó este mismo parámetro con dos programas estadísticos, el Arlequin 3.1 [20] y PowerStatsV12, (Promega® Corporation) [21] donde los resultados obtenidos en todos los casos tienen que ser los mismos.

La heterocigosidad observada y esperada se calculó también mediante estos dos últimos programas, PowerStatsV12 y Arlequin 3.1. Éste último también sirvió como herramienta para calcular el equilibrio Hardy-Weinberg y el ligamiento entre los distintos loci, que se calcularon con un test con 5000 repeticiones experimentales para comprobar la asociación entre cada dos loci.

Todos los parámetros forenses más relevantes se calcularon mediante la hoja de cálculo PowerStats [21] obteniendo el poder de discriminación, el índice de paternidad, el poder de exclusión, y el PIC (polymorphism information content).

La distribución de los genotipos observados en cada grupo se comparó calculando el estadístico F , el análisis molecular de la varianza (AMOVA: Analysis of Molecular Variance), las distancias genéticas de Reynolds [22], la transformación lineal del F_{st} [22] y distancia de Slatkin [23] así como el test exacto.

Para calcular la distancia genética con poblaciones tomadas desde la bibliografía se utiliza la aplicación *Gendist*, del paquete informático, Phylip 3.66 que calcula las distancias entre poblaciones a partir de las frecuencias alélicas de cada una de las poblaciones estudiadas (los datos se tomaron de las publicaciones que aparecen en la tabla II.4).

Tabla II.4. Poblaciones, abreviatura, lengua, situación geográfica, marcadores estudiados y autor del estudio publicado

POBLACION	Abreviatura	LOCALIZACIÓN/PAÍS	STRs	N	Referencias
Kaqchiquel	Kaqchiquel	GUATEMALA	15	50	Estudio actual
K'iche'	K'iche'	GUATEMALA	15	50	Estudio actual
Mam	Mam	GUATEMALA	15	50	Estudio actual
Q'eqchi'	Q'eqchi'	GUATEMALA	15	50	Estudio actual
K'iche'	KicGU	GUATEMALA	15	27	Ibarra-Rivera et al. 2008 [24]
Kaqchiquel	KaqGU	GUATEMALA	15	59	Ibarra-Rivera et al. 2008 [24]
Yucatan	YucMX	MÉXICO	15	151	Ibarra-Rivera et al. 2008 [24]
Campeche	CamMX	MÉXICO	15	48	Ibarra-Rivera et al. 2008 [24]
Huastecos	HuaMX	MÉXICO	15	135	Barrot et al. 2005 [25]
Otomi (Sierra Madre)	OtoSMX	MÉXICO	15	91	Barrot et al. 2005 [25]
Otomi (Valle de Ixmiquilpan)	OtoIMX	MÉXICO	15	83	Barrot et al. 2005 [25]
Apache	ApaUSA	USA	13	198	Budowle et al. 2001 [26]
Navajo	NavUSA	USA	13	182	Budowle et al. 2001 [26]
Chukchi	ChuSI	Noreste Asiático	15	15	Zhivotovsky et al. 2009 [27]
Koryaks	KorSI	Noreste Asiático	15	32	Zhivotovsky et al. 2009 [27]
Even	EveSI	Noreste Asiático	15	14	Zhivotovsky et al. 2009 [27]
Altaianos-Kizhi	AltSI	Sur Siberiano	15	80	Zhivotovsky et al. 2009 [27]
Mongolia	MonSI	Sur Siberiano	15	42	Zhivotovsky et al. 2009 [27]
Tuvinianos	TuvSI	Sur Siberiano	15	80	Zhivotovsky et al. 2009 [27]
Khamniganos	KhaSI	Sur Siberiano	15	95	Zhivotovsky et al. 2009 [27]
Corea	Kor	COREA	15	49	Zhivotovsky et al. 2009 [27]
Belgorod	BelRU	RUSIA	15	66	Zhivotovsky et al. 2009 [27]
Pskov	PsKRU	RUSIA	15	62	Zhivotovsky et al. 2009 [27]
Velikiy	VelRU	RUSIA	15	59	Zhivotovsky et al. 2009 [27]
Asian SIB	Asi	Asia	15	13	Zhivotovsky et al. 2009 [27]
China Han in Jilin	Chi	CHINA	15	200	Yang et al. 2005 [28]
Kuala Lumpur	Mal	MALASIA	15	110	Maruyama et al. 2008 [29]
Mestizos	MesGU	GUATEMALA	15	200	Martinez-Espin et al. 2006 [30]
Chihuahua	MesMX	MEXICO	15	193	Martinez-Gonzalez et al. 2005 [31]
Nayarit	MXNaya	MÉXICO	15	231	Alvarez-Cubero et al. 2011 [32]
Descendientes africanos	AfUSA	USA	15	157	Budowle et al. 2001 [26]
Descendientes europeos	EuUSA	USA	15	160	Budowle et al. 2001 [26]
Descendientes hispanos	HiUSA	USA	15	151	Budowle et al. 2001 [26]
Descendientes de Nativos Americanos	NaUSA	USA	15	203	Budowle et al. 2001 [26]
Springfield	MesPR	PUERTO RICO	15	205	Zuñiga et al. 2006 [33]
Belem	MesBR	BRASIL	13	325	Ribeiro Rodrigues et al. 2006 [34]
España	Spain	ESPAÑA	15	200	Estudio actual
Flanders	Belg	BELGICA	15	222	Decorte et al. 2003 [35]
Croacia	Croat	CROACIA	17	200	Haliti et al. 2009 [36]
Cabinda	Cab	ANGOLA	17	110	Beleza et al. 2004 [37]
Tanzania	Tan	TANZANIA	15	225	Forward et al. 2008 [38]
Tonga	Ton	MADAGASCAR	15	51	Regueiro et al. 2007 [39]

Una vez obtenidas las distancias para facilitar su visualización se utilizaron distintos programas.

En el caso de las poblaciones de Guatemala y España, con las distancias genéticas, se realizó un MDS (mayas, mestizos y españoles). Para hacer el cálculo multivariante se utilizó el programa SPSS v15.0. En el caso del resto de las poblaciones la representación de las distancias se hizo mediante los programas SPSS v15.0 y STATISTICA v8. Con éste último se hizo el análisis de correlaciones entre las frecuencias de cada población. Para realizar el MDS las distancias de partida fueron las de Nei [40], Reynolds [22] y Cavalli-Sforza [41], obtenidas con gendist y los cálculos se representaron con ambos programas.

Para estudiar las poblaciones nativas americanas también se realizó una representación en dendograma. Se realizó mediante el software TREEVIEW for Win16, [42] (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>). Para poder realizar un dendograma hay que ir haciendo conjuntos, cluster, por similitudes e ir agrupando las poblaciones; esta agrupación para realizar los árboles se realiza por el método del Neighbor-Joining [43].

Se estimó el mestizaje producido en el grupo “mestizos de Guatemala” a partir de cada una de las 5 poblaciones parentales (los 4 grupos mayas y el español) haciendo los cálculos a partir de cada una de las poblaciones mayas y la española y posteriormente una última comparación con el conjunto de las 4 poblaciones mayas y la española. Para el cálculo del mestizaje en este tipo de población en Guatemala se utilizaron dos métodos distintos, una estimación por mínimos cuadrados (WLS), mediante el software ADMIX.PAS [44] y también un segundo cálculo con el programa ADMIX95 [45,46]. Un segundo cálculo de la estructura poblacional de los mestizos, así como de los otros 5 grupos parentales, se realizó con el programa STRUCTURE 2.3.1 [47-50] y se estimó con él las proporciones de individuos ancestros en cada una de las poblaciones (<http://pritch.bsd.uchicago.edu/structure.html>). Para realizar estos cálculos se hizo con un número de interacciones altas, debido a la similitud de las muestras, periodos “burn-in” de 25.000 interacciones, seguidos de 15.000

interacciones posteriores. Para las muestras mestizas se calcularon un número de poblaciones de partida (K) igual a 1, 2, 3, 4 y 5. Cuando se estudia el conjunto total de los datos (n=600) se hizo para un número de poblaciones 6, 3 y 2, pretendiendo ver si se encontraban diferencias significativas entre los grupos. En esta misma línea, se hizo un cálculo K= 1, 2 y 5, incluyendo a todos los guatemaltecos (n=400). Con el grupo maya completo (n=200) se calculó K=1,2 y 4. Por último se realizaron 8 experiencias más, calculando K=1 y K=2 en los grupos poblacionales maya (n=50).

3. RESULTADOS

3.1 ESTUDIO POBLACIONAL

En las siguientes tablas II.5, II.6, II.7, II.8 y II.9 se observan de forma resumida los resultados. Estas tablas contienen la mayoría de los datos más relevantes desde el punto de vista de la identificación genética en cada una de las poblaciones: las frecuencias alélicas de los 15 loci STR, la heterocigosidad observada (Het), el poder de discriminación (PD), el índice típico de paternidad (PI) y el poder de exclusión (PE).

Tabla II. 5: Frecuencia para los alelos observados en 15 loci incluidos en el kit Identifiler^R en la población Kaqchiquel. Het = Heterozigosidad observada, PD= poder de discriminación, PE = poder de exclusión.

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S739	D251338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
6						0.69									
7						0.22								0.08	
8							0.04					0.40			
9			0.02	0.02		0.02	0.25	0.14						0.06	
9.3						0.07									
10	0.06		0.19	0.31			0.15	0.30					0.01	0.01	
10.3			0.03												
11	0.03		0.35	0.15			0.13	0.29				0.36	0.01	0.60	
12	0.17		0.37	0.41			0.17	0.26		0.05		0.22	0.03	0.22	
12.2										0.05					
13	0.29		0.04	0.09			0.17	0.01		0.19		0.02	0.11	0.03	
13.2										0.30					
14	0.35			0.02	0.05		0.09			0.16	0.05		0.19		
14.2										0.02					
15	0.10				0.59					0.08	0.05		0.16		
15.2										0.08					
16					0.28					0.01	0.45		0.13		
16.2										0.05					
17					0.07				0.05	0.01	0.28		0.13		
18					0.01				0.10		0.13		0.15		
19									0.27		0.04		0.04		0.06
20									0.16				0.03		0.07
21									0.04				0.01		0.07
22									0.10						0.07
23									0.22						0.06
24									0.05						0.18
25									0.01						0.20
26															0.22
27															0.06
27.2		0.01													
28		0.03													0.01
29		0.31													
30		0.27													
30.2		0.03													
31.2		0.12													
32		0.02													
32.2		0.16													
33.2		0.05													
Het	0.091	0.011	0.626	0.029	0.110	0.723	0.240	0.030	0.920	0.935	0.510	0.901	0.042	0.926	0.648
PD	0.833	0.890	0.854	0.854	0.753	0.660	0.918	0.837	0.939	0.923	0.862	0.790	0.951	0.788	0.950
PE	0.527	0.590	0.448	0.461	0.315	0.245	0.661	0.492	0.659	0.667	0.461	0.378	0.730	0.352	0.709

Tabla II.6: Frecuencia para los alelos observados en 15 loci incluidos en el kit Identifiler^R en la población K'iche'.

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S739	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
6						0.53									
7						0.31								0.11	
8			0.01			0.02	0.03					0.47			
9				0.04			0.27	0.11				0.02		0.07	
9.3						0.14									
10	0.03		0.20	0.21			0.12	0.27						0.05	
11	0.02		0.28	0.22			0.09	0.25				0.24		0.55	
11.2										0.01					
12	0.2		0.41	0.45			0.25	0.30		0.03		0.25	0.04	0.20	
12.2			0.09							0.04					
13	0.34			0.07			0.12	0.07		0.16		0.02	0.15	0.02	
13.2										0.19					
14	0.24		0.01	0.01	0.02		0.12			0.29	0.06		0.23		
14.2										0.04					
15	0.13				0.61					0.17	0.04		0.19		
15.2										0.04					
16	0.04				0.25					0.02	0.42		0.07		
16.2										0.01					
17					0.11				0.09		0.35		0.15		
18					0.01				0.05		0.10		0.11		0.01
19									0.27		0.03		0.03		0.07
20									0.17				0.01		0.03
21									0.01				0.01		0.07
22									0.11				0.01		0.11
23									0.25						0.09
24									0.05						0.17
25															0.24
26															0.18
27															0.03
28		0.05													
29		0.22													
30		0.28													
31		0.08													
31.2		0.13													
32		0.01													
32.2		0.18													
33.2		0.05													
Het	0.206	0.274	0.847	0.821	0.044	0.874	0.468	0.079	0.260	0.812	0.177	0.942	0.767	0.952	0.964
PD	0.907	0.922	0.861	0.850	0.703	0.780	0.929	0.900	0.918	0.934	0.831	0.811	0.949	0.834	0.948
PE	0.553	0.631	0.456	0.460	0.306	0.331	0.631	0.529	0.628	0.647	0.439	0.391	0.693	0.415	0.700

Tabla II.7: Frecuencia para los alelos observados en 15 loci incluidos en el kit Identifiler^R en la población Mam.

Allele	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S739	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
6						0.51						0.01			
7						0.40								0.11	
8			0.01				0.03					0.42		0.01	
9			0.02	0.01			0.38	0.18						0.05	
9.3						0.08									
10	0.03		0.26	0.10		0.01	0.19	0.35						0.13	
11	0.03		0.25	0.32			0.06	0.23				0.27		0.50	
11.2										0.01					
12	0.12		0.41	0.30			0.17	0.21				0.30	0.04	0.16	
12.2										0.07					
13	0.43		0.05	0.10	0.01		0.11	0.03		0.09	0.01		0.07	0.04	
13.2										0.34					
14	0.22				0.08		0.06			0.27	0.07		0.25		
14.2										0.04					
15	0.17				0.54					0.10	0.01		0.15		
15.2										0.06					
16					0.25						0.39		0.11		
17					0.08				0.11	0.02	0.34		0.16		
18					0.04				0.05		0.11		0.11		
19									0.27		0.07		0.03		0.07
20									0.19				0.01		0.05
21									0.01				0.02		0.03
22									0.12						0.08
23									0.18				0.01		0.13
24									0.06						0.13
25									0.01				0.03		0.30
26													0.01		0.15
27															0.03
28		0.07													0.03
29		0.32													
30		0.23													
31		0.06													
31.2		0.10													
32		0.01													
32.2		0.13													
33.2		0.08													
Het	0.416	0.844	0.401	0.897	0.360	0.559	0.083	0.931	0.788	0.392	0.722	0.478	0.180	0.779	0.501
PD	0.875	0.912	0.824	0.864	0.781	0.716	0.890	0.876	0.940	0.913	0.849	0.807	0.953	0.873	0.941
PE	0.495	0.621	0.443	0.471	0.395	0.285	0.572	0.512	0.656	0.592	0.473	0.373	0.716	0.481	0.685

Tabla II.8: Frecuencia para los alelos observados en 15 loci incluidos en el kit Identifiler^R en la población Q'eqchi'.

Allele	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	TH01	D13S317	D16S739	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
6						0.55									
7						0.27						0.01		0.10	
8			0.02			0.01	0.02					0.47			
9			0.04	0.04			0.25	0.20				0.01		0.19	
9.3						0.17									
10	0.06		0.20	0.15			0.12	0.28						0.05	
10.3			0.05												
11	0.03		0.37	0.32			0.12	0.18				0.26		0.50	
12	0.18		0.27	0.43			0.27	0.30		0.03		0.24	0.09	0.11	
12.2										0.02					
13	0.34		0.04	0.05			0.15	0.04		0.12	0.02	0.01	0.07	0.05	
13.2										0.33					
14	0.18		0.01	0.01	0.05		0.07			0.15	0.05		0.22		
14.2										0.07					
15	0.14				0.61					0.15	0.09		0.14		
15.2										0.07					
16	0.04				0.26					0.03	0.37		0.07		
16.2										0.03					
17	0.03				0.05				0.05		0.27		0.19		
18					0.03				0.05		0.12		0.05		
19									0.16		0.06		0.10		0.08
20									0.17		0.02		0.04		0.04
21									0.05				0.02		0.08
22									0.08						0.06
22.2															0.01
23									0.39						0.11
24									0.05						0.11
25													0.01		0.22
26		0.01													0.23
27		0.01													0.05
28		0.01													0.01
29		0.25													
30		0.29													
31		0.06													
31.2		0.21													
32		0.01													
32.2		0.07													
33.2		0.08													
Het	0,280	0,304	0,262	0,101	0,776	0,155	0,944	0,037	0,545	0,809	0,231	0,548	0,127	0,687	0,718
PD	0,918	0,904	0,870	0,821	0,726	0,722	0,930	0,905	0,910	0,934	0,883	0,822	0,951	0,866	0,948
PE	0,605	0,600	0,521	0,436	0,313	0,326	0,622	0,526	0,592	0,656	0,556	0,382	0,728	0,472	0,711

Tabla II.9: Frecuencia para los alelos observados en 15 loci incluidos en el kit Identifiler^R en la población Mestiza.

Allele	D8S1179	D21S11	D7S820	CSFIPO	D3S1358	THO1	D13S317	D16S739	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
4						0.005									
5						0.003									
6						0.420									
7			0.008	0.005		0.250								0.068	
8	0.005		0.068	0.010		0.050	0.055	0.005				0.498		0.005	
9	0.008		0.048	0.023		0.090	0.248	0.113				0.045	0.003	0.083	
9.3						0.168									
10	0.060		0.208	0.235		0.015	0.118	0.240		0.005		0.043		0.043	
10.2										0.003					
10.3			0.005												
11	0.040		0.308	0.263			0.185	0.255		0.008		0.258	0.003	0.448	
11.2										0.003					
12	0.143		0.283	0.400			0.190	0.255		0.063		0.153	0.100	0.240	
12.2										0.015					
13	0.330		0.068	0.060	0.003		0.138	0.120		0.210	0.005	0.005	0.115	0.105	
13.2										0.160					
14	0.265		0.008	0.003	0.060		0.068	0.010		0.260	0.080		0.230	0.010	
14.2										0.040					
15	0.118			0.003	0.483			0.003	0.003	0.118	0.078		0.148		
15.2										0.085					
16	0.030				0.238				0.008	0.018	0.368		0.088		
16.2					0.003					0.010					
17	0.003				0.125				0.115	0.005	0.310		0.145		
18					0.085				0.033		0.105		0.070		0.003
19					0.005				0.180		0.043		0.048		0.098
20									0.160		0.010		0.023		0.050
21									0.023		0.003		0.023		0.085
22									0.123				0.005		0.120
23									0.235				0.003		0.118
24									0.063						0.168
25									0.045						0.005
26		0.003							0.015						0.170
27		0.005													0.145
28		0.075													0.035
29		0.255													0.003
30		0.235													0.003
30.2		0.003													
31		0.053													
31.2		0.143													
32		0.010													
32.2		0.148													
33		0.003													
33.2		0.058													
34		0.003													
34.2		0.005													
35		0.005													
Het	0.855	0.728	0.392	0.463	0.189	0.884	0.447	0.437	0.921	0.565	0.774	0.754	0.145	0.241	0.653
PD	0.858	0.878	0.948	0.960	0.957	0.877	0.945	0.908	0.914	0.870	0.950	0.897	0.920	0.826	0.967
PE	0.454	0.494	0.657	0.731	0.703	0.502	0.656	0.560	0.576	0.465	0.676	0.526	0.582	0.414	0.743

Tabla II.10: Frecuencia para los alelos observados en 15 loci incluidos en el kit Identifiler[®] en la población Española.

Allele	D8S1179	D21S11	D7S820	CSF1P	D3S1358	TH01	D13S317	D16S739	D2S1338	D19S433	VWA	TPOX	D18S51	D5S818	FGA
6						0.18						0.00			
7			0.02			0.16						0.00			
8	0.01		0.16	0.00		0.15	0.16	0.02				0.53		0.01	
9	0.00		0.13	0.01		0.22	0.07	0.12				0.11		0.03	
9.3						0.26									
10	0.11		0.28	0.28		0.01	0.05	0.07				0.05	0.00	0.05	
11	0.07		0.20	0.32			0.27	0.26				0.26	0.01	0.38	
<12					0.00										
12	0.13		0.15	0.31			0.28	0.29		0.10		0.02	0.16	0.31	
13	0.30		0.03	0.05			0.11	0.20		0.25			0.11	0.18	
13.2										0.01					
14	0.22		0.00	0.00	0.12		0.03	0.03		0.29	0.11		0.15	0.01	
14.2										0.03					
15	0.10				0.26					0.17	0.11		0.14		
15.2										0.06					
16	0.02				0.23				0.04	0.04	0.23		0.15		
16.2										0.02					
17	0.00				0.17				0.31	0.00	0.27		0.11		
18					0.18				0.06		0.15		0.07		0.01
19					0.01				0.10		0.08		0.04		0.06
20									0.13		0.01		0.01		0.12
21									0.05		0.00		0.01		0.19
22									0.04				0.00		0.19
22.2															0.00
23									0.09						0.17
24									0.07						0.13
25									0.07						0.07
26		0.00							0.01						0.02
27		0.02													0.00
28		0.14													
29		0.20													
29.2		0.00													
30		0.26													
30.2		0.03													
31		0.07													
31.2		0.09													
32		0.02													
32.2		0.10													
33.2		0.03													
Het	0.82	0.88	0.75	0.70	0.78	0.83	0.81	0.76	0.78	0.77	0.82	0.67	0.87	0.72	0.83
PD	0.93	0.94	0.93	0.86	0.92	0.92	0.92	0.92	0.95	0.93	0.93	0.79	0.96	0.86	0.95
PE	0.63	0.76	0.51	0.42	0.56	0.65	0.62	0.52	0.56	0.54	0.63	0.39	0.74	0.46	0.65

Al observar estas tablas se pueden ver los marcadores que ofrecen una mayor información para la identificación, así como para los análisis de parentesco, son el D18S51 y el FGA, en todas las poblaciones americanas. En la población española el primero de ellos es también el más informativo, pero el segundo más informativo es en este caso el D2S1338. Para todas las poblaciones el marcador menos descriptivo es el TH01.

El poder combinado de discriminación, calculado respecto del dato obtenido en cada locus, en todas las poblaciones es más alto que 0.99999. Pero si se observan todas las poblaciones, se ve que el valor obtenido en la población española y en la mestiza (de Guatemala) es mayor que los valores de PD totales en las poblaciones Kaqchiquel, K'iche', Mam y Q'eqchi'.

No se observa ningún tipo de desviación del equilibrio Hardy-Weinberg en ninguno de los grupos estudiados. La heterocigosidad se encuentra dentro de los valores normales para estos marcadores aunque desciende en las poblaciones mayas. Tampoco se encuentran signos de ligamiento entre ninguno de los loci.

3.2 DISTANCIAS GENÉTICAS

Las distancias calculadas a partir de los genotipos, Reynolds [22] y Slatkin [23], proporcionan resultados muy similares. Entre los grupos mayas y la población española la distancia siempre fue mayor de 0,046. Los grupos Kaqchiquel, K'iche', Mam y Q'eqchi' siempre tuvieron distancias pequeñas entre sí comprendidas entre el 0,007 y el 0,0007.

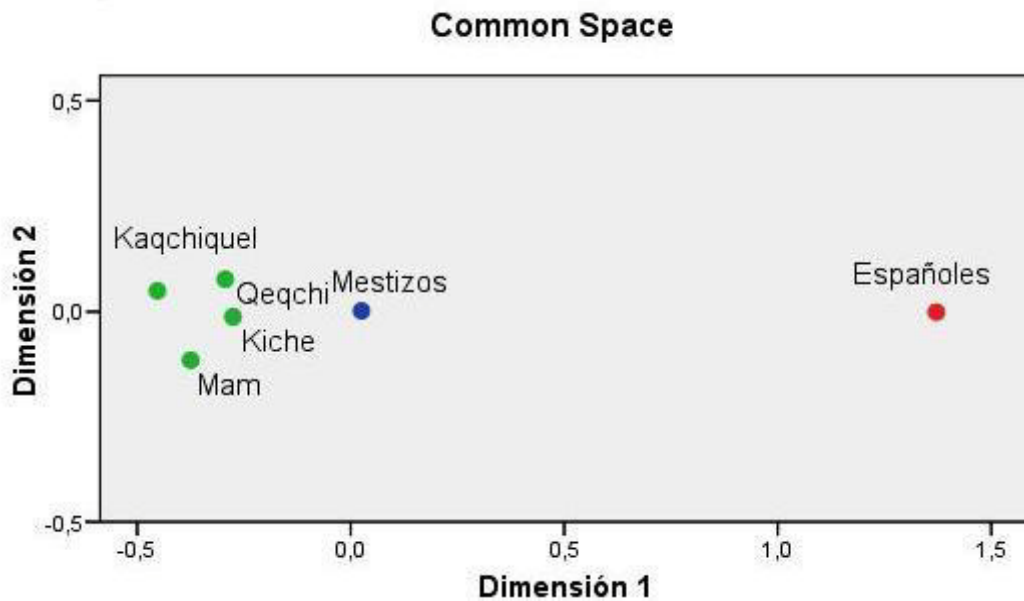


Figura II.6: MDS de las seis poblaciones estudiadas. Calculado a partir de las distancias de Reynolds et al. Stress = 0.0098

Para poder observar mejor las relaciones entre las cuatro poblaciones mayas, la mestiza y la española se calcula, a partir de la matriz de las distancias genéticas, se ha elaborado un MDS (escalamiento multidimensional), donde a la izquierda están las poblaciones de origen maya, muy cercanas a ésta, la población mestiza y más alejada la española.

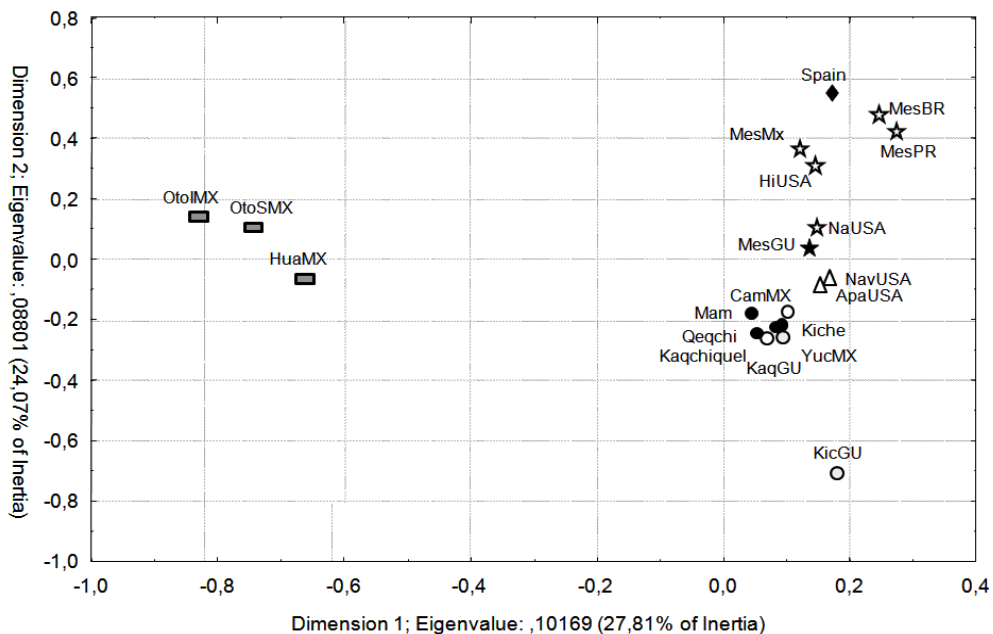


Figura II.7: Análisis de correspondencias de las frecuencias obtenidas en las seis poblaciones caracterizadas en este estudio y todas las poblaciones americanas, tanto nativas como mestizas.

Utilizando las frecuencias publicadas por otros autores se comparan nuestras poblaciones con otras poblaciones del continente americano. Dentro del análisis de distancias se realiza un análisis de correspondencias, mediante el programa informático Statistica. Los resultados de este análisis (fig. II.7) muestran cuan distantes se encuentran algunas poblaciones nativas americanas, como los Huastecos y los Oto-Manguelianos, que siendo el primero de ellos una población de origen maya está mucho más alejado de éstas con respecto a las poblaciones mestizas o incluso a la población española.

En la representación del análisis de correspondencias, las poblaciones se representan junto a las frecuencias de los marcadores, en la figura II.7 no se representan, porque al ser tantas, no se podrían distinguir la posición que toman las poblaciones. Las frecuencias que más influyen son la D3 13 repeticiones, que sitúa a los Huastecos y Otomis, y las frecuencias D2, 17 repeticiones y D5 13 que separan a los europeos y los nativos, respectivamente.

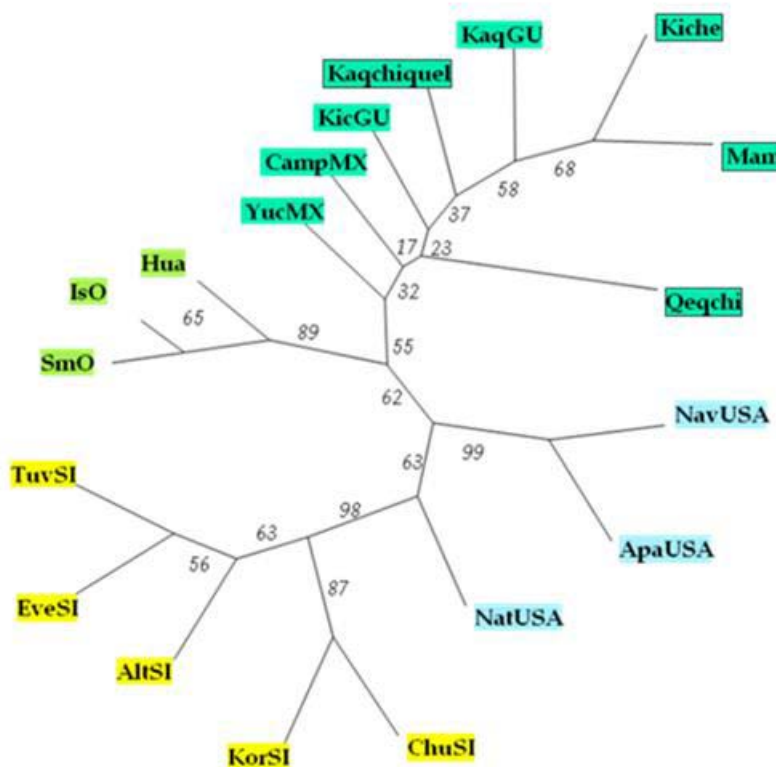


Figura II.8: Dendrograma de las poblaciones nativas americanas y siberianas obtenido a partir de las distancias genéticas de Reynolds con Treeview. Calculado a partir de las frecuencias con las distintas aplicaciones del programa phylip 3.66. En verde las poblaciones mayas, en azul nativas americanas del norte, en amarillo las siberianas.

Se calculan las distancias de Nei [40] de las poblaciones nativas americanas y se incluye en la comparación, las frecuencias de los grupos siberianos más cercanos. Se puede representar en un árbol filogenético estas distancias (fig. II 8), los resultados en primer lugar muestran cuatro cluster muy claros: siberianos, en color amarillo, na-dene (apaches y navajos) en color azul en la figura, nativos del norte de Méjico (Otomis y Huastecos) con lengua Oto-Manguiana y Maya (ambas proto mayas), en color verde en la representación y luego un último el grupo maya, en color turquesa.

Si se representa el MDS de estas distancias se observa cómo se agrupan en conjuntos las poblaciones. Al representar las poblaciones americanas con algunas siberianas (fig. II.9), se observa que al incluir las poblaciones mestizas

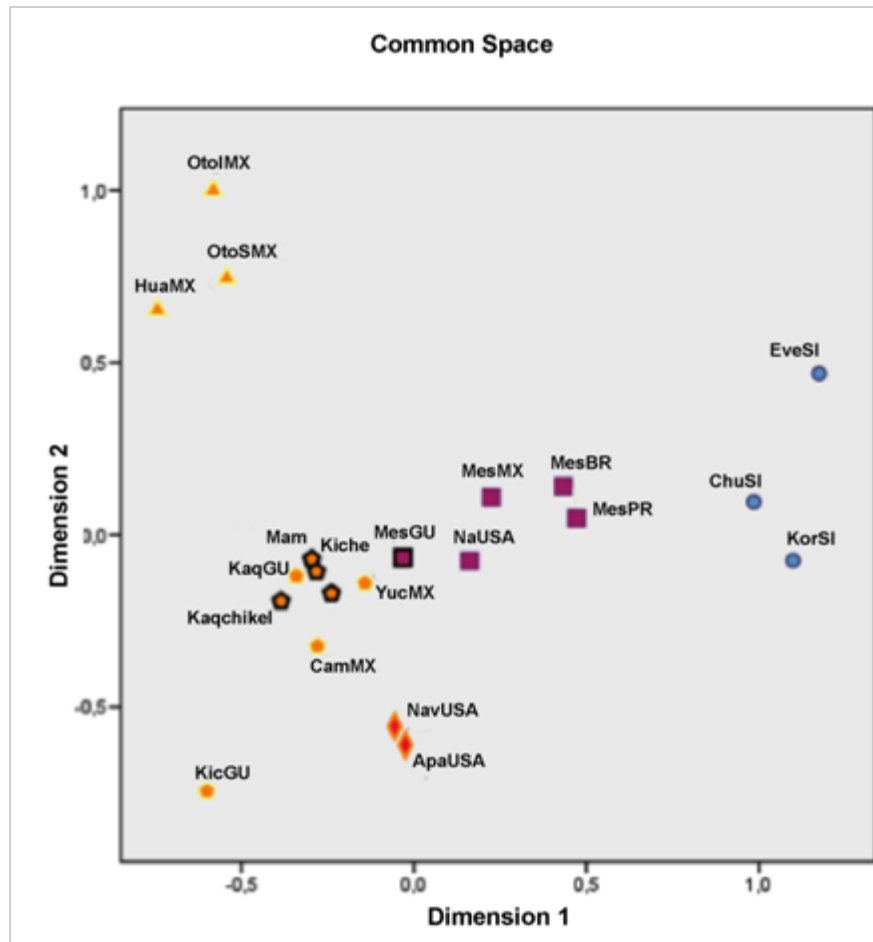


Figura II.9: Plot del MDS calculado a partir de las distancias de Nei de las poblaciones americanas nativas y mestizas. También se incluyen tres poblaciones de origen siberiano. Stress = 0.0134

éstas se agrupan en un punto intermedio entre las poblaciones siberianas y las nativas americanas. En la figura II.9 se puede ver como al igual que ocurría en el análisis de correspondencias las poblaciones nativas norte mejicanas se alejan bastante del resto de las poblaciones.

Se realiza un cálculo de la varianza molecular (AMOVA), tabla II.11. Comparando las poblaciones nativas americanas. Haciendo dos grupos, uno donde se agrupan las poblaciones por el origen de su lengua; y un segundo donde lo hace por su situación geográfica. El cálculo muestra una variación menor dentro de los grupos realizados por lengua (2,51%) que en los realizados por su situación geográfica (4,34%).

Tabla II.11: Estudio AMOVA de las poblaciones nativas americanas, teniendo en cuenta su lengua y su situación geográfica.

Fuente de variación	d.f.	s.Cuadrados	Varianza componentes	Porcentaje de variación	Índices de fijación
Grupos con mismo origen lingüísticos					
1. Gran grupo Maya: Kaqchiquel, Mam, K'iche', Q'eqchi', Yucatán, Huastecos y Campeche.					
2. Oto-Mangueno: Otomi (Sierra Madre y Valle de Ixmiquilpan).					
3. Na-Dene: Descendientes de nativos de EEUU, Apaches y Navajos					
4. Eskimo: Altaians-Kizhi, Tuvinians, Chukchi, Koryaks y Even.					
Entre los grupos	3	59.790	0,08994 Va	21,47%	FSC:0.03202
Dentro de cada grupo	15	12.698	0,01053 Vb	2,51%	FST:0.23987 FCT:0.21473
Grupos con situación geográfica próxima					
1. Asia y Siberia: Altaians-Kizhi, Tuvinians, Chukchi, Koryaks y Even.					
2. EEUU: Descendientes de nativos de EEUU, Apaches y Navajos.					
3. CentroAmerica: Kachiquel, Mam, Ki'che', Q'eqchi'.					
4. México: Campeche, Yucatan, Huastecos y Otomi (Sierra Madre y Valle de Ixmiquilpan).					
Entre los grupos	3	54.237	0,07218 Va	17,67%	FSC:0.05330
Dentro de cada grupo	14	18.252	0,01793 Vb	4,39%	FST:0.22058 FCT:0.17669

Por último, se calcula la matriz de distancias de Nei [40], Reynolds [22] y Cavalli-Sforza [41], de poblaciones norte, centro y sur americanas (nativas y mestizas), poblaciones norte europeas, africanas y asiáticas (éstas últimas, en su mayoría siberianas); todas estas poblaciones están incluidas en la tabla II.4.

Si se realiza una representación euclídea incluyendo también la matriz de distancias, principalmente se resaltarán las diferencias y semejanzas entre las

poblaciones. En el caso de realizar la matriz de distancias con las 41 poblaciones se obtendría una agrupación en cinco conjuntos (fig. II.10). Un primer conjunto, en el cuadrante superior derecho, que muestra todas las poblaciones mayas, de Guatemala y del sur de Méjico. En el segundo conjunto, situado junto al anterior en el cuadrante inferior, se ven los nativos americanos del norte, de Estados Unidos de América y del Norte de Méjico. En el cuadrante inferior izquierdo se agrupan todas las poblaciones siberianas y en superior las europeas y africanas. En el centro del gráfico se encuentran las poblaciones mestizas, las asiáticas (más occidentales), las norteamericanas descendientes de europeos e hispanos y la española.

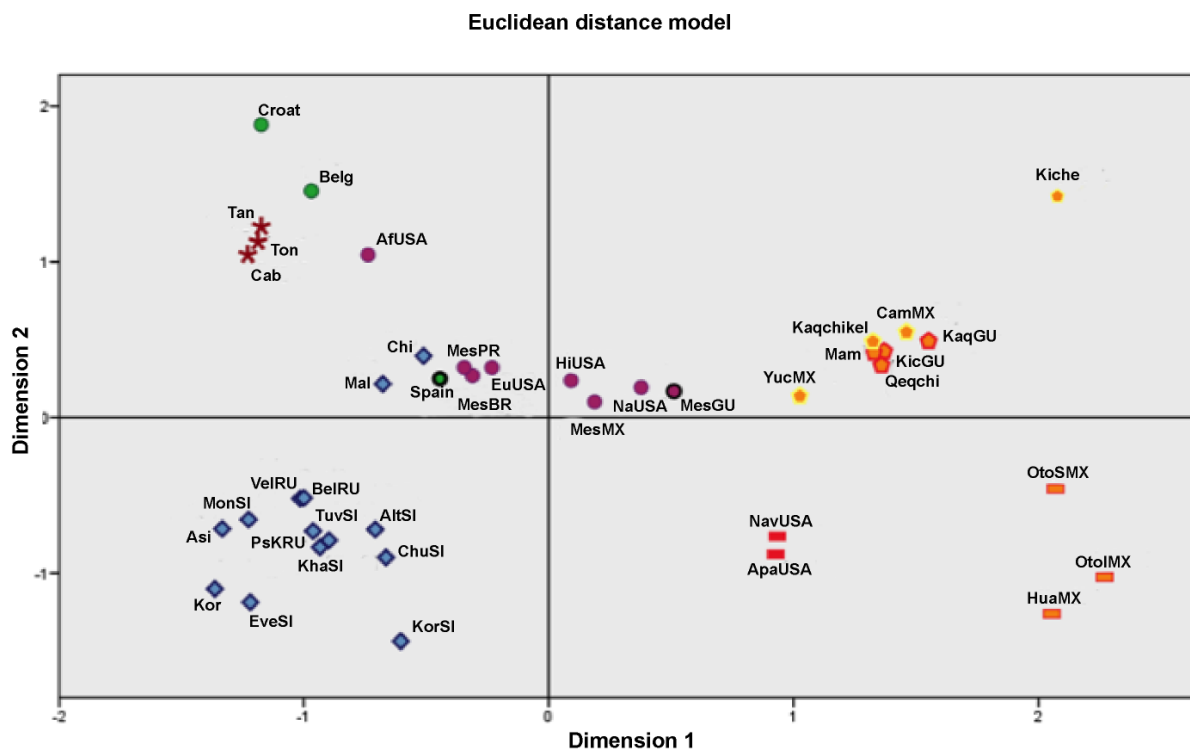


Figura II.10: Plot de las distancias euclídeas calculado a partir de las distancias de Nei de las poblaciones. En este análisis se comparan: Poblaciones Europeas, Americanas del norte y del sur, Africanas y asiáticas (incluyendo un gran número de poblaciones siberianas). Stress = 0.1021

3.3 ESTUDIO DEL MESTIZAJE

Uno de los análisis más interesantes que se incluyen al estudiar la población de Guatemala, es cuál ha sido el aporte de mayas y de españoles a la población

mestiza de este país. Este análisis se realiza, comparando las frecuencias de la población mestiza con las parentales.

La tabla siguiente es un resumen de las proporciones de mezcla de cada grupo. Para realizar esta comparación se toman como parentales cada una de las poblaciones mayas y la española. También se toma la suma de las frecuencias de las cuatro poblaciones mayas (nombrada como maya) para poder estimar el aporte total indígena con respecto al español.

Tabla II.12: Cálculo del mestizaje. Aportación al grupo mestizo por parte de los españoles y cada uno de los grupos poblacionales mayas (filas 1 a 12) y la población maya en general (última columna) Calculado mediante dos programas estadísticos Chakravorty y Long.

	Identidad Genética /Chakravorty		Mínimos cuadrados /Long	
Población	Kaqchiquel	Española	Kaqchiquel	Española
Proporción	0,532	0,468	0,565	0,436
Desviación	0,011	0,011	0,052	0,052
Población	K'iche'	Español	K'iche'	Española
Proporción	0,504	0,499	0,520	0,480
Desviación	0,014	0,014	0,058	0,058
Población	Mam	Española	Mam	Española
Proporción	0,553	0,447	0,5267	0,4733
Desviación	0,006	0,006	0,068	0,068
Población	Q'eqchi'	Española	Q'eqchi'	Española
Proporción	0,597	0,403	0,618	0,382
Desviación	0,006	0,006	0,049	0,049
Población	Mayas	Español	Mayas	Española
Proporción	0,627	0,373	0,660	0,340
Desviación	0,008	0,008	0,048	0,048

Los resultados que se muestran han sido calculados por dos métodos distintos, el método de identidad genética [45,51,52] y el de los mínimos cuadrados [44]. Como resultados con el modelo biparental (españoles y mayas), se obtiene un intervalo 62-66% de mayas y de un 34-37% de españoles. Resultados muy similares se han obtenido cuando el análisis se ha realizado con cada uno de los subgrupos poblacionales mayas (Tabla II.12).

Para poder observar estos mismos resultados con otros métodos que completan la información actual, se estudió la composición de nuestra población mediante

el programa STRUCTURE 2.3.1 [48,50]. Gracias a este programa se detectan la presencia de uno o varios grupos poblacionales dentro de cada una de las poblaciones. Ésta es una forma de ver si en un grupo muestral existen diferencias significativas como para poder diferenciarlos en poblaciones distintas.

Lo más interesante en este caso es ver si en la población mestiza existen individuos con diferencias significativas como para poder subdividirlo en subgrupos. Pero también, esta herramienta permite ver si detecta diferencias en la totalidad de las muestras (españoles mayas y mestizos).

Por esta razón los análisis se realizaron de la siguiente manera:

- Cada uno de los grupos muestrales por separado (Mestizos, Kaqchiquel, K'iche', Mam y Q'eqchi') con $K=1$, $K=2$ and $K=3$ (K asume el tamaño de la población)
- Todas las muestras procedentes de Guatemala (los cinco grupos poblacionales) en este caso con $K=1$, $K=2$ y $K=5$. Intentando encontrar diferencias entre cada una de ellas.
- Por último se somete a todo el grupo muestral a un análisis ($n=600$) a un análisis asumiendo $K=1$, $K=2$, $K=3$ o $K=6$.

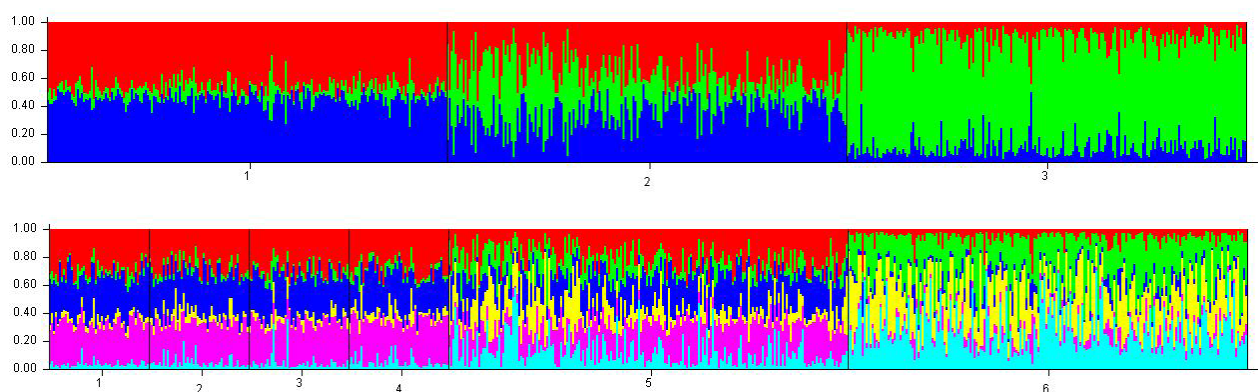


Figura II.11: Imagen obtenida en Structure, asumiendo una población formada por tres grupos poblacionales, $K=3$. Figura II.13: Imagen obtenida en Structure, asumiendo una población formada por seis grupos poblacionales, $K=6$.

Como parámetros para el uso del programa se predefine antes de cada uno de los análisis un número de 50.000 interacciones y un periodo “*burn-in*” de 10.000 interacciones adicionales. Los resultados se resumen en la tabla II.13.

Tabla II.13: Probabilidad obtenida asumiendo un número de poblaciones (k), partiendo de un modelo de muestra mixta.

Población de estudio	STRs	n	Población asumida	Ln P (D)	Var [Ln] P (D)]
Mestizos	15	200	K=1	-10468,5	68,0
Mestizos	15	200	K=2	-10796,4	854,0
Mestizos	15	200	K=3	-10763,4	900,1
Mestizos	15	200	K=4	-10760,9	957,7
Mestizos	15	200	K=5	-10997,0	1428,7
4 Mayan	15	200	K=1	-9625,6	57,4
4 Mayan	15	200	K=2	-9670,6	201,2
4 Mayan	15	200	K=4	-9763,6	405,3
5 pop Guatemala	15	400	K=1	-20307,7	688,5
5 pop Guatemala	15	400	K=2	-20241,4	70,4
5 pop Guatemala	15	400	K=5	-21685,3	3755,4
Spain-Guatemala	15	600	K=1	-31674,3	71,8
Spain-Guatemala	15	600	K=2	-30700,8	413,0
Spain-Guatemala	15	600	K=3	-30767,3	727,0
Spain-Guatemala	15	600	K=6	-30672,7	1302,8
Kaqchiquel	15	50	K=1	-2353,9	39,9
Kaqchiquel	15	50	K=2	-2364,4	65,6
K'iche'	15	50	K=1	-2457,2	43,6
K'iche'	15	50	K=2	-2462,6	56,6
Mam	15	50	K=1	-2385,7	46,5
Mam	15	50	K=2	-2396,4	48,9
Q'eqchi'	15	50	K=1	-2381,8	41,9
Q'eqchi'	15	50	K=2	-2386,0	51,8

Cuando se toma el grupo muestral completo (n=600) la mayor probabilidad la se encuentra en K=2, lo que significa que se podrían diferenciar mejor dos subgrupos poblacionales, el análisis encuentra mayor probabilidad de que haya dos poblaciones ancestrales. Igualmente ocurre cuando se toman mestizos y mayas juntos (n=400), la población de Guatemala se podría diferenciar en dos grupos.

Cuando se estudia el grupo de los españoles, así como los grupos mayas individualmente, no se encuentran, durante el análisis, signos de más de un grupo poblacional. Sin embargo, cuando se hace lo mismo con el grupo mestizos, K=1 y K=2 tienen unos valores similares, lo que pone en duda la asignación del número de poblaciones en ésta. Los mestizos podrán estar formados por una o varias poblaciones ancestrales, que perduran en la estructura poblacional actual.

4. DISCUSIÓN

La mayoría de los estudios poblacionales a nivel genético han demostrado que existe un bajo nivel de diferenciación en todas las poblaciones del mundo (Excoffier 2003). Este bajo nivel se pone aún más de manifiesto cuando se estudian este tipo de marcadores, que presentan un alto grado de polimorfismo. Se encuentran mayor número de diferencias al comparar dos individuos de la muestra que cuando se comparan grupos muestrales completos. Por tanto, siempre habrá una mayor diferencia interpoblacional que intrapoblacional.

Aunque estos niveles de diferenciación son bajos en general; factores locales, diferencias geográficas, lingüísticas, culturales, etc. pueden ir creando una serie de variaciones en unas poblaciones que no se produzcan en otras, creando una serie de discontinuidades o saltos que permiten estudiar cómo se segregan o se hibridan los distintos pueblos.

En este capítulo se muestran estos quince marcadores del genoma nuclear (kit Identifiler) como una herramienta muy poderosa para realizar la identificación genética. De igual modo, estos loci expresan la estructura de la población de Guatemala mostrando las diferencias entre los distintos grupos poblacionales. El estudio de estos marcadores supone un beneficio potencial para Guatemala, tanto para su uso en identificación humana, como para conocer su estructura poblacional. Lo que permite hacer estudios diagnósticos y también descubrir, desde el punto de vista genético, el desarrollo de la historia de los pueblos mayas.

4.1 APORTACIÓN A LA CIENCIA FORENSE

En primer lugar, el análisis de este número de muestras incrementa el número de STRs estudiados, aumentando así las bases de datos generales de STRs, y en particular permite hacer estudios con estos marcadores en la población de

Guatemala. Estudiar por separado las poblaciones mayas de la población mestiza permite tener una base de datos específica para estos grupos poblacionales, y poder hacer estudios individualizados dentro de estas comunidades. Los loci analizados D3S1358, TH01, D21S11, D18S51, D2S1338, D5S818, D13S317, D7S820, D16S539, CSF1PO, D19S433, vWA, D8S1179, TPO X, y FGA ponen de manifiesto que el análisis del CODIS, y en particular del kit Identifiler, con dos marcadores más que el CODIS, es una herramienta muy potente para la identificación humana, incluso en poblaciones aisladas. Las seis tablas de resultados, con las frecuencias encontradas y los parámetros calculados pueden ser ya utilizadas para poder resolver miles de casos forenses. El conjunto de marcadores tiene un gran poder de discriminación, ninguno de los marcadores se encuentra ligado y en todas las poblaciones se encuentran en equilibrio HW. Pero se observa la bajada de algunos parámetros como la heterozigosis, el poder de discriminación, el poder de exclusión,... en las poblaciones maya; Esto obliga a tomar precauciones a la hora de enfrentarse a casos en estas regiones. Precauciones tanto en la toma de las muestras, donde habrán de buscarse familiares con relación directa. Y precauciones en el laboratorio donde habrá que incrementar el número de marcadores a estudiar y las características de los mismos.

El estudio de comparación y el de mestizaje también aportan al campo forense una gran cantidad de datos, que facilitarán la aplicación de las herramientas (tablas parámetros y probabilidades). Todos estos datos ayudan a enfrentarse a los distintos casos que se pueden encontrar. Cuando se estudian las distancias poblacionales se ven las enormes diferencias que existen dentro de algunas poblaciones, como se puede observar tanto en los estudios realizados tanto en Estados Unidos de América como en Méjico. Diferencias, sobretodo, acentuadas en poblaciones indígenas o aisladas. En cambio se ve como las poblaciones mestizas, tienden a homogeneizarse mostrando frecuencias similares.

Estudiando la estructura genética de Guatemala mediante Structure, se ve cuanta similitud existe entre los pueblos mayas, de Guatemala. Existe tanta similitud, que los distintos programas estadísticos no son capaces de encontrar

suficientes rasgos como para diferenciarlas. Esta semejanza hace que se pueda agrupar la población maya en un solo grupo, para poder utilizar los datos obtenidos para la población maya en general y no en subgrupos poblacionales.

4.2 APORTACIÓN A LA ANTROPOLOGÍA

Durante el periodo prehispánico, los grupos mayas que habitaban centro América eran los grupos más numerosos, poderosos y fuertes desde un punto de vista político, religioso y social. Esto no tiene nada que ver con el resto de los grupos indígenas del continente americano, tanto del norte como del sur. Por eso, este grupo poco a poco se fue haciendo más numeroso y este desarrollo supuso también el de la lengua maya, que se fue extendiendo y divergiendo por todo Centroamérica, persistiendo hasta nuestros días.

La caracterización genética de estos cuatro grupos mayas revela la gran asociación entre estos y las diferencias en la diversidad genética con otras poblaciones, incluso con otras poblaciones nativas americanas, que no sean de origen maya. Ésta es una señal de la gran cohesión social durante muchos años de las poblaciones mayas. Resultados muy similares fueron descritos por Ibarra-Ribera [24] con grupos mayas mejicanos y guatemaltecos. Apoyan también los resultados obtenidos en Méjico por Silva-Zolezzi [53] mediante el uso de SNPs.

También en los análisis realizados mediante el estudio AMOVA se confirma, que la distancia que se encuentra entre las poblaciones que poseen una lengua común es muy similar, frente a las distancias entre grupos con distinto origen que poseen una gran divergencia.

Como se ve, las poblaciones más cercanas a la población mestiza en Guatemala son la K'iche' y la Q'eqchi', la primera por ser la más numerosa en Guatemala y la segunda por ser la que más extensión ocupa, esto ha hecho que a lo largo de los años sean las que más se entremezclen con las poblaciones europeas.

En general, el estudio de la población mestiza, con respecto a sus distancias genéticas, pone de manifiesto que no existen grandes diferencias entre el grupo mestizo y otros grupos mestizos tanto del centro como del sur de América que también se encuentran cercanos a los norteamericanos de origen latino.

Por otra parte, estos grupos de Guatemala, se encuentran muy alejados de las poblaciones Africanas, y también de aquellas poblaciones descendientes de origen africano. Estos datos genéticos apoyan los datos históricos, que afirman que los esclavos se introdujeron en la costa Atlántica, tanto en las islas como en Sudamérica, en las grandes plantaciones, y posteriormente se fueron extendiendo en el continente, en este mismo eje, ocupando las zonas costeras más deshabitadas. También explica, el bajo número de mulatos y zambos que actualmente residen en Guatemala.

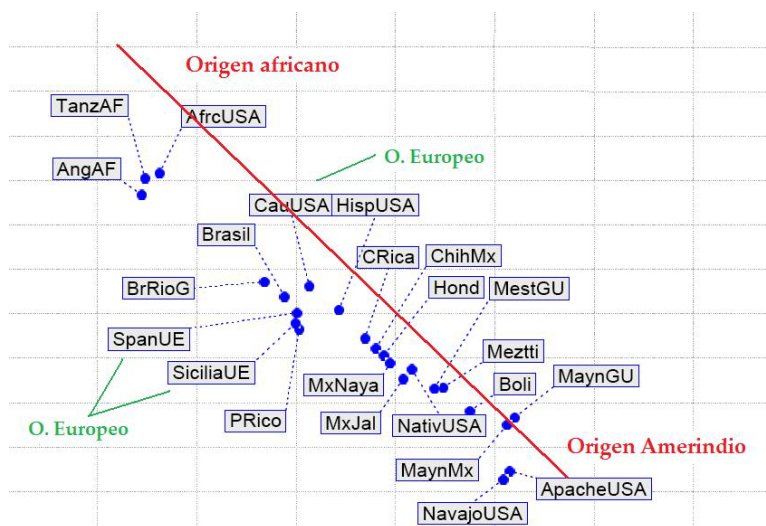


Figura II.12. Análisis de Correspondencias de la población mestiza americana, frente a poblaciones de origen europeo, africano y amerindio. Se puede ver como poblaciones como Brasil, Costa Rica u Honduras se sitúan cercanas a las poblaciones de origen africano. En cambio, Bolivia o Guatemala se sitúan más próximas a las poblaciones amerindias.

Otra de las dudas que se quería resolver es el origen de estas poblaciones mayas. Es difícil buscar relaciones entre estas poblaciones mayas con las poblaciones de origen asiático, como pueden ser poblaciones esquimales y siberianas, con estos marcadores.

Si se eliminan de nuestros estudios las poblaciones mestizas (Fig. II 9), se observa como del grupo siberiano el cluster más cercano son las poblaciones Na-Dene y los más alejados a estos son los Mayas de Guatemala. Dentro de

estos dos grupos se encuentran los premayas Huastecos (Otomis y Huastecos) y los mayas de la rama lingüística del Yucatán (Yucatán y Campeches). Lo que apoyaría más a aquellos que defienden que el origen de las poblaciones americanas se produjo con una sola migración desde Asia, ya que cada uno de los cluster se va agrupando y diferenciando según se avanza en el eje Norte→Sur.

Se pone de manifiesto con este estudio que este tipo de marcador, STRs autosómicos hipervariables, es muy bueno para comparar sucesos “recientes”, de este último milenio, pero es difícil establecer relaciones ancestrales con poblaciones asiáticas. Lo que hace que no se pueda confirmar la teoría de Goebel, pero si es un dato más para apoyar una única migración.

Es difícil estudiar las relaciones ancestrales de los nativos americanos. Al estudiar todas estas poblaciones que han estado durante tanto tiempo alejadas, se observa una gran divergencia, como ocurre entre mayas y premayas (Otomanes y Huastecos), o con Navajos y Apaches. La gran recombinación de estos marcadores y su tasa de mutación produce este tipo de resultados. El estudio de la historia, así como datos genéticos obtenidos mediante otros marcadores, [54,55], ayudará a poner de manifiesto el origen común de algunas poblaciones. Las diferencias acumuladas durante su historia ancestral, el número de mutaciones acumuladas, la endogamia, el efecto fundador en las distintas aéreas del continente Americano, pueden explicar los resultados obtenidos en estos marcadores de tipo STR autosómicos y su segregación en la estructura en los nativos americanos.

En la figura II.13, se ve como se pone de manifiesto esta última afirmación. Los marcadores estudiados diferencian los pueblos mayas, en los mismos grupos que lo hace la historia de su lengua. Pero en cambio, es difícil encontrar el origen común de K'iche', Kaqchiquel y mayas, cuando según el origen de su lengua es el mismo. O de igual modo, debería de aparecer un grupo común Yucatán y Campeche, etc.

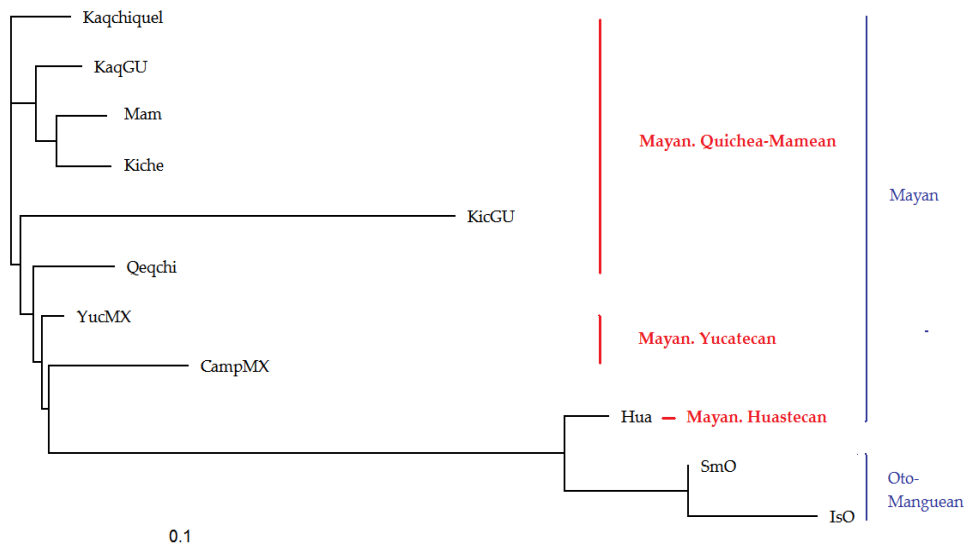


Figura II.13: Filograma que muestra las relaciones y evolución de poblaciones nativas mesoamericanas mayas y otomangues, en este filograma se han incluido las lenguas habladas por cada grupo poblacional. Se realizó con la distancia de Nei. Phylip 3.66 y Treeview

Estos primeros datos revelan la alta calidad de los marcadores STR para realizar estudios de la historia moderna de las poblaciones. También muestra cómo esta herramienta disminuye su poder discriminativo a la hora de hacer estudios ancestrales, donde se deberá utilizar otro tipo de marcadores con menor tasa de mutación, como pueden ser algunos SNPs, o buscar zonas del ADN con la recombinación más limitada, como el ADNmt o el cromosoma Y.

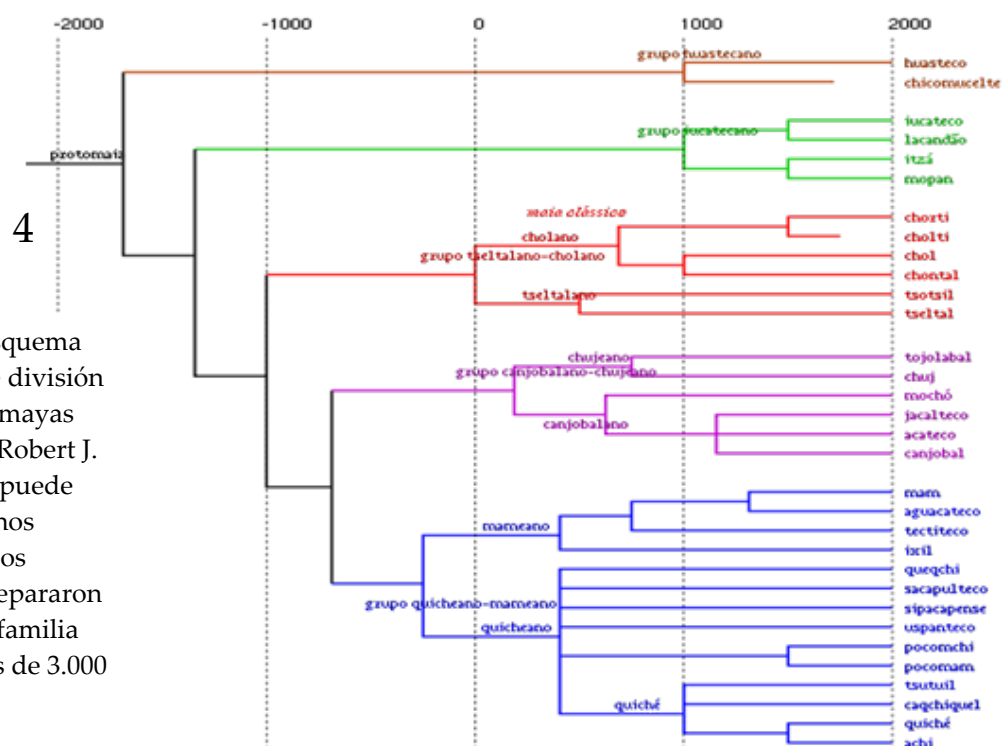


Figura II.14. Esquema genealógico de división de las lenguas mayas. Realizado por Robert J. Sharer [56]. Se puede ver como algunos grupos, como los huastecos, se separaron del resto de la familia maya hace más de 3.000 años.

.3 ESTUDIOS DE MESTIZAJE

Los estudios de comparación realizados en este trabajo, indican que el pool genético en la población mestiza de Guatemala es predominantemente una mezcla equilibrada entre nativos americanos de origen maya y la población europea, en su gran mayoría española; datos concordantes con la historia de este país durante estos últimos 500 años y con los resultados realizados en otras poblaciones latinoamericanas [56-61].

Como se deriva de este trabajo, se ve que mediante el análisis de marcadores autosómicos, el aporte a la población mestiza es más o menos equilibrado, en todos los marcadores. Así como en el caso en el que se toma de referencia, cualquiera de las poblaciones mayas, el aporte de las poblaciones nativas americanas es mayor que el aporte de la población europea. Esto puede deberse a que el número de individuos que por migración llegaron a Guatemala fue siempre menor al número de nativos indígenas; o a que siempre haya existido un flujo entre las poblaciones mayas y la población mestiza. Esto difiere de estudios realizados en países vecinos, como en México [53].

CONCLUSIONS

Every single haplotype has been found as unique. Power of Discrimination (PD) in Identifiler Kit is a very powerful tool for genetic identification both in the Mestizo Guatemalan and Spanish populations, but also when considering the Mayan populations. However a decrease of many of the parameters analyzed has been found, which means that a special precaution in the genetic identification of these small populations is needed.

This study represents one of the first steps to understand the populational structure of Guatemala. This is useful for forensic and /or anthropological analysis, but also it is indicative for tackling new genetic analysis (in others scientific fields such as diagnosis).

The Guatemalan population is divided in two subpopulations, one with Mayan origins (isolated and particular), and some others known as Mestizo, which are the result of the mixture between Mayan and Spanish people. This feature, allows us to sketch a genetic map of Guatemala.

5. BIBLIOGRAFÍA

- [1] Central Intelligence Agency, *The World Factbook*, (2009).
- [2] Instituto Nacional de Estadística, *XI Censo Nacional de Población y VI de Habitación*, 2002.
- [3] G.F. Simons, J. Gordon R.G., *Ethnologue*, in: Keith Brown (Ed.), *Encyclopedia of Language & Linguistics*, Elsevier, Oxford, 2006, pp. 250-253.
- [4] J.M. Butler, *Genetics and genomics of core short tandem repeat loci used in human identity testing*, *J.Forensic Sci.* 51 (2006) 253-265.
- [5] E. Meyer, P. Wiegand, B. Brinkmann, *Phenotype differences of STRs in 7 human populations*, *Int.J.Legal Med.* 107 (1995) 314-322.
- [6] E. Bosch, F. Calafell, A. Pérez-Lezaun, D. Comas, H. Izaabel, O. Akhayat, et al., *Y chromosome STR haplotypes in four populations from northwest Africa*, *Int.J.Legal Med.* 114 (2000) 36-40.
- [7] E. Bosch, F. Calafell, A. Pérez-Lezaun, J. Clarimón, D. Comas, E. Mateu, et al., *Genetic structure of north-west Africa revealed by STR analysis*, *European Journal of Human Genetics.* 8 (2000) 360-366.
- [8] A. Corella, F. Bert, A. Pérez-Pérez, M. Gené, D. Turbón, HUMTH01, HUMVWA31A, HUMCSF1PO and HUMTPOX polymorphisms in Amerindian populations living in the Beni Department of Bolivia, *Ann.Hum.Biol.* 35 (2008) 556-564.
- [9] H. Oota, B. Pakendorf, G. Weiss, A. Von Haeseler, S. Pookajorn, W. Settheetham-Ishida, et al., *Recent origin and cultural reversion of a hunter-gatherer group*, *PLoS Biology.* 3 (2005) 0536-0542.
- [10] M.A. Jobling, P. Gill, *Encoded evidence: DNA in forensic analysis*, *Nature Reviews Genetics.* 5 (2004) 739-751.
- [11] H. Liu, W.-. Yu, F. Fang, X.-. Wang, G. Yang, B. Liu, et al., *Polymorphism of microsatellite DNA vWA in random population and schizophrenic patients*, *Chinese Journal of Clinical Rehabilitation.* 9 (2005) 252-254.
- [12] J. Liu, Y. Shi, J. Tang, T. Guo, X. Li, Y. Yang, et al., *SNPs and haplotypes in the S100B gene reveal association with schizophrenia*, *Biochem.Biophys.Res.Commun.* 328 (2005) 335-341.

- [13] V.L. Martinez-Marignac, A. Valladares, E. Cameron, A. Chan, A. Perera, R. Globus-Goldberg, et al., Admixture in Mexico City: Implications for admixture mapping of Type 2 diabetes genetic risk factors, *Hum.Genet.* 120 (2007) 807-819.
- [14] K.B. Mullis, Target amplification for DNA analysis by the polymease chain reaction, *Ann.Biol.Clin.(Paris)*. 48 (1990) 579-582.
- [15] P.J. Collins, L.K. Hennessy, C.S. Leibel, R.K. Roby, D.J. Reeder, P.A. Foxall, Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: The AmpFlSTR[®] identifier[®] PCR amplification kit, *J.Forensic Sci.* 49 (2004) 1265-1277.
- [16] S. Paabo, Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification, *Proc.Natl.Acad.Sci.U.S.A.* 86 (1989) 1939-1943.
- [17] ABI PRISM[®]310 Genetic Analyzer User's Manual Part Number 903565D Printed in the USA, 12/2001 © Copyright 2001, Applied Biosystems,.
- [18] GeneScan[®] Reference Guide Chemistry Reference for the ABI PRISM[®] 310 Genetic Analyzer Part Number 4303189B Printed in the USA, 10/2000 © Copyright 2000, Applied Biosystems,.
- [19] ABI PRISM[™] Genotyper 2.0P/N 904649 © Copyright 1996, The Perkin-Elmer Corporation,.
- [20] M.B. Richards, V.A. Macaulay, H.J. Bandelt, B.C. Sykes, Phylogeography of mitochondrial DNA in Western Europe, *Ann.Hum.Genet.* 61 (1998) 251-254.
- [21] A. Tereba, Tools for Analysis of Population Statistics, Profiles in DNA. (2001).
- [22] 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.
- [23] M. Slatkin, Linkage disequilibrium in growing and stable populations, *Genetics*. 137 (1994) 331-336.
- [24] V.M. Sarich, A.C. Wilson, Generation time and genomic evolution in primates, *Science*. 179 (1973) 1144-1147.
- [25] B. Budowle, M.R. Wilson, J.A. DiZinno, C. Stauffer, M.A. Fasano, M.M. Holland, et al., Mitochondrial DNA regions HVI and HVII population data, *Forensic Sci.Int.* 103 (1999) 23-35.
- [26] B. Budowle, B. Shea, S. Niezgoda, R. Chakraborty, CODIS STR loci data from 41 sample populations, *J.Forensic Sci.* 46 (2001) 453-489.

- [27] L.A. Zhivotovsky, B.A. Malyarchuk, M.V. Derenko, M. Wozniak, T. Grzybowski, Developing STR databases on structured populations: The native South Siberian population versus the Russian population, *Forensic Science International: Genetics*. 3 (2009) e111-e116.
- [28] B. Yang, G. Wang, Y. Liu, W. Yang, Population data for the AmpF/STR[®] Identifiler[™] PCR Amplification Kit in China Han in Jilin Province, China, *Forensic Sci.Int.* 151 (2005) 293-297.
- [29] S. Maruyama, K. Minaguchi, N. Takezaki, P. Nambiar, Population data on 15 STR loci using AmpF/STR Identifiler kit in a Malay population living in and around Kuala Lumpur, Malaysia, *Leg.Med.* 10 (2008) 160-162.
- [30] W.A. Silva Jr., S.L. Bonatto, A.J. Holanda, A.K. Ribeiro-dos-Santos, B.M. Paixão, G.H. Goldman, et al., Mitochondrial genome diversity of native americans supports a single early entry of founder populations into America, *Am.J.Hum.Genet.* 71 (2002) 187-192.
- [31] F.R. Santos, N.O. Bianchi, S.D.J. Pena, Worldwide distribution of human Y-chromosome haplotypes, *Genome Res.* 6 (1996) 601-611.
- [32] M.J. Alvarez-Cubero, L.J. Martinez-Gonzalez, D. Virgen-Ponce, E.A. Delgado-Najar, G.H. Moscoso-Caloca, J.C. Alvarez, et al., Genetic variation of 15 autosomal microsatellite loci in a Nayarit population (Mexico), *Leg.Med.*
- [33] J. Zúñiga, M. Ilzarbe, V. Acunha-Alonzo, F. Rosetti, Z. Herbert, V. Romero, et al., Allele frequencies for 15 autosomal STR loci and admixture estimates in Puerto Rican Americans, *Forensic Sci.Int.* 164 (2006) 266-270.
- [34] Ribeiro Rodrigues E. M, Ferreira Palha T. J, Batista dos Santos S.E., Allele frequencies data and statistic parameters for 13 STR loci in a population of the Brazilian Amazon Region, *For Sci Int.* 168 (2007) 244-247.
- [35] R. Decorte, A. Gilissen, J.-. Cassiman, Allele frequency data for 15 STR loci (AmpF/STR[®] SGM plus[™] and AmpF/STR[®] Profiler[™]) in the Belgian population, *Int.Congr.Ser.* 1239 (2003) 219-222.
- [36] N. Haliti, M. Čarapina, M. Mašić, D. Strinović, I.M. Klarić, M. Kubat, Evaluation of population variation at 17 autosomal STR and 16 Y-STR haplotype loci in Croatians, *Forensic Science International: Genetics*. 3 (2009) e137-e138.
- [37] S. Belezá, 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.

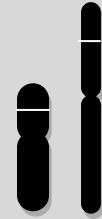
- [38] B.W. Forward, M.W. Eastman, T.B. Nyambo, R.E. Ballard, AMPFISTR® Identifiler™ STR Allele Frequencies in Tanzania, Africa, *J.Forensic Sci.* 53 (2008) 245-247.
- [39] Regueiro M., Mirabal S., L.H. , Caeiro J. L., Garcia-Bertrand R. L., Herrera R. J., Austronesian genetic signature in East African Madagascar and Polynesia, *J.Hum.Genet.* 53 (2008).
- [40] M. Nei, L. Jin, Variances of the average numbers of nucleotide substitutions within and between populations. *Mol.Biol.Evol.* 6 (1989) 290-300.
- [41] L.L. Cavalli-Sforza, M.W. Feldman, The application of molecular genetic approaches to the study of human diversity, *Nat.Genet.* 33 (2003) 267.
- [42] R.D.M. Page, TREEVIEW: an application to display phylogenetic trees on personal computers, *Comput.Appl.Biosci.* 12 (2001) 357-358.
- [43] J.E. Lygo, P.E. Johnson, D.J. Holdaway, S. Woodroffe, J.P. Whitaker, T.M. Clayton, et al., The validation of short tandem repeat (STR) loci for use in forensic casework, *Int.J.Legal Med.* 106 (1994) 302-311.
- [44] R. Chakraborty, Power of assaying inbreeding through sampling of phenotypes and mating types, *Humangenetik.* 29 (1975) 217-222.
- [45] I.W. Evett, C. Buffery, G. Willott, A guide to interpreting single locus profiles of DNA mixtures in forensic cases, *Journal of the Forensic Science Society.* 31 (1991) 41-47.
- [46] T.J. Parsons, J.A. Irwin, Questioning evidence for recombination in human mitochondrial DNA, (2000).
- [47] Espejo Arias, T. y Arias Torres, J.P, "El Corán de Cutar. Una joya del patrimonio escrito andalusí", Edición facsímil y Estudios, Ed. Junta de Andalucía -Consejería de Obras Públicas y Transportes, y Consejería de Cultura- y Fundación Tres Culturas del Mediterráneo, SEVILLA, 2009.
- [48] E. Martínez-Espin, Análisis y comparación de 16 loci STR en cromosoma Y de varios grupos poblacionales (apellidados Colón y no apellidados Colón) en la cuenca noroeste mediterranea, (2008).
- [49] S. Finnilä, I.E. Hassinen, K. Majamaa, Phylogenetic analysis of mitochondrial DNA in patients with an occipital stroke: Evaluation of mutations by using sequence data on the entire coding region, *Mutation Research - Mutation Research Genomics.* 458 (2001) 31-39.
- [50] A.C. Wilson, L.R. Maxson, V.M. Sarich, Two types of molecular evolution. Evidence from studies of interspecific hybridization, *Proc.Natl.Acad.Sci.U.S.A.* 71 (1974) 2843-2847.

- [51] P. Gill, D.J. Werrett, Exclusion of a man charged with murder by DNA fingerprinting, *Forensic Sci.Int.* 35 (1987) 145-148.
- [52] P. Gill, J.E. Lygo, S.J. Fowler, D.J. Werrett, An evaluation of DNA fingerprinting for forensic purposes, *Electrophoresis.* 8 (1987) 38-44.
- [53] 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.
- [54] E. Tamm, T. Kivisild, M. Reidla, M. Metspalu, D.G. Smith, C.J. Mulligan, et al., Beringian standstill and spread of native American founders, *PLoS ONE.* 2 (2007).
- [55] S. Wang, N. Ray, W. Rojas, M.V. Parra, G. Bedoya, C. Gallo, et al., Geographic patterns of genome admixture in latin American mestizos, *PLoS Genetics.* 4 (2008).
- [56] D.A. Merriwether, S. Huston, S. Iyengar, R. Hamman, J.M. Norris, S.M. Shetterly, et al., Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating, *Am.J.Phys.Anthropol.* 102 (1997) 153-159.
- [57] A.K.C. Ribeiro-Dos-Santos, J.M. Pereira, M.R.F. Lobato, B.M. Carvalho, J.F. Guerreiro, S.E. Batista Dos Santos, Dissimilarities in the process of formation of Curiaú, a semi-isolated Afro-Brazilian population of the Amazon region, *Am.J.Hum.Biol.* 14 (2002) 440-447.
- [58] M. Sans, T.A. Weimer, M.H.L.P. Franco, F.M. Salzano, N. Bentancor, I. Alvarez, et al., Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay, *Am.J.Phys.Anthropol.* 118 (2002) 33-44.
- [59] B. Bertoni, B. Budowle, M. Sans, S.A. Barton, R. Chakraborty, Admixture in Hispanics: Distribution of ancestral population contributions in the continental United States, *Human Biology.* 75 (2003) 1-11.
- [60] C. Bonilla, B. Bertoni, S. González, H. Cardoso, N. Brum-Zorrilla, M. Sans, Substantial Native American female contribution to the population of Tacuarembó, Uruguay, reveals past episodes of sex-biased gene flow, *Am.J.Hum.Biol.* 16 (2004) 289-297.
- [61] Rocco P.P, Morales C.G, Moraga M.V, Miquel J.F.P, Composición genética de la población chilena: distribución de polimorfismos de DNA mitocondrial en grupos originarios y en la población mixta de Santiago. *Rev.Med.Chile.* 130 (2002) 125-131.

Capítulo III: Genotipado de STRs en el Cromosoma Y.

Chapter III: Y-chromosome

STRs Genotyping.



ABSTRACT

As many American populations, the origins of Guatemala began with ancestral migrations from Asia through the Bering Strait and posterior migrations with the admixture process during the colonization period. To characterize the genetic relationships among Guatemalan inhabitants, we studied the genetic structure of five Guatemalan populations and the Spanish population using Y chromosome microsatellite markers, AmpFISTR® Yfiler PCR Amplification Kit (Applied Biosystems).

The detection of modern populations represents the identification of admixture process. Population structure defines the presence of individuals with significant differences in admixture components. These differences are important for diverse purposes: admixture and association mapping and evolutionary studies.

During the 500 years of admixture among Native Americans and Europeans, principally Spaniards, the majority of Guatemalan contemporary population was formed.

Guatemala is the largest country in Central America and its total population is approximately 13.000.000 individuals. In Guatemala, different population groups whose structure expresses that they are isolated because of the influence of geographic, linguistic and cultural factors can be found.

In this work, we analyzed the genetic data of seventeen short tandem repeat (STR) polymorphisms from the Y Chromosome (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385, DYS393, DYS391, DYS439, DYS635, DYS392, YGATAH4, DYS437, DYS438, and DYS448) in 397 unrelated donors representing a total of six population samples. These were analyzed in Kaqchiquel (n = 22), K'iche' (n = 37), Mam (n = 27), Qeqchi (n = 24), Mestizos (mixed Indigenous-Spanish). (n=115) of Guatemala and we also studied the Spanish population (n=172). The allele frequencies and the genetic diversity among the four native groups were compared, and also compared with Guatemala Mestizos. Moreover, we study the Mestizos group, which represents ≈60% of the total of Guatemalan population in the present day. Admixture among Amerindians and the Spanish has principally come to shape the present day gene pool of Guatemalan Mestizos, so we study the contribution of Mayans and Spanish samples to the Mestizo genetic diversity.

Mutation rates per generation in STRs are difficult for us to calculate the distance between ancestral populations studied, but STRs are a good tool to study modern history. The distribution, frequencies and diversity of these markers in Guatemala indicate to us how close the Mestizo population is to the Spanish population and the contribution of the parental populations to Mestizo.

1. INTRODUCCIÓN

La caracterización de marcadores genéticos presentes en el cromosoma Y permite establecer relaciones entre la población aprovechando su transmisión directa de padres a hijos. Tras realizar el análisis de los STRs autosómicos de la población de Guatemala, se advirtió que continuar el estudio con marcadores de cromosoma Y, proporcionaría nuevos datos para poder caracterizar la muestra, tanto desde el punto de vista forense, como desde el punto de vista de la antropología molecular.

La recombinación limitada de los marcadores STRs incluidos en este estudio y el hecho de que posean una herencia en bloque, que permite establecer el linaje paterno de las muestras, permitirá establecer relaciones paterno-filiales. Esto ayudará a esclarecer relaciones entre individuos cercanos, muy interesante desde el punto de vista forense, al igual que relaciones mucho más alejadas, interesante también desde el punto de vista antropológico. Todos los parientes masculinos pertenecientes a la línea paterna del individuo tendrán el mismo haplotipo cromosoma Y ya que se trasmite directamente y sin modificaciones de padres a hijos.

Como ya se dijo en capítulos anteriores, Guatemala posee una población de 13.276.517 habitantes [1]. En el último censo de Guatemala, donde la población total era de 11.237.196, la proporción de varones comprendía el 48,92% (5.496.839 varones); variando desde la menor de 44,21% en Concepción Chiquirichapa y la mayor en San Andrés de 52,32% [2].

Junto a la población de Guatemala, como se hizo en el capítulo anterior, se va a caracterizar individuos españoles, que ayudarán a comparar muchos de los parámetros, forenses y antropológicos, de la población de Guatemala.

En estos seis grupos poblacionales se estudiarán los siguientes diecisiete marcadores microsatélites de cromosoma Y: DYS456, DYS389I, DYS389II, DYS390, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, GATA H4.1, DYS437, DYS438 y DYS448 incluidos en un kit tipo multiplex: AmpF/STR® Yfiler™ (Applied Biosystems, Foster City, CA). Todos los marcadores estudiados son tetranucleótidos, excepto el DYS436 que es un trinucleótido. Se encuentran en distintas zonas del cromosoma Y y, como ya se describió en el primer capítulo, dentro de este kit existen dos marcadores de zonas duplicadas DYS389I, DYS389II y DYS385 (ver pág. 12).

Tabla III.1: Locus Y-STRs incluidos en el Kit Yfiler, unidad de repetición, número de alelos, número de repeticiones y tamaños entre los que oscilan los alelos más frecuentes.

Locus	Unidad de repetición	Nº de repeticiones	Tamaño de alelos (pb)	Nº de alelos
DYS19	(GATA) _n	10-19	174-210	10
DYS385a/b	[GAAA] _n	7-23	241-304	20
DYS388			125-143	
DYS389I	(GATA) _n (GACA) _n	9-17	235-267	9
DYS389II	(GATA) _n (GACA) _n	25-34	351-387	10
DYS390	(GATA) _n (GACA) _n	17-28	187-231	12
DYS392	(ATT) _n	6-16	233-263	11
DYS393	(GATA) _n	9-17	108-140	9
DYS434	(ATAG) ₄ (ATAC) ₁ (ATAG) ₂	8-11	110-122	4
DYS435	CTAT	15-17	218-226	3
DYS436	GTT	10-15	128-143	6
DYS437	(TCTA) _m (TCTG) _n (TCTA) ₄	13-17	180-196	5
DYS438	(TTTTC) ₁ (TTTTA) _{0,1} (TTTTC) _n	6-13	201-236	8
DYS439	(GATA) _n	9-14	236-256	6

En el caso de estos marcadores, al ser haploides, muchos de los parámetros estudiados en el capítulo anterior no se realizarán, como son todos aquellos los relacionados con la heterocigosidad. La transmisión en bloque también limita el equilibrio **HW**, ya que muchos de los marcadores podrían encontrarse ligados.

Aunque el uso más frecuente de los STRs es la genética forense [3]; también es posible realizar estudios de estimación de distancias genéticas poblacionales y la estimación del mestizaje con las bases de datos de Y-STRs [4-8].

Para poder aplicar estos marcadores, desde el punto de vista forense, es necesario conocer el poder de discriminación de cada locus en estas poblaciones. La validez del Kit es proporcional con el poder total de discriminación obtenido en cada una de las seis poblaciones.

Aun más crítico que con el estudio de cromosomas autosómicos es el estudio de cromosoma Y en poblaciones como las mayas. A causa de su herencia en bloque se debe comprobar los efectos de consanguinidad y los de deriva genética en los grupos más aislados. En primer lugar, se han de ver si existen haplotipos repetidos en las poblaciones indígenas y también se han de observar si existe pérdida de variabilidad en la población maya.

Tras estudiar estas poblaciones se puede observar la distancia genética que separa a las poblaciones tanto mayas entre sí, como la maya con la mestiza y éstas con la española. Con ello, se pretende confirmar si estos datos concuerdan con lo esperado teniendo en cuenta la geografía, factores culturales y lingüísticos,... Del mismo modo, se comprueba el aporte de los indígenas mayas y de los españoles a la población mestiza [9]. En el primer capítulo se define el mestizaje como el fruto de la migración de españoles a Guatemala tras la llegada de estos a América. Al ser mayor la proporción de varones que viajaban a Guatemala es de esperar que en este caso la población mestiza se encuentre algo más cercana a la española que en los estudios realizados con marcadores autosómicos.

Finalmente, se comparan los resultados encontrados en estos grupos con otras poblaciones. Los datos poblacionales se tomarán de la bibliografía publicada sobre poblaciones caracterizadas con estos mismos marcadores de cromosoma Y. Se buscarán las similitudes y las distancias que las separan o aproximan a otras poblaciones, lo que ayudará a realizar, como en el caso de STRs autosómicos, estudios de evolución humana, mapas de asociación, identificación geográfica e incluso la predicción de riesgos de enfermedades asociadas a una población [10-15].

2. MATERIAL Y MÉTODOS

2.1 LA MUESTRA

Como se explica en el capítulo anterior, para toma de la muestra se seleccionaron indistintamente hombres y mujeres, un total de 400 individuos en Guatemala y 200 españoles, no relacionados entre sí. De los 400 individuos tomados en Guatemala 225 muestras eran varones, 115 procedían de mestizos, y 110 de los grupos mayas: Kaqchiquel 22, K'iche' 37, Mam 27 y Q'eqchi' 24. De los españoles, 174 muestras eran varones.

La toma de la muestra se detalla en el capítulo II.

2.2 ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA

La toma de las muestras y de los controles, su análisis y registro siguieron las recomendaciones de la ISFG, tanto intralaboratorio como interlaboratorio.

2.2.1 Extracción

La purificación del ADN, al llegar en soporte FTA, se realizó siguiendo los protocolos del fabricante (Whatman, Clifton, NJ, USA), adaptados a este tipo de muestra en el laboratorio de Identificación Genética, protocolo en el ANEXO 1.

2.2.2 Amplificación

Se tomó como ADN molde un punch (1.2mm Ø) de papel FTA. Se amplificaron 17 zonas de interés, incluidas en el kit validado Yfiler® (Applied Biosystems, Foster City, CA) [16]. La amplificación se realizó en un termociclador, Biometra (T3 Thermocycler).

El Kit AmpFlSTR® Yfiler TM [17] permite el análisis simultáneo de 17 loci del cromosoma Y. Uno de los primers de cada pareja está marcado con un

fluorocromo con un color (verde, amarillo, azul y rojo) para ser analizados simultáneamente en el aparato de electroforesis capilar. Los fluorocromos que se usan para cada marcador son los siguientes:

- 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 III.2: Componentes del kit de amplificación:

Reactivo	Componentes	Volumen
AmpFISTR® Yfiler® Reaction Mix	MgCl ₂ , dATP, dGTP, dCTP, and dTTP, suero de albúmina bovina y 0.05% NaO ₃	4,05µl
AmpFISTR® Yfiler® Primer Set	Parejas de primers marcados para los 17 loci Y-STR	2,2µl
AmpliAq Gold® DNA Polymerase	Enzima Polimerasa 5U/µL	1,08µl
AmpFISTR® Control DNA 9947A	ADN humano femenino 10ng/ µL	4 µl
AmpFISTR® Control DNA 007	ADN humano masculino 0.10 ng/µL	4 µl

La cantidad de ADN por muestra osciló entre 0.25-0.5ng y el volumen final de reacción fue de 10µl.

En las amplificaciones, además de las muestras, se amplificaron sistemáticamente dos controles: uno positivo y otro negativo [18].

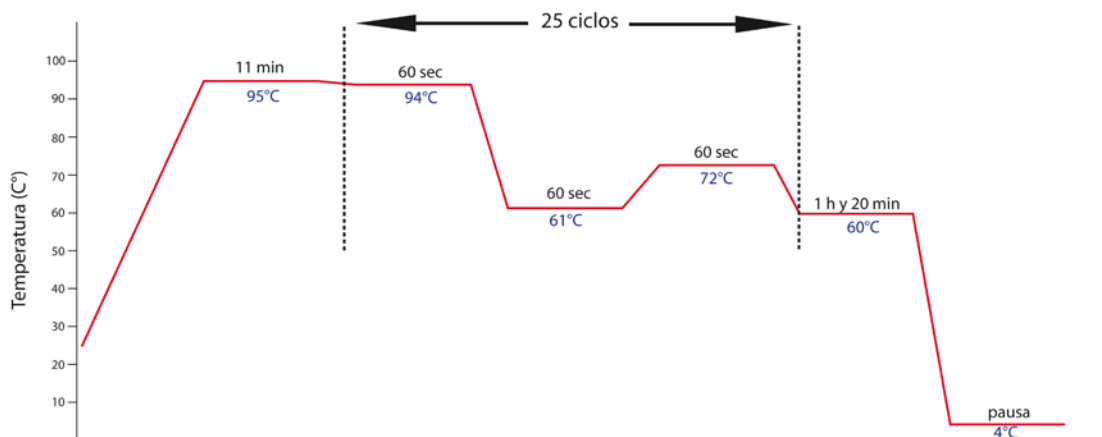


Figura III.1: Número de ciclos y temperaturas de desnaturalización, hibridación y extensión utilizadas para la amplificación con el kit Yfiler.

2.2.3 Electroforesis Capilar

Las muestras amplificadas con el kit fueron visualizadas en un analizador genético ABI Prism 310 de Perkin Elmer. Las condiciones fueron las mismas que en capítulo anterior (Ver generalidades en capítulo II).

- SOFTWARE

En este caso el software utilizado fue GeneMapper® ID-X v1.1 Este software ofrece mayor fiabilidad a la hora de genotipar muestras de tipo forense comparado con los softwares mencionados anteriormente u otros softwares actuales con otras características.

Con respecto a lo mencionado en el capítulo I, este mismo software permite analizar el tamaño de los alelos, diferenciar los colores de los fluorocromos y asignar los alelos a partir de un ladder de referencia. Todo esto es posible en un mismo paso.

Además de esto, este nuevo software es una poderosa herramienta para la identificación genética, disminuyendo el tiempo de análisis de las muestras así como dando mayor fiabilidad y reproducibilidad a los resultados [19]. Todo esto es gracias a la introducción en el software de los denominados *expert system* (sistemas expertos), sistemas que integran parámetros de revisión y certificación de la calidad de los resultados obtenidos [20].

Esta nueva herramienta bloquea la manipulación por parte del usuario, no permite realizar cambios o modificaciones de una serie de parámetros establecidos, sin registrar el usuario, el cambio, etc. En investigación, donde se realizan muchas variaciones en los módulos de análisis, puede resultar complejo, pero tiene la ventaja de que siempre se podrá reproducir un ensayo en las mismas condiciones.

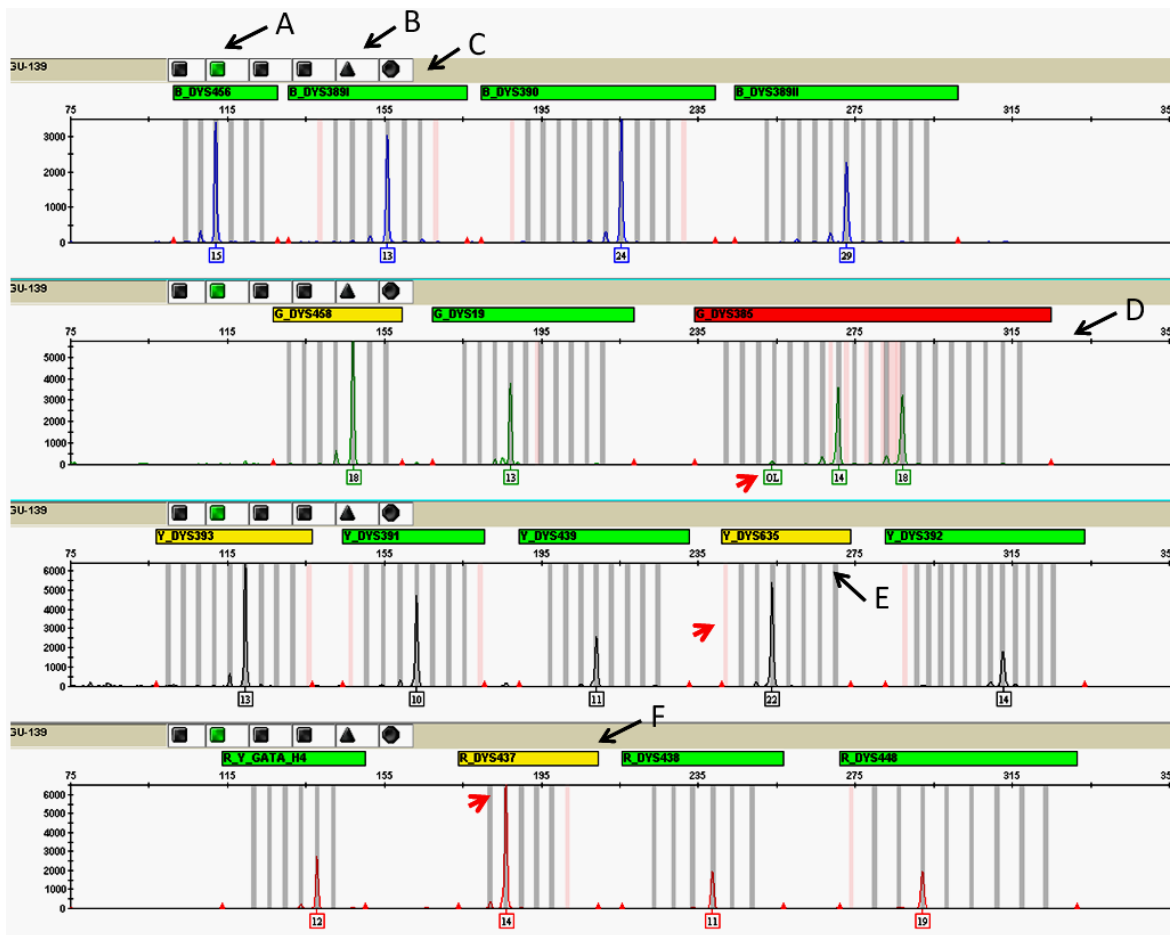


Figura III.2: Gráfico generado por el programa GeneMapper® ID-X v1.1. En él se puede apreciar las señales de alerta producidas por los *expert system*. La revisión de los parámetros de calidad advierten mediante señales de colores: A: señales de aviso con respecto al análisis y comparación con el ladder. B: señales sobre la electroforesis y estándar de tamaño. C, D, E y F: señales sobre los resultados del perfil obtenido; en verde donde todo está correcto, en amarillo alerta sobre posibles errores por alteraciones en las medidas de intensidad y en rojo donde los errores son tales que el analista debe revisar y/o modificar el resultado obtenido.

2.3 ANÁLISIS ESTADÍSTICO DE LA MUESTRA

- El cálculo de las frecuencias de las repeticiones encontradas en cada una de las poblaciones se realizó mediante una hoja de cálculo (Microsoft Office Excel 2007) siendo el programa base de ésta, la hoja de cálculo para STRs de uso forense denominada PowerStatsV12, (Promega® Corporation) [21].
- 23-Haplogroup Beta Program (On-line Software): [22,23]. Mediante este programa 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. 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:

<http://www.hprg.com/hapest5/>. Hay que tener cautela con el uso de este tipo de programas ya que al ser una inferencia a través de un cálculo estadístico presenta un porcentaje con un cierto error [24,25].

- Network: [26] (DNA Alignment © Fluxus Technology) Software de alineamiento de ADN y proteínas. El análisis producido genera redes filogenéticas que proporcionan una estimación de cómo han transcurrido los sucesos. Buscando un ancestro y desde este antepasado común se originarán el resto de genotipos estudiados.
- Software Arlequin 3.1: [27] Comprobación de los datos obtenidos mediante hoja de cálculo; comparación de la distribución de los genotipos observados en cada grupo mediante el estadístico F; las distancias genéticas de Reynolds [28] y la transformación linear del Fst de Slatkin [29], el test exacto. Comparación con bibliografía publicada (siempre que se incluyeran en la publicación los perfiles genéticos).

- Phylip v3.66: [30,31] (the PHYLogeny Inference Package, software gratuito disponible en:

<http://evolution.genetics.washington.edu/phylip.html> . De este paquete de software informático se utilizaron las siguientes aplicaciones:

Gendist → Este programa calcula las distancias más frecuentes en antropología molecular a partir de un conjunto de frecuencias de los genes en diferentes poblaciones.

Neighbor → Este programa incluye los cálculos de Nei and Saitou (1987) así como el método para realizar clúster UPGMA. Neighbor realiza distintos conjuntos con los linajes encontrados dependiendo de su distancia genética, construyendo así ramas de distinta longitud que representen la historia filogenética de los linajes encontrados. [32]

- Stadistica v8: StatSoft® (StatSoft ibérica) distribuye este software de análisis estadístico que posee un conjunto de herramientas para análisis, gestión y visualización de bases de datos. Sus técnicas incluyen una amplia selección de modelación predictiva, agrupamientos (clustering) y herramientas exploratorias.
- SPSS v15.0: Software de análisis estadístico, que al igual que el programa anterior posee herramientas para el análisis multivariante, así como herramientas de modelación y agrupamientos. Igualmente permite de una forma sencilla la visualización de bases de datos y los resultados generados.
- ADMIX.PAS, [33] Cálculo del mestizaje a partir de frecuencias mediante una estimación por mínimos cuadrados (WLS).
- ADMIX95, [34,35]. (*Gene identity method*) Recopilación de algoritmos matemáticos que a partir de las frecuencias obtenidas en cada población estima la proporción de población parental que ha heredado una población híbrida.

- STRUCTURE 2.3.1 [36-39]. Con él se han estimado las proporciones de individuos ancestros en cada una de las poblaciones (<http://pritch.bsd.uchicago.edu/structure.html>). Para realizar estos cálculos se hizo con un número de interacciones altas, debido a la similitud de las muestras, periodos “burn-in” de 25.000 interacciones, seguidos de 15.000. Los cálculos realizados se hicieron siguiendo los dos modelos de predicción: partiendo de la base de que la población estudiada es una mezcla de varias poblaciones, y un segundo modelo en el que se supone que es una población en equilibrio sin mezclas ni grupos poblacionales entre sus individuos.

El estudio de distancias con respecto a otras poblaciones se realizó con distintas poblaciones tomadas de la bibliografía ya que se disponía de las frecuencias de los marcadores estudiados, así como de los perfiles genéticos. Las poblaciones escogidas para realizar estos estudios de comparación se muestran en la tabla III.3. Las frecuencias obtenidas para los marcadores de cada una de las poblaciones se estudian en primer lugar mediante el programa Statistica v8, con el que se realizó un análisis de correspondencias. Este programa genera un gráfico sobre la distribución de los marcadores y de las poblaciones dependiendo de las frecuencias obtenidas. De manera paralela, se toman los perfiles genéticos de cada población y se calculan las distancias genéticas mediante Arlequin v3.1. Las distancias obtenidas posteriormente son sometidas a un análisis multidimensional mediante el programa SPSS v15.0 para poder representar más fácilmente los datos obtenidos con anterioridad.

Tabla III.3: Poblaciones, abreviatura, situación geográfica, marcadores estudiados y autor del estudio publicado.

POBLACION	Abreviatura	PAÍS	Y-STRs	N	Referencias
Mestizo	Mestizo	GUATEMALA	17	115	Estudio actual
Maya	Maya	GUATEMALA	17	110	Estudio actual
Colla	CollaAR	ARGENTINA	17	29	Toscanini et al. 2008 [40]
Toba	TobaAR	ARGENTINA	17	31	Toscanini et al. 2008 [40]
Bolivia	BolivBO	BOLIVIA	15	87	Tirando et al. 2009 [41]
Cartagena	CartaCO	COLOMBIA	16	173	Builes et al. 2006 [42]
Rio de Janeiro	RioJBR	BRASIL	17	121	Silva et al. 2006 [43]
Rio Grande do Sul	RioGBR	BRASIL	17	255	Schwengber et al. 2009 [44]
Mestizo	MestEC	ECUADOR	12	58	Gonzalez-Andrade et al 2009 [45]
Kichwe	KichEC	ECUADOR	12	63	Gonzalez-Andrade et al 2009 [45]
Descendientes Africanos	AfrUSA	USA	27	258	Butler et al. 2008 [46]
Descendientes Europeos	EuroUSA	USA	27	257	Butler et al. 2008 [46]
Descendientes Hispanos	HispUSA	USA	27	139	Butler et al. 2008 [46]
Algeria	ArgAF	Algeria	17	545	Robino et al. 2008 [47]
Guinea	GuinAF	Guinea Ecuatorial	16	100	Arroyo-Pardo et al. 2005 [48]
Ovambo	OvamAF	Namibia	27	52	Fijuhara et al. 2009 [49]
Japón	JapAS	Japón	16	1058	Budowle et al. 2009 [50]
Coreanos	KorAS	Corea	22	200	Lee et al. 2007 [51]
España	SpainEU	España	17	153	Gomez-Martín et al. 2009 [52]
Portugal	PortuEU	Portugal	17	178	Bento et al. 2009 [53]
Italia	ItalyEU	Italia	15	154	Turrina et al. 2006 [54]
Croacia	CroacEU	Croacia	16	200	Haliti et al. 2009 [55]
Serbia	SerbEU	Serbia	17	185	Veselinovic et al. 2008 [56]
Austria	AustrEU	Austria	17	135	Berger et al. 2005 [57]

3. RESULTADOS Y DISCUSIÓN

3.1 ESTUDIO HAPLOTÍPICO

Se caracterizaron 397 individuos obteniendo los haplotipos para los 17 STRs en cromosoma Y. Todos los haplotipos encontrados en el conjunto son únicos, lo que implica una diversidad haplotípica de 1 tanto en el conjunto, como para cada uno de los 6 grupos (españoles, mestizos, Kaqchiquel, K'iche', Mam y Q'eqchi'). Los resultados completos se muestran en la tabla III.4.

Tabla III.4: Estudio haplotípico de las poblaciones.

POBLACIÓN	n	Nº Haplotipos encontrados	p, Probabilidad de coincidencia.	Diversidad Haplotípica $(n/n-1)(1-\sum p_i^2)$	Capacidad de discriminación $(n^\circ \text{ haplotipos}/ n) \times 100$
Espanoles	172	172	0,005814	1	100%
Mestizos	115	115	0,008696	1	100%
Kaqchiquel	22	22	0,045454	1	100%
K'iche'	37	37	0,027027	1	100%
Mam	24	24	0,041667	1	100%
Q'eqchi'	27	27	0,037037	1	100%

Es frecuente, cuando se estudian grupos poblacionales con una identidad cultural afianzada en un pequeño marco geográfico, como son algunas de las poblaciones de nuestro estudio (Kaqchiquel, K'iche', Mam y Q'eqchi'); que aparezcan más de un individuo con igual información genética. El éxito de discriminación del Kit Yfiler es la inclusión de gran cantidad de marcadores y muchos de ellos con una tasa de mutación relativamente alta, como son el DYS635, DYS439, DYS456 y el DYS458. Si se elimina del conjunto total estos marcadores, aparecen con frecuencia individuos con un mismo haplotipo.

3.2 CALCULO DE HAPLOGRUPOS

Los haplogrupos determinados mediante el 23-Haplogroup Beta Program fueron los que se presentan en la tabla III.5 y en las figuras III.3 y III.4.

Tabla III.5: Frecuencia de los haplogrupos estimados en las distintas poblaciones.

Grupo	Q	R1b	E1b1b	J2a	I	T	G2a	J1	J2b	L	R1a	N
Q'echi'	0,7630	0,1250	0,1093	0,0000	0,0000	0,0013	0,0000	0,0001	0,0000	0,0004	0,0009	0,0000
Mam	0,8654	0,0000	0,0000	0,0741	0,0000	0,0585	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000
K'iche'	0,9490	0,0000	0,0031	0,0000	0,0002	0,0110	0,0000	0,0296	0,0000	0,0028	0,0011	0,0032
Kaqchiquel	0,7595	0,1364	0,0915	0,0000	0,0001	0,0096	0,0000	0,0019	0,0000	0,0010	0,0000	0,0000
Mestizo	0,4489	0,2696	0,0790	0,0441	0,0330	0,0249	0,0344	0,0089	0,0261	0,0107	0,0019	0,0001
Español	0,0000	0,6339	0,0843	0,0582	0,1058	0,0131	0,0632	0,0272	0,0073	0,0000	0,0066	0,0000
Maya	0,8342	0,0654	0,0510	0,0185	0,0001	0,0201	0,0000	0,0079	0,0000	0,0010	0,0005	0,0008

Tras estudiar la población maya, el haplogrupo mayoritario es el haplogrupo Q, apareciendo en la población K'iche' hasta en un 95% de las muestras y en todos los casos superior al 76% de la población Kaqchiquel. Estos resultados son los esperados sabiendo que el haplogrupo Q es uno de los mayoritarios en Sudamérica, donde alcanza el 30 % del total de la población. El haplogrupo Q tiene un origen siberiano y se distribuyó por toda América, encontrándose presente en las poblaciones nativas. Por tanto, los cálculos realizados por el 23-Haplogroup Beta Program concuerdan con la historia y evolución de estos grupos poblacionales.

Aparte de esto, en los cuatro grupos mayas se observa que aparecen otros como el R1b y el E1b1b (tanto en la población Q'echi', como en la población Kaqchiquel). El haplogrupo R1b y el E1b1b son típicos en las poblaciones mediterráneas. El primero de ellos tiene una distribución europea y asiática y el segundo aparece en África y se distribuye por el sur de Europa. En las poblaciones Mam y K'iche' aparecen los haplogrupos J2a, J1 y T. El haplogrupo T es uno de los haplogrupos ancestros. Aparece en Mesopotamia y se distribuye por el sur de Europa. De él aparecerán posteriormente el Q y el R. El J es también un haplogrupo que aparece en Mesopotamia, Anatolia y norte de África. La aparición de los haplogrupos R1b y E1b1b en estas poblaciones se puede explicar por el flujo entre éstas y la población mestiza en los últimos 500 años. Por ello, es normal su aparición en el grupo Q'echi', un grupo poblacional disperso en un amplio territorio. La aparición de los haplogrupos J2a, J1 y T es más difícil de explicar. Aunque podría ser prehispánica, parece más probable que la llegada de estos sea por la misma razón que el R1b y el

E1b1b, puesto que hoy en día no se encuentran en gran número (J2a, J1 y T) en poblaciones del norte de Asia ni en Siberianas.

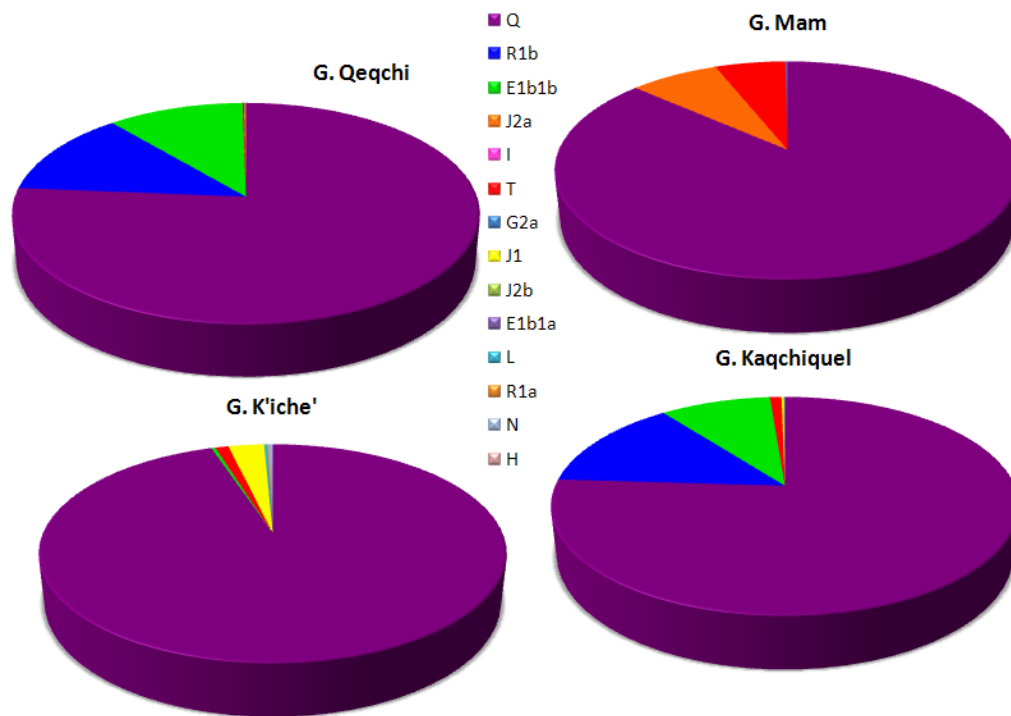


Figura III.3: Gráficos de sectores de las frecuencias de los haplogrupos en las poblaciones Maya. Cada gráfico circular representa los haplogrupos obtenidos en un grupo poblacional y cada sector el haplogrupo que señala la leyenda.

Si se observan los haplogrupos obtenidos en la población mestiza de Guatemala y en la española, se ve que hay mayor variabilidad que la que se encuentra entre las poblaciones mayas. En la población española, el haplogrupo mayoritario es el R1b, muy frecuente en la población mediterránea y el más frecuente en España y Portugal. Tras éste, en la población española, aparecen el I (norte europeo, aunque en España es frecuente en Galicia, Cataluña, Asturias, Cantabria, País Vasco y Aragón), E1b1b (europeo), J2a (europeo y norteafricano) y el G2a (euroasiático). En la población mestiza de Guatemala aparece mayoritariamente en un 45% el haplogrupo Q, seguido de un 27% del R1b. El resto de los haplogrupos que aparecen en la población mestiza son los citados anteriormente en los grupos español y maya, explicando perfectamente como el grupo mestizo aparece de la mezcla de los haplogrupos nativos y de los provenientes del sureste de Europa. Así después del Q y del R1b aparecen el

E1b1b, J2a, G2a e I. Todos ellos muy presentes en la población española y los dos primeros también en la población maya (aunque seguramente la aparición de estos en la comunidad maya se deba al flujo con los mestizos).

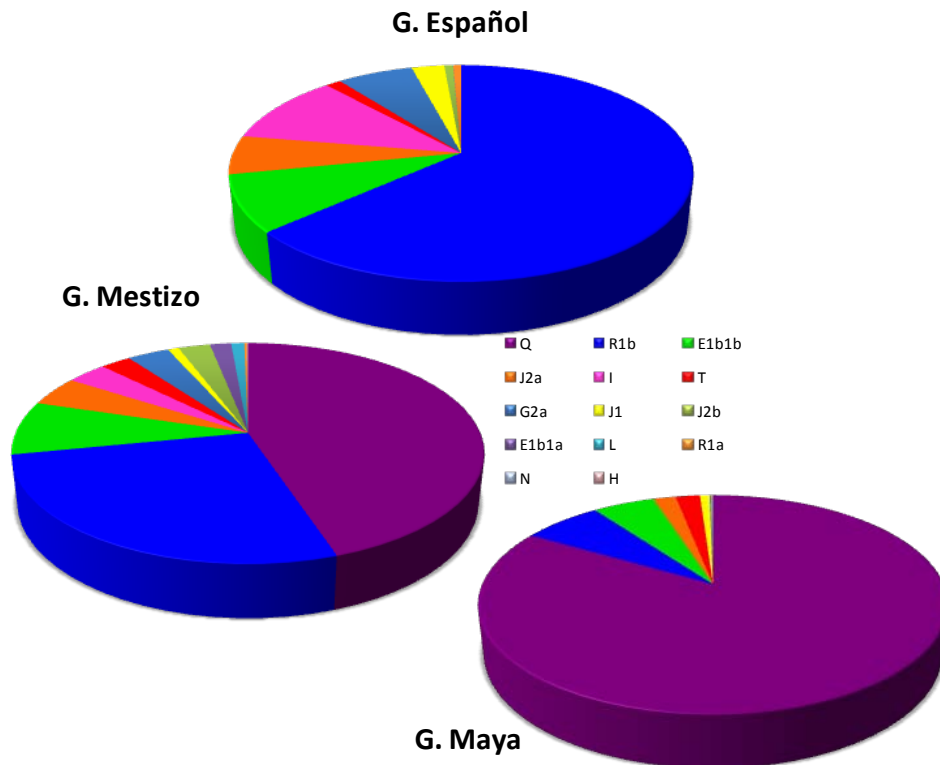


Figura III.4: Gráficos de sectores de haplogrupos en las poblaciones estudiadas. El gráfico de la población maya es el resultado del conjunto de los cuatro subgrupos poblacionales.

3.2.1 Estudio del Haplogrupo Q en Guatemala

El haplogrupo que va a dar mayor información sobre la historia y evolución del pueblo de Guatemala es el haplogrupo Q. Así se podrá o puede saber cuáles pueden ser los ancestros de esta población y qué individuos son los que poseen haplotipos más antiguos.

Para realizar los estudios sobre este haplogrupo, se tomaron los individuos que poseían un haplogrupo Q de todos los individuos de Guatemala. Se hizo con ellos una red Network para buscar aquellos haplotipos más antiguos.

De igual modo, se puede comprobar si las distintas poblaciones estudiadas difieren lo suficiente, como para separarse en ramas, o si por lo contrario aparecen distribuidas uniformemente en el diagrama de redes.

Como ya se ha citado anteriormente, el Kit utilizado para analizar estas muestras posee 17 marcadores. En estos varía mucho la tasa de mutación [58], hay una diferencia de hasta 25 veces mayor en unos marcadores con respecto a otros. Por esta razón, no se tomaron todos los marcadores que ofrece el Kit para realizar este estudio, sino que fueron aquellos con menor tasa los que se utilizaron. Así se ve que algunos individuos ahora sí comparten un haplotipo común. Los marcadores utilizados en este estudio fueron: DYS390, DYS385a, DYS393, DYS392, DYS437, DYS438 y DYS448, todos ellos con una tasa de mutación menor a 0.0020.

Los resultados de los estudios Network se pueden observar en las figuras (III.5 y III.6). En primer lugar se realizó un estudio, con los individuos mayas. Posteriormente se realizó un estudio incluyendo todos los individuos de Guatemala.

- Haplogrupo Q en la población Maya

Dentro del grupo maya, como ya se ha visto, existe una variación en cuanto a la proporción de individuos con haplogrupo Q desde el grupo K'iche', con un 94%, hasta el Kaqchiquel donde el número de individuos con haplogrupo Q desciende hasta el 75% (n=16). Cuando se observan por tanto los resultados, se tiene que tener en cuenta que el grupo K'iche' es el que posee un número mayor de individuos (n=37) y además es el que posee mayor número de individuos con haplogrupo Q. Los Mam representados son 22 y los Q'echi' 18.

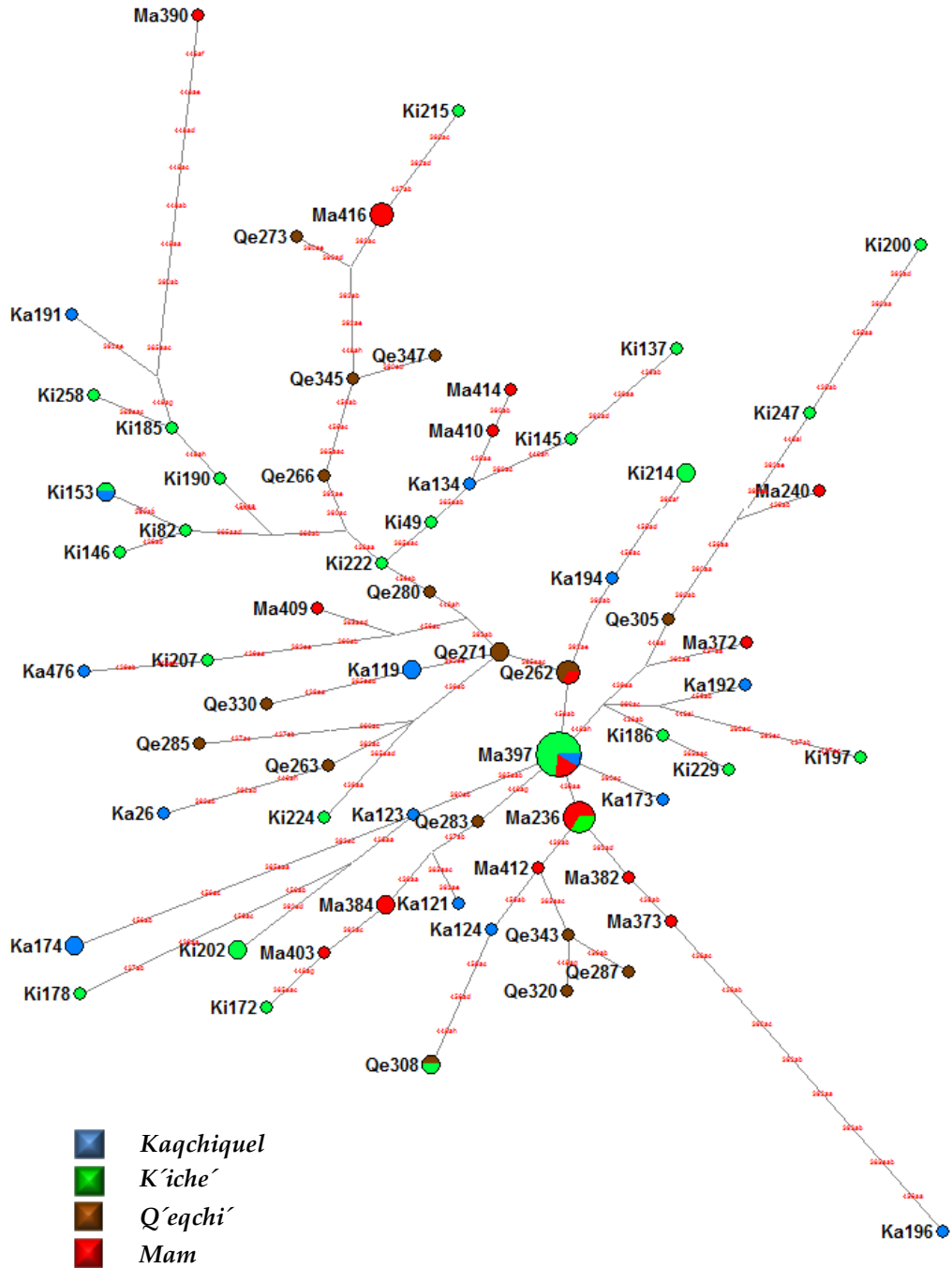


Figura III.5: Gráfico de redes calculado mediante Network tomando como base los individuos Mayas.

En el centro de la imagen, se encuentra el comienzo de la formación de la red, con tres haplotipos: Q271 (n=2), Q262 (n=3) y Ma397 (n=11). A partir de estos individuos, por acúmulo de variaciones, van apareciendo el resto de los individuos. Encontrando los más alejados los individuos Ka196 y Ma390.

En la figura III. 5 se puede ver como la población maya, tiene una distribución uniforme, sólo se puede destacar el caso de la población Kaqchiquel, donde la mayoría de sus representantes están en los extremos de las ramas. Pero no se puede decir que diverjan del resto, ya que no comparten ramas comunes.

- Haplogrupo Q en la población de Guatemala

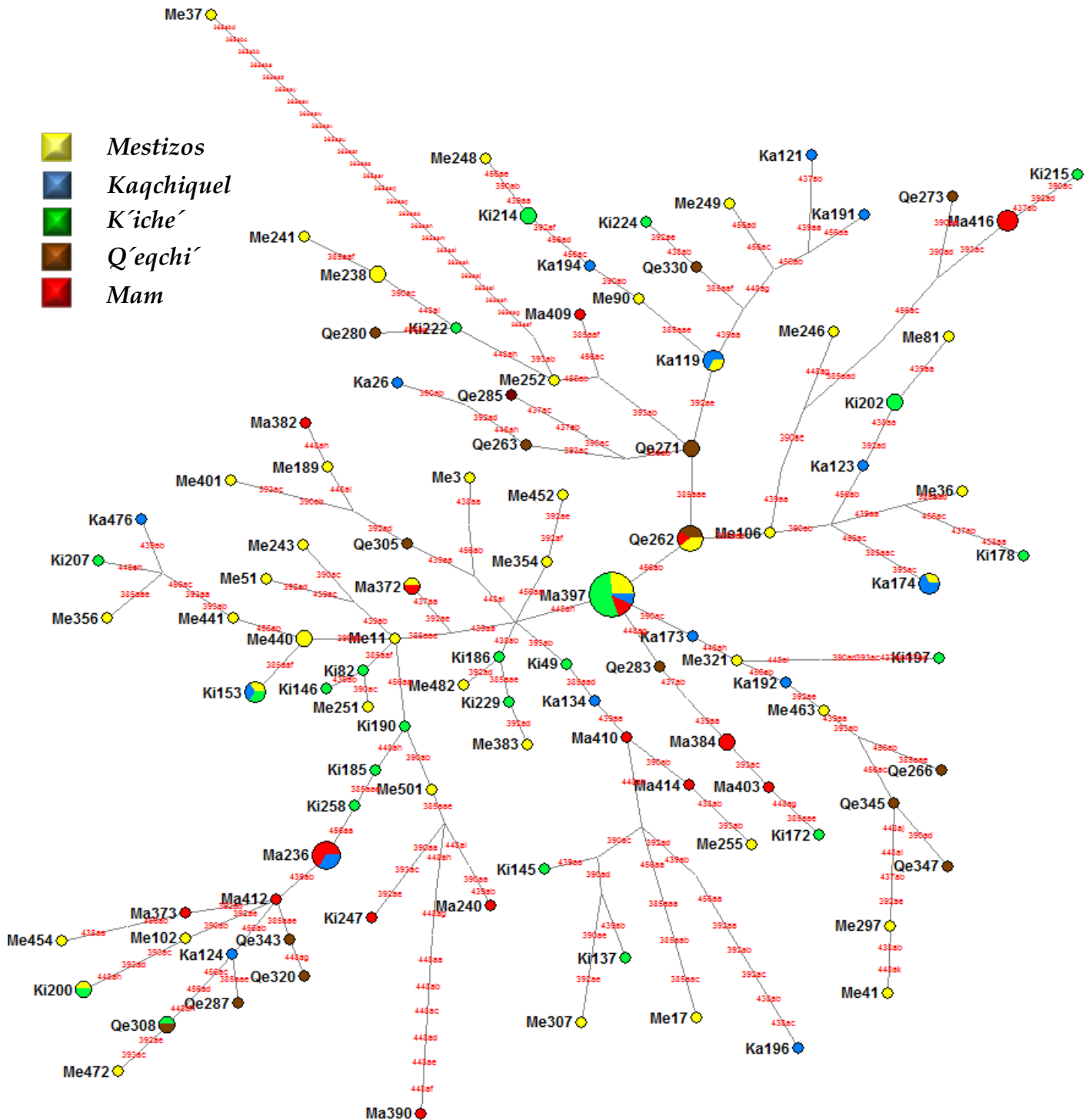


Figura III.6: Árbol Network, de los haplotipos Q de la población general de Guatemala.

Al incluir los mestizos en el análisis networks, figura III.6, se ve que el estudio se hace más complejo, pero sigue teniendo una distribución similar y uniforme entre las poblaciones. El grupo mestizo, aporta a este estudio 51 individuos, (un 44,9% de este grupo).

El grupo central, sigue formado por los haplotipos Qe271, el Qe262, que ahora cuenta con dos individuos mestizos que se suman a los dos Q'eqchi' y al Mam que tenía en la figura III.5. También dentro de este grupo, de donde nace la red, se encuentra al haplotipo Ma397, ahora formado por quince individuos, con cuatro individuos mestizos más.

Al ver la figura, III.6, en el exterior no se observan los individuos Kaqchiquel, como ocurría en el estudio anterior, ahora son los mestizos, los que en su gran mayoría están en las posiciones más alejadas de la red. Siendo el individuo Me37, el más alejado al resto. No existe una divergencia señalada del grupo mestizo, éste se puede ver también en posiciones centrales, y en ramas. Pero si se distinguen muchas posiciones, como la Me17, Me41, Me307, Me401,... que se encuentran distanciados, ocupando muchas de las posiciones finales, cuando su representación en este estudio no supera el 35%.

El grupo mestizo, no parece tener una mayor proximidad a ninguna de las poblaciones mayas estudiadas. El origen de estos no parece provenir de una única población Maya, sino más bien, de la mezcla homogénea de los colonizadores españoles con los distintos grupos mayas que habitaban en Guatemala.

3.3 DISTANCIAS GENÉTICAS

Para comprobar la distancia genética entre nuestras poblaciones se realizó un cálculo de las distancias de Reynolds [28] y transformación lineal del F_{st} [29]. Mediante el programa Arlequin, se tomaron todos los perfiles obtenidos y se calcularon la distancia entre estos. Las distancias obtenidas a partir de los perfiles ponían de manifiesto la similitud de la población española con la población mestiza (la distancia entre españoles y mestizos era del orden de

1×10^{-4} ; mientras que el resto de las distancias se aproximaban a 5×10^{-3}). Tras esto se realizó una representación de la matriz de distancias obtenida mediante el nmMDS usando el software SPSS.

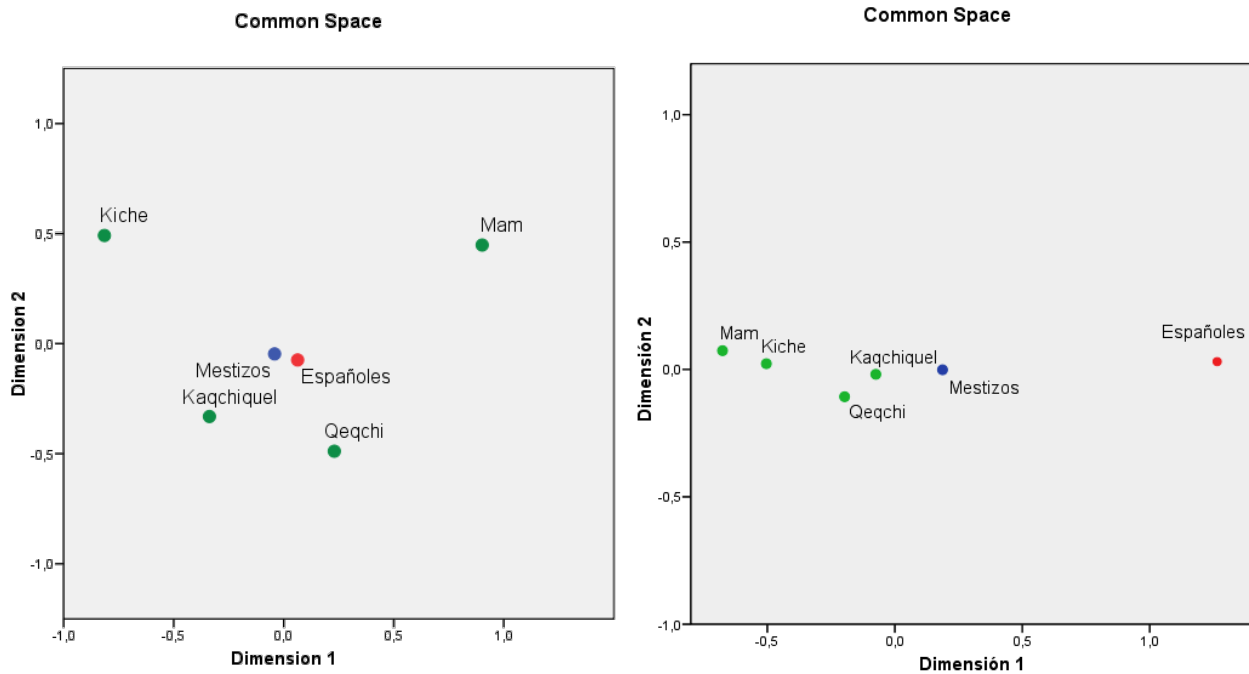


Figura III.7: Gráfico de distancias genéticas de Reynolds, representadas realizando un cálculo MDS bidimensional. A la izquierda, representación utilizando las distancias obtenidas con 17-Y-STRs. A la derecha representación obtenida con 13 marcadores Y-STRs. Stress= 0.0986 y 0.00102.

Al ver las distancias obtenidas y su representación tras hacer el MDS, se observó la pequeña distancia que existía entre todas las poblaciones y en especial entre los españoles y los mestizos. Tras esto se realizaron de nuevo los mismos cálculos utilizando 4 marcadores menos (eliminando los datos obtenidos de los marcadores con mayor tasa de mutación) DYS635, DYS439, DYS456 y DYS458. Al eliminar estos marcadores, el resultado obtenido fue bastante distinto. La distancia entre las poblaciones de Guatemala y la española se hicieron mayores y la población mestiza tomó una posición intermedia entre la población española y las poblaciones mayas (ver figura III.7). Estos nuevos resultados revelaron la necesidad de evaluar los marcadores Y-STRs utilizados para realizar los cálculos de distancias entre nuestras poblaciones, al igual que para evaluar distancias con otras poblaciones posteriormente.

3.4 ESTUDIO DE FRECUENCIAS ALÉLICAS

Una vez caracterizados los haplotipos y estudiadas las distancias entre nuestras poblaciones, se calcularon las frecuencias de cada repetición y el poder de discriminación (PD) de cada marcador para las poblaciones estudiadas. Para realizar estas tablas y teniendo en cuenta los datos anteriores, como los generados al hacer los estudios de aSTRs, se calcularon tomando tres grupos poblacionales (mestizos y mayas de Guatemala y españoles). Los datos obtenidos se muestran en las tablas III.6, III.7 y III.8. En la tabla III.7 se exponen (ponen de manifiesto) los datos de la población Maya. La población Maya engloba a los cuatro grupos poblacionales, Kaqchiquel, K'iche', Mam y Q'eqchi'. En este caso, al no ser muy alto el número de muestras de cada población ($n = 22$, $n = 37$, $n = 27$ y $n = 24$), todos estos grupos se han insertado dentro de un mismo grupo, Maya ($n=110$), para aumentar la validez de los resultados del estudio poblacional. Como se menciona anteriormente esta agrupación se realizó teniendo en cuenta todos los datos anteriores así como que las frecuencias poblacionales obtenidas en cada uno de los grupos no eran distintas estadísticamente significativas.

Tabla III.6: Frecuencias de las repeticiones identificadas en los marcadores contenidos en el kit Yfiler en la población Mestiza de Guatemala (n=115).

Allele	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS393	DYS391	DYS439	DYS635	DYS392	YGATAH4	DYS437	DYS438	DYS448	Allele	DYS385A/B
8														0,00870		10,14	0,00870
9							0,04348							0,10435		11,11	0,00870
10							0,66087	0,02609			0,00870	0,03478		0,21739		11,12	0,00870
11							0,00870	0,29565	0,45217		0,23478	0,30435		0,39130		11,13.2	0,00870
11.2																11,14	0,14783
12		0,19130			0,02609	0,15652			0,38261			0,60000		0,27826		11,15	0,04348
13	0,02609	0,48696			0,40000	0,73913			0,12174		0,33913	0,06087	0,00870			11,16	0,01739
13.2																11,18	0,00870
<14					0,00870											11.2,12	0,00870
14	0,06957	0,31304			0,02609	0,40870	0,08696		0,00870		0,31304		0,65217			12,14	0,03478
15	0,50435	0,00870			0,21739	0,11304	0,00870		0,00870		0,08696		0,23478			12,16	0,01739
16	0,29565				0,22609	0,03478					0,01739		0,10435			13,14	0,02609
17	0,06957				0,31304	0,01739										13,15	0,03478
18	0,02609				0,15652										0,095652	13,16	0,02609
18.2					0,00870											13,17	0,02609
>18	0,00870															13,18	0,02609
19					0,04348					0,00870						13,20	0,00870
20										0,00870					0,313043	13.2,18	0,00870
21			0,01739							0,13913					0,130435	14,14	0,01739
22			0,05217							0,35652					0,034783	14,15	0,01739
23			0,26957							0,37391					0,008696	14,16	0,07826
24			0,47826							0,10435						14,17	0,06957
25			0,16522							0,00870						14,18	0,07826
26			0,00870													14,19	0,01739
27			0,00870													14,20	0,00870
28				0,10435												15,15	0,01739
29				0,35652												15,16	0,07826
30				0,37391												15,17	0,02609
31				0,12174												15,18	0,03478
32				0,02609												16,16	0,02609
33				0,01739												16,17	0,03478
PD	0,64711	0,62820	0,66813	0,70639	0,77641	0,65800	0,42147	0,47395	0,63350	0,70261	0,72393	0,54246	0,50858	0,71123	0,70034		0,93899

Tabla III.7: Frecuencias de las repeticiones identificadas en los marcadores contenidos en el kit Yfiler en la población Maya de Guatemala (n=110).

Allele	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS393	DYS391	DYS439	DYS635	DYS392	YGATAH4	DYS437	DYS438	DYS448	Allele	DYS385A/B
9								0,06364						0,02727		11,13	0,01818
10		0,01818						0,77273	0,03636		0,00909			0,09091		11,14	0,02727
11		0,00909					0,01818	0,16364	0,41818		0,05455	0,15455		0,76364		12,12	0,00909
12		0,02727				0,00909	0,15455		0,39091			0,70909		0,10000		12,16	0,00909
13		0,70909				0,59091	0,718182		0,154545			0,136364	0,1181818	0,00909	0,018182	13,14	0,00909
13.2						0,02727										13,15	0,05455
>13												0,01818				13,16	0,00909
<14					0,00909											13,17	0,05455
14	0,08182	0,20909			0,00909	0,3	0,10000				0,66364		0,854545			13,18	0,01818
15	0,53636	0,01818			0,17273	0,05455	0,00909				0,11818		0,11818			13,19	0,01818
16	0,22727				0,34545	0,01818					0,01818		0,01818			14,14	0,03636
17	0,11818	0,00909			0,28182											14,15	0,00909
17.2					0,00909											14,16	0,14545
18	0,03636				0,11818										0,072727	14,17	0,14545
19					0,03636										0,581818	14,18	0,13636
2					0,018182										0,318182	14,19	0,00909
21										0,04545					0,027273	14,20	0,00909
21.3										0,00909						15,15	0,04545
22			0,02727							0,65455						15,16	0,06364
23			0,21818							0,23636						15,17	0,04545
24			0,60909							0,04545						15,18	0,04545
25			0,10909							0,00909						15,19	0,00909
26			0,02727	0,00909												16,16	0,01818
27			0,00909	0,00909												16,17	0,02727
28				0,02727												16,18	0,01818
29				0,27273												17,17	0,00909
30				0,34545													
31				0,22727													
32				0,09091													
33				0,01818													
PD	0,63868	0,45190	0,56793	0,74512	0,75554	0,55669	0,44992	0,37207	0,64711	0,51140	0,52364	0,45901	0,25537	0,39752	0,554215		0,91769

Tabla III.8: Frecuencias de las repeticiones identificadas en los marcadores contenidos en el kit Yfiler en la población española (n=173).

Allele	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS393	DYS391	DYS439	DYS635	DYS392	YGATAH4	DYS437	DYS438	DYS448	Allele	DYS385A/B
9								0,05848	0,00585					0,09357		11,11	0,00581
10							0,00585	0,44444	0,07018		0,00585	0,02924		0,26316		12,11	0,00581
11		0,00585						0,46784	0,31579		0,34503	0,41520		0,04678		12,12	0,01163
12		0,15789				0,01170	0,17544	0,02924	0,45029		0,03509	0,50877		0,57895		13,11	0,01744
<13	0,01170															13,12	0,04070
13	0,01170	0,59064				0,05848	0,67251		0,15205		0,57310	0,04094		0,01754		13,13	0,01163
>13												0,00585				14,9	0,00581
<14					0,00585											14,10	0,02326
14	0,09942	0,24561			0,02339	0,65497	0,12281		0,00585		0,03509		0,32164			14,11	0,40116
15	0,43275				0,12865	0,22807	0,02339				0,00585		0,56725			14,12	0,05233
16	0,37427				0,29240	0,04094							0,11111			14,13	0,04651
17	0,07018				0,28655	0,00585									0,01170	14,14	0,03488
17.2					0,00585											15,10	0,01163
18					0,21637										0,15789	15,11	0,07558
19					0,02339										0,46199	15,12	0,01163
19.2					0,00585											15,13	0,03488
20										0,07018					0,24561	15,14	0,02326
>20					0,01170											16,11	0,00581
21			0,01754							0,18129					0,11696	16,13	0,04070
22			0,11111							0,09357						16,14	0,01163
23			0,21637							0,53801					0,00585	16,15	0,00581
24			0,53801							0,09942						16,16	0,00581
25			0,09942							0,01170						17,13	0,00581
26			0,01170							0,00585						17,14	0,00581
27			0,00585	0,00585												17,15	0,00581
28				0,16374												17,16	0,01163
29				0,46199												18,12	0,00581
30				0,26316												18,13	0,00581
31				0,08187												18,14	0,03488
32				0,01754												18,15	0,01744
33				0,00585												18,16	0,01163
																19,13	0,00581
																19,17	0,00581
PD	0,65757	0,56585	0,64102	0,68342	0,76769	0,51373	0,50128	0,57932	0,66940	0,65394	0,54998	0,56619	0,56243	0,58432	0,68746		0,81842

Si se comparan las tablas con las del capítulo anterior, el contenido de éstas ofrece menor cantidad de datos aplicables. Únicamente se pueden utilizar como parámetros estadísticos el poder de discriminación (PD) y la frecuencia de cada variante alélica en la población. Debido a su herencia en bloque no se debe utilizar el cromosoma Y para realizar cálculos de identificación, pero sí como una herramienta que discierna y pueda aumentar o reducir el campo de búsqueda, siempre acompañado de otros datos.

Si se observan estas tablas se puede ver que los marcadores que ofrecen un mayor poder de discriminación en las tres poblaciones son el DYS 385A/B y el DYS458. Sí se ven diferencias marcadas entre ellas, como el aumento del PD del marcador DYS 389II en la población maya o el descenso del poder de discriminación del DYS 385A/B en la población española.

El poder combinado de discriminación, calculado del obtenido en cada locus, en todas las poblaciones es más alto que 0.99999. Pero si se observan todas las poblaciones, se ve que el valor obtenido en la población mestiza (de Guatemala) es el más alto, seguido de la población española.

Tabla III.9: Poder combinado de discriminación.

P. Mestiza Guatemala	n=115	0,999999989
P. Maya Guatemala	n=110	0,999999403
P. Española	n=173	0,999998297

3.5 DISTANCIAS CON OTRAS POBLACIONES

Una vez calculadas las frecuencias de nuestras poblaciones, se pasó a observar las relaciones genéticas de éstas con las frecuencias de otras poblaciones, obtenidas de la bibliografía consultada (ver tabla III.14). Mediante el paquete Estadística v8 de StatSoft® se realizó una comparación de las distintas frecuencias. En este análisis se redujo el número de marcadores a 13, omitiendo

los resultados de los marcadores DYS635, DYS439, DYS456 y DYS458. Esta comparación se realizó mediante un análisis de correspondencias.

La población maya, una vez realizado el cálculo, figura III.8, se situó cerca del resto de las poblaciones nativas americanas. Las poblaciones nativas americanas se situaron entre las muestras asiáticas y las mestizas americanas. Todas las poblaciones mestizas se situaron en posiciones intermedias entre las nativas americanas y las europeas. Las europeas se sitúan en el cuadrante superior derecho, situación debida a las frecuencias de los marcadores DYS438 (alelos 9 y 12) y al DYS437 (alelo 16). Estos marcadores correlacionan a las muestras europeas, muy homogéneas y distintas al resto. La posición de las muestras asiáticas, Japón, Korea, Tíbet y Seúl en el cuadrante inferior la alcanzan por los alelos 12 del marcador DYS389I y las frecuencias de las repeticiones 27 y 28 del DYS389II.

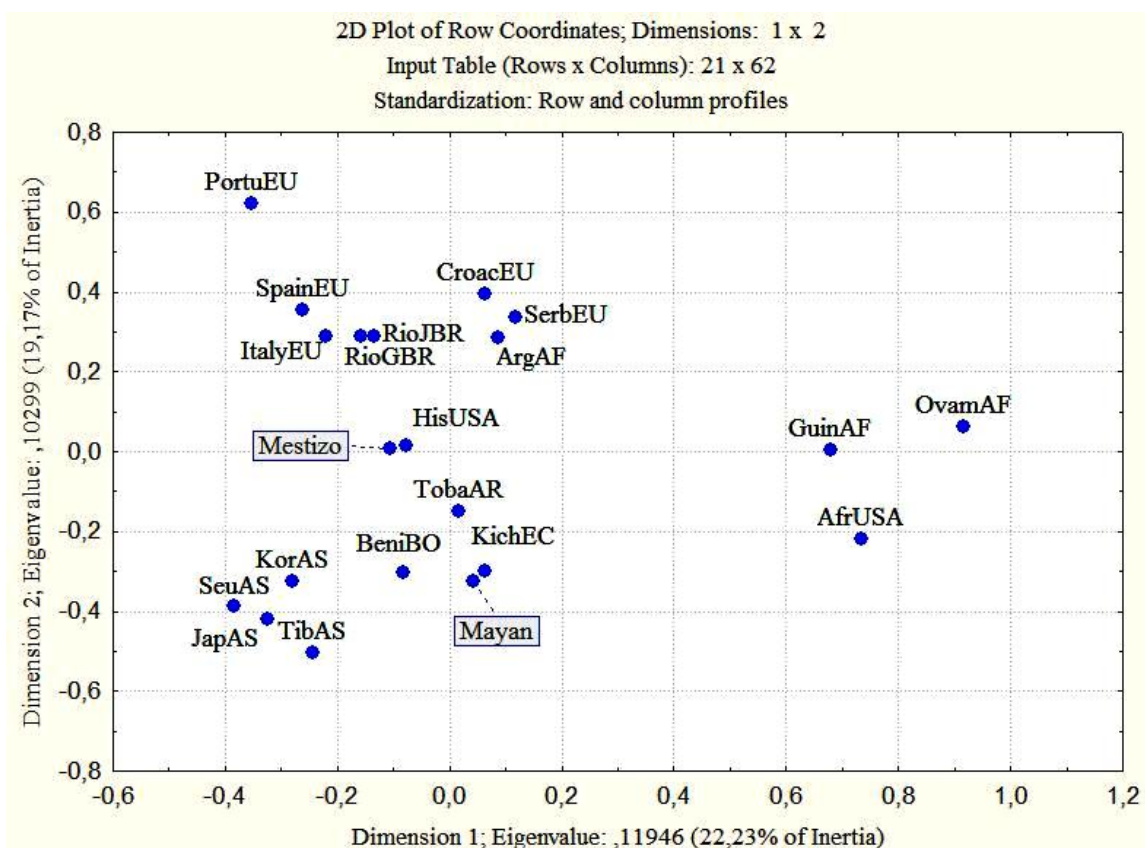


Figura III.8: Análisis de correspondencias a partir de las frecuencias obtenidas en los marcadores del kit Yfiler.

Las poblaciones africanas, excepto Argelia, se alejan del resto, situándose a la derecha del gráfico. Este caso se explica por las frecuencias obtenidas en los marcadores: DYS390 (alelo 21), DYS393 (alelo 15), DYS437 (alelo 13), DYS392 (alelo 12), DYS389II (alelo 32) y DYS19 (alelo 17). Sin embargo, Argelia está mucho más próxima a las poblaciones europeas.

Como se puede ver en este gráfico, la posición de las poblaciones se aleja mucho de las distancias geográficas entre cada una de ellas. Pero sí que refleja la historia de cada una de ellas. Así se puede entender que las poblaciones nativas americanas se encuentren cercanas a las poblaciones asiáticas y que las poblaciones mestizas estén en una posición intermedia entre las nativas americanas y las mestizas. De igual modo no es de extrañar la posición de Argelia, mucho más relacionada, a lo largo de su historia, con poblaciones europeas que con africanas.

Como ya se describió en el capítulo anterior no se pone de manifiesto componente africana entre los mestizos de Guatemala, es decir, en el pool génico de Guatemala no ha habido intromisión de genes africanos. En otras poblaciones americanas es muy usual encontrarlas debido a la introducción de esclavos en América.

Tras realizar el análisis de las frecuencias, se pasó a realizar un análisis de distancias. A partir de los perfiles genéticos se hallaron las distancias de Reynolds [28] y Slatkin [29]. Para realizar este análisis se utilizaron 11 marcadores, eliminando DYS456, DYS458, DYS385ab, DYS635 y DYS439. Se toman estas distancias genéticas y mediante un análisis multidimensional y se eligió el modelo euclideo para la representación (ver figura III.9). Los resultados obtenidos fueron similares a los obtenidos en el análisis de correspondencias.

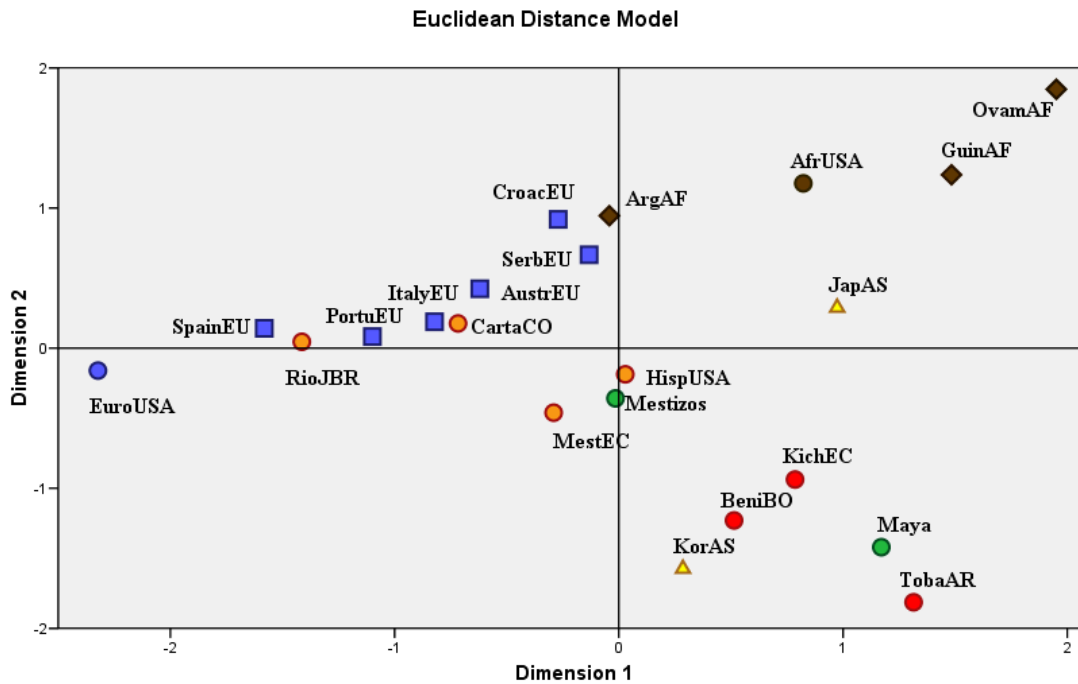


Figura III.9: Plot de las distancias euclideas, MDS calculado a partir de las distancias de Reynolds entre las poblaciones. En este análisis se comparan: poblaciones europeas, americanas del norte y del sur, africanas y asiáticas. 0.3817

En este nuevo análisis se observa como la población maya se sitúa cercana a las poblaciones nativas americanas, en el cuadrante inferior derecho. Junto a éstas se sitúa la población de Korea y sobre este conjunto, ya en el cuadrante superior, se ve la población de Japón (en este caso no se representan las poblaciones de Seúl y Tíbet, ya que solo se disponía de las frecuencias). En el cuadrante superior derecho, se observan las poblaciones africanas, como en el análisis de correspondencias, y la población argelina se sitúa más cercana a las poblaciones europeas. En el cuadrante superior izquierdo se ven todas las poblaciones de origen europeo y algunas mestizas (poblaciones de algunas ciudades sudamericanas, Rio de Janeiro y Cartagena). Las poblaciones mestizas, así como la población hispana de los EEUU, se posicionan en el centro del gráfico, una posición intermedia entre las poblaciones nativas americanas y las europeas.

3.6 ESTUDIO DEL MESTIZAJE

Como ya se dijo en el capítulo anterior, es muy interesante realizar un estudio del mestizaje al estudiar la población de Guatemala. En el caso del cromosoma Y se puede estudiar el aporte de los conquistadores españoles a la población mestiza de este país y también observar cuánto conservan los mestizos de sus vecinos nativos.

En primer lugar, se va a realizar un análisis de distancias genéticas, como el realizado anteriormente, omitiendo todas las poblaciones que pueden hacer que se dispersen o agrupen nuestras poblaciones. Como en los análisis anteriores se utilizaron 11 marcadores, eliminando DYS456, DYS458, DYS385ab, DYS635 y DYS439.

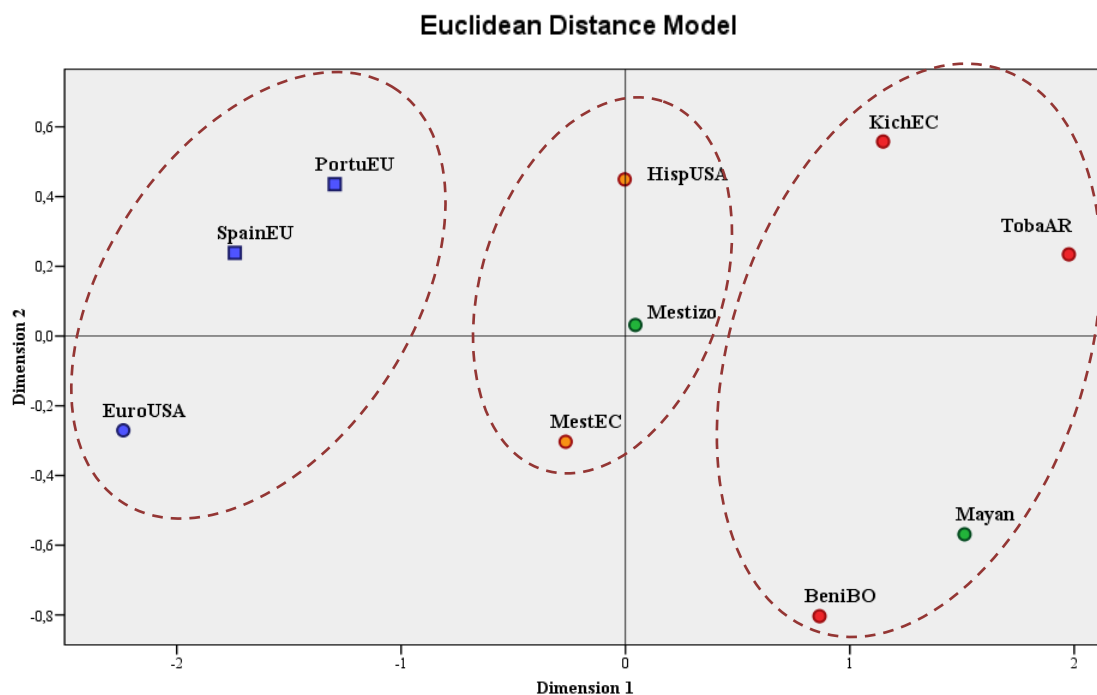


Figura III.10: Representación euclídea del cálculo MDS a partir de las distancias de Reynolds entre las poblaciones Europeas y Americanas, nativas y mestizas. Stress 0.1407

Si se observa la figura III.10 se ve el resultado de este análisis multidimensional y su representación euclídea que distribuyen a las poblaciones en tres ejes; a la izquierda las europeas, en el centro las mestizas y a la derecha las poblaciones

nativas americanas. En las poblaciones americanas, se observa cómo aumentan las diferencias entre ellas. Estas distancias representan cuán aisladas han estado estas poblaciones y como van divergiendo unas de otras aunque el origen de las mismas sea común.

Para conocer el aporte de la población española y maya a la población mestiza, se comparan las frecuencias de los distintos haplogrupos de la población mestiza con las parentales utilizando los programas Admix Pas y Admix 95.

Tabla III.10: Contribución Maya y Española a la población Mestiza de Guatemala.

Programa	Población Parental	Porcentaje de mezcla	Error
ADMIX.PAS	Maya	0,5084	0,0580
	Española	0,4916	0,0580
ADMIX95	Maya	0,5138	0,0343
	Española	0,4862	0,0343

Los resultados que se muestran han sido calculados por dos métodos distintos, el método de identidad genética [34] y el de los mínimos cuadrados [33]. Los resultados con el modelo biparental (españoles y mayas) son similares, próximos al 50% por cada una de las poblaciones. Los dos programas difieren en un tanto por ciento mucho menor al error que puede producirse al realizar el cálculo. Estos resultados muestran el gran aporte de la población española en la estructura genética de la población mestiza de Guatemala a nivel de su cromosoma Y.

3.7 ESTRUCTURA DE LA POBLACIÓN

Para conocer la estructura de cada una de las poblaciones estudiadas se empleó el programa STRUCTURE 2.3.1 [37,38]. Con este programa se detectará la presencia de uno o varios componentes dependiendo de las poblaciones asumidas (K) dentro de cada una de las poblaciones. Ésta es una forma de ver si en un grupo muestral existen diferencias significativas como para poder diferenciar poblaciones distintas.

La variabilidad de los STRs estudiados en el caso del cromosoma Y pondrá de manifiesto si, mediante estos marcadores, es posible encontrar en la población mestiza componentes del resto de nuestras poblaciones.

Como parámetros para el uso del programa, se predefine un número de 50.000 interacciones y un periodo "burn-in" de 10.000 interacciones adicionales antes de cada uno de los análisis. Estos se realizaron siguiendo el siguiente diseño (tabla III.11):

Tabla III.11: Diseño y resultados obtenidos mediante el análisis en STRUCTURE 2.3.1.

POBLACIÓN	n	k	MÉTODO	Ln P(d)	k	Ln P(d)
				17-Y-STRs		17-Y-STRs
GUATEMALTECOS	225	1	Método Admixture	-4778,7	3	-4240,4
			Método Non admixture	-4776,2		-4166,1
		2	Método Admixture	-4433,6	4	-4203,6
			Método Non admixture	-4393,1		-4043,0
MESTIZOS	115	1	Método Admixture	-2566,4	3	-2223,6
			Método Non admixture	-2566,0		-2198,2
		2	Método Admixture	-2315,7	4	-2235,0
			Método Non admixture	-2312,3		-2219,6
MAYAS	110	1	Método Admixture	-2106,4	3	-1979,1
			Método Non admixture	-2106,4		-1966,5
		2	Método Admixture	-2020,2	4	-2015,1
			Método Non admixture	-2009,9		-1982,4
ESPAÑOLES	172	1	Método Admixture	-3153,7	3	2580,0
			Método Non admixture	-3154,6		-2569,8
		2	Método Admixture	-2623,5	4	-2551,2
			Método Non admixture	-2619,6		-2534,4
ESPAÑOLES GUATEMALTECOS	397	2	Método Admixture	-7308,3	6	-6755,2
			Método Non admixture	-7236,6		-6367,7
		3	Método Admixture	-6839,8	7	-6770,7
			Método Non admixture	-6700,1		-6277,3

Si se estudia la población de Guatemala, incluyendo tanto mayas como mestizos, se ve que la probabilidad va aumentando según se aumenta el número de poblaciones asumidas (K) hasta llegar a cuatro. En Mestizos la probabilidad más alta aparece en tres, al igual que en mayas. Pero si se observan los gráficos de barras, como se puede ver en la figura III.11, se comprueba que al distribirse las componentes a lo largo del grupo muestral Guatemala, las componentes de cada población sí separan al grupo poblacional mestizo del maya. Esto ocurre tanto asumiendo K=2, como 3 y 4.

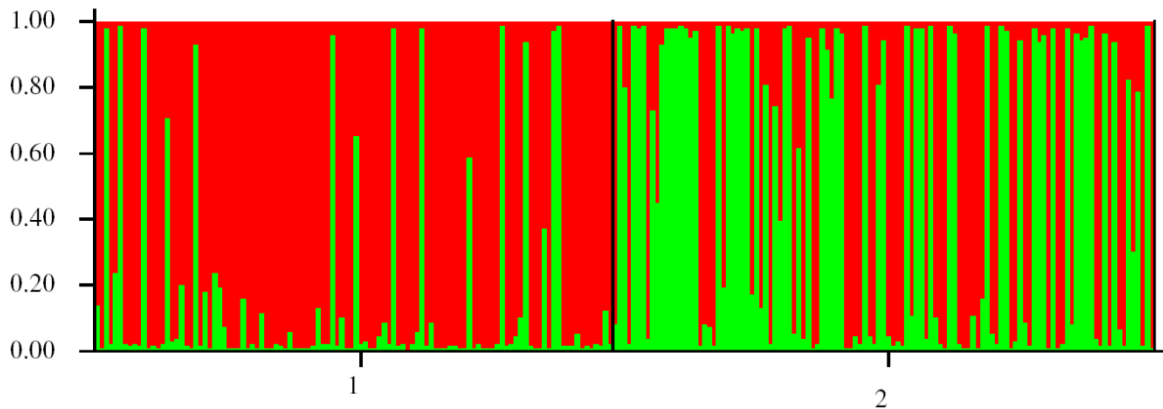


Figura III.11: Imagen obtenida en Structure. Tomando los distintos grupos poblacionales guatemaltecos. $K=2$

Si se observa la figura III.11, el análisis se ha hecho con dos poblaciones asumidas, $K=2$. El programa busca dos componentes que separen a las poblaciones, y se analizan los individuos buscando la presencia de éstas. Las dos componentes en esta imagen se representan en verde y en rojo, y los individuos se tiñen de verde o de rojo, o de ambos, según exista una mayor o menor presencia de la componente. Si se ve la figura, hay dos grupos, el 1, a la izquierda donde están los individuos maya y el 2 donde están los mestizos. En el grupo 1 aparece mayoritariamente la componente roja, mientras que en el grupo 2 aparece mayoritariamente la verde (aunque también aparece la roja en muchas ocasiones). Se puede, por tanto, diferenciar ambos grupos, ya que su comportamiento es distinto durante el análisis de ambas componentes.

El mismo fenómeno ocurre cuando se estudia todo el grupo muestral ($n=397$). La probabilidad va aumentando según se va asumiendo mayor número de poblaciones. Pero al observar cómo se distribuyen éstas en los individuos de cada una de las poblaciones se ve perfectamente cómo se pueden diferenciar los tres grandes grupos, mayas mestizos y españoles.

Si se observa la figura III.12, se ve cómo se diferencian los tres grupos. El análisis es de tres poblaciones asumidas, $K=3$, por lo que se distribuyen tres componentes (representadas en color verde, azul y rojo). El más característico es el maya (grupo 1 a 4) donde mayoritariamente se observa la componente

representada en verde. En cambio el grupo español (6) se distribuyen dos componentes, la roja y la azul, a lo largo de todo el grupo. El grupo mestizo (5), aparece como una mezcla de los tres componentes que distribuye el programa.

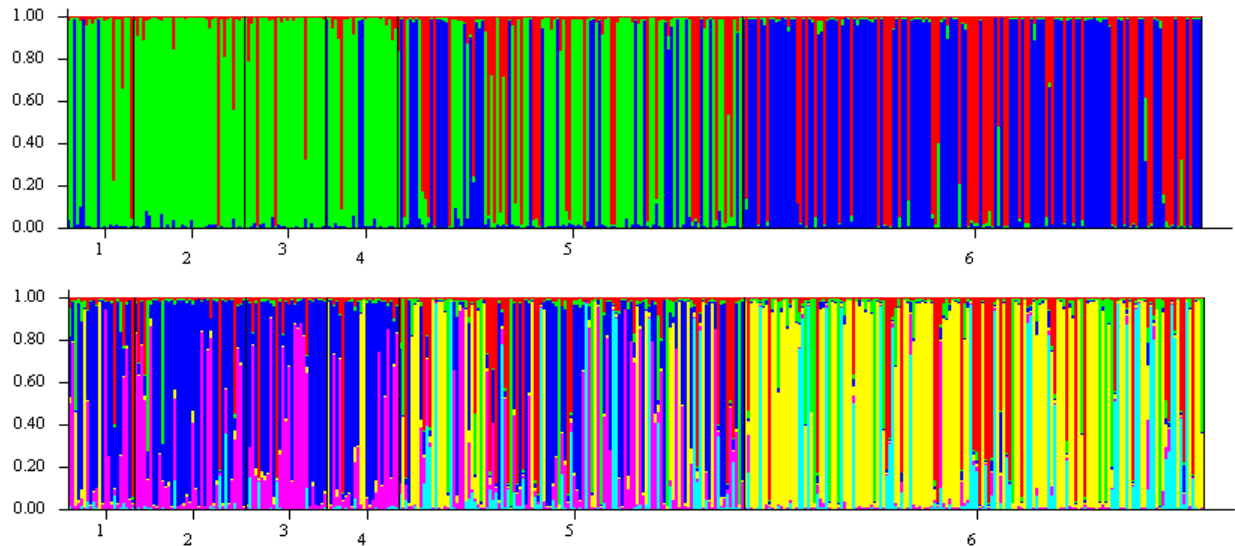


Figura III.12 y 13: Imagen obtenida en Structure. Tomando los distintos grupos poblacionales guatemaltecos y español ($K=3$). En la figura III.13 K es 6.

Cuando se toma el grupo muestral completo y se asume $K=6$, busca seis características distintas. Tiñendo los individuos en azul, violeta, amarillo, rojo, turquesa y verde. La forma de distribuir las es la que muestra la figura III.13. En el grupo 6, españoles, se ve como aparecen los colores amarillo, rojo, turquesa y minoritariamente el verde. En el grupo 5, mestizos, aparecen todos los componentes representados. En el maya mayoritariamente aparecen el azul y el violeta. Se ve como no encuentra diferencias entre los grupos que forman el grupo maya (1 a 4), tiñéndolos de azul y violeta, pero sin poder diferenciar ningún grupo con exactitud. El grupo español 6, toma ahora hasta cuatro de los componentes, distribuidos de manera bastante uniforme por toda la muestra. El grupo mestizo (5) como ocurría en el caso de estudiar 3 componentes aparecen las mismas componentes que en los grupos maya y español.

4. RESUMEN

El cromosoma Y, así como los STRs representados en él, presentan una serie de características de herencia y recombinación que permiten realizar una serie de análisis que no es posible realizar con otros marcadores.

Tras realizar este estudio, se confirma el alto poder de discriminación de estos marcadores incluidos en el Kit Yfiler, en todas las poblaciones, incluso en las más pequeñas ya que, como se mostró, no se ha encontrado ningún haplotipo repetido. Aunque se explicaron las diferencias entre las tres poblaciones, todos los marcadores estudiados ofrecen un poder de discriminación alto, siempre mayor al 50%, lo que hace de su poder combinado un instrumento eficaz en identificación genética forense.

La variabilidad encontrada en sus marcadores y su poder de discriminación total (en todas las poblaciones) combinado con las técnicas de extracción adecuadas, permiten resolver numerosos y arduos casos forenses.

Esta variabilidad se debe a su tasa de mutación, que es la que facilita la aparición de nuevas variantes alélicas que permitan diferenciar nuevos haplotipos. Esta tasa “elevada” de mutación es la que hace que estos marcadores no sean tan robustos en el caso de hacer estudios de antropología molecular. Para poder realizar estudios con estos marcadores se han tenido que omitir muchos de los loci analizados. Una vez excluidos estos, los análisis realizados se han ajustado completamente a todo lo que se podía esperar debido a la historia de nuestras poblaciones.

De los distintos análisis realizados se desprende que la población maya comparte distancias bajas con poblaciones asiáticas y más elevadas con poblaciones europeas y africanas. La población española se encuentra cercana a sus poblaciones vecinas, Portugal e Italia, pero también a muchas de las poblaciones mestizas americanas. La población Mestiza de Guatemala se

encuentra en una posición intermedia entre las poblaciones nativas americanas y las europeas, al igual que muchas de las poblaciones mestizas americanas.

El haplogrupo mayoritario en las poblaciones maya de Guatemala, es el haplogrupo Q. En la población mestiza, también es el Q (44%), pero seguido por el R1b1 (27%). En la población española el haplogrupo mayoritario es el R1b1.

En los análisis realizados en cromosoma Y de la población mestiza se puede ver un ligero aumento del porcentaje parental de la población española, con respecto a los análisis realizados en a-STRs.

Al igual que con los análisis realizados en el capítulo anterior se observa que la población de Guatemala está claramente dividida en dos subpoblaciones, una de origen maya, particular y aislada, y otra mestiza fruto de la mezcla de la población española y la anterior.

CONCLUSIONS

Y chromosome, as the STRs' represented in it, has many heredity and recombination characteristics that enable us to perform some analysis that are impossible to be carried out with other kind of markers.

After finishing this study, the high discrimination power of these markers included in Yfiler kit is confirmed in all the populations, even in the smaller ones as no repeated haplotype was found (page 146). Although the differences between the discrimination power of the markers were explained in the three populations, all the studied markers offer a high discrimination power, higher than 50%. This fact makes the combined discrimination power an efficient tool in forensic genetics identification.

The variability found in the markers and the total discrimination power (in all the populations), combined with the suitable extraction procedures let us solve several and arduous forensic cases.

This variability is due to its mutation rate that facilitates the appearance of new allelic variants that differentiates between new haplotypes. The "high" mutation rate is the cause that makes these markers not a robust tool in molecular anthropology studies. To perform these kind of studies with Y-STR markers, some analyzed loci had to be omitted. Once these markers are omitted, the analyses were completely adjusted to the expected results due to the history of the populations.

From the different analyses it emerges that the Mayan population shares low distances with Asiatic population, but high with European and African populations. The Spanish population is near from its neighbours, Portugal and Italy, but it is also close to several American Mestizo populations. The

Guatemalan Mestizo population is located between Native American populations and European ones, as many of the American Mestizo populations.

The most common haplogroup in the Guatemalan Mayan populations is the Q haplogroup. Q haplogroup (44%) is the most representative haplogroup in the Mestizo population but R1b1 haplogroup is significant (27%). In the Spanish population, the most common haplogroup is R1b1.

In the Y-STRs analysis performed within the Mestizo population, a slight increase in the parental percentage of the Spanish population can be noted compared to the analysis in a-STRs.

As in previous analysis, the Guatemalan population is clearly divided in to two subpopulations, one with Mayan origins (particular and isolated) and one mixed result from the admixture of the Spanish population and the Mayan one

5. BIBLIOGRAFÍA

- [1] Central Intelligence Agency, *The World Factbook*, (2009).
- [2] Instituto Nacional de Estadística, *XI Censo Nacional de Población y VI de Habitación*, 2002.
- [3] J.M. Butler, Genetics and genomics of core short tandem repeat loci used in human identity testing, *J.Forensic Sci.* 51 (2006) 253-265.
- [4] F. González-Andrade, D. Sánchez, J. González-Solórzano, S. Gascón, B. Martínez-Jarreta, Sex-specific genetic admixture of Mestizos, Amerindian Kichwas, and Afro-Ecuadorans from Ecuador, *Human Biology.* 79 (2007) 51-77.
- [5] Y.M. Chang, R. Perumal, P.Y. Keat, D.L.C. Kuehn, Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia, *Forensic Sci.Int.* 167 (2007) 70-76.
- [6] J.U. Palo, M. Pirttimaa, A. Bengs, V. Johnsson, I. Ulmanen, M. Lukka, et al., The effect of number of loci on geographical structuring and forensic applicability of Y-STR data in Finland, *Int.J.Legal Med.* 122 (2008) 449-456.
- [7] A. Gómez, S.J. Ávila, I. Briceño, Correlation analysis of surnames and Y-chromosome genetic heritage in 3 provinces of southwestern Colombia, *Biomedica.* 28 (2008) 359-370.
- [8] Y.M. Chang, Y. Swaran, Y.K. Phoon, K. Sothirasan, H.T. Sim, K.B. Lim, et al., Haplotype diversity of 17 Y-chromosomal STRs in three native Sarawak populations (Iban, Bidayuh and Melanau) in East Malaysia, *Forensic Science International: Genetics.* 3 (2009) e77-e80.
- [9] M.A. Jobling, P. Gill, Encoded evidence: DNA in forensic analysis, *Nature Reviews Genetics.* 5 (2004) 739-751.
- [10] O. Lao, T.T. Lu, M. Nothnagel, O. Junge, S. Freitag-Wolf, A. Caliebe, et al., Correlation between Genetic and Geographic Structure in Europe, *Current Biology.* 18 (2008) 1241-1248.
- [11] L. Venkatachalam, R.V. Sreedhar, N. Bhagyalakshmi, The use of genetic markers for detecting DNA polymorphism, genotype identification and phylogenetic relationships among banana cultivars, *Mol.Phylogenet.Evol.* 47 (2008) 974-985.
- [12] J.U. Palo, I. Ulmanen, M. Lukka, P. Ellonen, A. Sajantila, Genetic markers and population history: Finland revisited, *European Journal of Human Genetics.* 17 (2009) 1336-1346.

- [13] H. Liu, W.-. Yu, F. Fang, X.-. Wang, G. Yang, B. Liu, et al., Polymorphism of microsatellite DNA vWA in random population and schizophrenic patients, *Chinese Journal of Clinical Rehabilitation*. 9 (2005) 252-254.
- [14] J. Liu, Y. Shi, J. Tang, T. Guo, X. Li, Y. Yang, et al., SNPs and haplotypes in the S100B gene reveal association with schizophrenia, *Biochem.Biophys.Res.Commun.* 328 (2005) 335-341.
- [15] V.L. Martinez-Marignac, A. Valladares, E. Cameron, A. Chan, A. Perera, R. Globus-Goldberg, et al., Admixture in Mexico City: Implications for admixture mapping of Type 2 diabetes genetic risk factors, *Hum.Genet.* 120 (2007) 807-819.
- [16] J.J. Mulero, C.W. Chang, L.M. Calandro, R.L. . Green, Li Y, C.L. Johnson, et al., Development and Validation of the AmpFISTR® Yfiler™ PCR Amplification Kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System, *J.Forensic Sci.* 5 (2006) 64-75.
- [17] AmpFISTR® Yfiler™, PCR Amplification Kit, User's Manual, Part Number 4358101. Rev. C 08/2006,.
- [18] P.M. Schneider, Scientific standards for studies in forensic genetics, *Forensic Sci.Int.* 165 (2007) 238-243.
- [19] D. Frumkin, A. Wasserstrom, A. Davidson, A. Grafit, Authentication of forensic DNA samples, *Forensic Science International: Genetics*. 4 (2010) 95-103.
- [20] T. Power, B. McCabe, S.A. Harbison, FaSTR DNA: A new expert system for forensic DNA analysis, *Forensic Science International: Genetics*. 2 (2008) 159-165.
- [21] A. Tereba, Tools for Analysis of Population Statistics, Profiles in DNA. (2001).
- [22] T.W. Athey, Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach, *Journal of Genetic Genealogy*. 2 (2006) 34-39.
- [23] T.W. Athey, Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach, *Journal of Genetic Genealogy*. 1 (2005) 1-7.
- [24] 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.
- [25] W. Athey, Comments on the article, "Software for Y Haplogroup Predictions, a Word of Caution", *Int.J.Legal Med.* (2010) 1-3.
- [26] W. Ludwig, O. Strunk, R. Westram, L. Richter, H. Meier, A. Yadhukumar, et al., ARB: A software environment for sequence data, *Nucleic Acids Res.* 32 (2004) 1363-1371.

- [27] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: An integrated software package for population genetics data analysis, *Evolutionary Bioinformatics Online*. 1 (2005) 47.
- [28] 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.
- [29] M. Slatkin, Linkage disequilibrium in growing and stable populations, *Genetics*. 137 (1994) 331-336.
- [30] J. Felsenstein, An alternating least squares approach to inferring phylogenies from pairwise distances, *Syst.Biol.* 46 (1997) 101-111.
- [31] J. Felsenstein, Comparative methods with sampling error and within-species variation: Contrasts revisited and revised, *Am.Nat.* 171 (2008) 713-725.
- [32] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* 4 (1987) 406-425.
- [33] J.C. Long, The genetic structure of admixed populations, *Genetics*. 127 (1991) 417-428.
- [34] R. Chakraborty, Estimation of race admixture. A new method, *Am.J.Phys.Anthropol.* 42 (1975) 507-511.
- [35] B. Bertoni, B. Budowle, M. Sans, S.A. Barton, R. Chakraborty, Admixture in Hispanics: Distribution of ancestral population contributions in the continental United States, *Human Biology*. 75 (2003) 1-11.
- [36] M. Stephens, P. Donnelly, Inference in molecular population genetics, *R.Stat.Soc.Ser.B.* (2000).
- [37] J.K. Pritchard, M. Stephens, N.A. Rosenberg, P. Donnelly, Association mapping in structured populations, *Am.J.Hum.Genet.* 67 (2000) 170-181.
- [38] M.J. Hubisz, D. Falush, M. Stephens, J.K. Pritchard, Inferring weak population structure with the assistance of sample group information, *Molecular Ecology Resources*. 9 (2009) 1322-1332.
- [39] 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.
- [40] U. Toscanini, L. Gusmão, G. Berardi, A. Amorim, Á. Carracedo, A. Salas, et al., Y chromosome microsatellite genetic variation in two Native American populations from Argentina: Population stratification and mutation data, *Forensic Science International: Genetics*. 2 (2008) 274-280.

- [41] M. Tirado, A.M. López-Parra, C. Baeza, F. Bert, A. Corella, A. Pérez-Pérez, et al., Y-chromosome haplotypes defined by 17 STRs included in AmpFISTR Yfiler PCR Amplification Kit in a multi ethnical population from El Beni Department (North Bolivia), *Leg.Med.* 11 (2009) 101-103.
- [42] J.J. Builes, B. Martínez, A. Gómez, L. Caraballo, C. Espinal, D. Aguirre, et al., Y chromosome STR haplotypes in the Caribbean city of Cartagena (Colombia), *Forensic Sci.Int.* 167 (2007) 62-69.
- [43] D.A. Silva, E. Carvalho, G. Costa, L. Tavares, A. Amorim, Gusmao L., Y-Chromosome Genetic Variation in Rio de Janeiro Population, *Am.J.Hum.Biol.* 18 (2006) 829-837.
- [44] S.P. Schwengber, T. Kommers, C.H.F. Matte, P.E. Raimann, B.A. Carvalho, F.P.N. Leite, et al., Population data of 17 Y-STR loci from Rio Grande do Sul state (South Brazil), *Forensic Science International: Genetics.* 4 (2009) e31-e33.
- [45] F. González-Andrade, L. Roewer, S. Willuweit, D. Sánchez, B. Martínez-Jarreta, Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis, *Forensic Science International: Genetics.* 3 (2009) e83-e91.
- [46] A. Edwards, H.A. Hammond, L. Jin, C.T. Caskey, R. Chakraborty, Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups, *Genomics.* 12 (1992) 241-253.
- [47] 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.
- [48] 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.
- [49] 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.
- [50] B. Budowle, J. Ge, J. Low, C. Lai, W.H. Yee, G. Law, et al., The effects of Asian population substructure on Y STR forensic analyses, *Leg.Med.* 11 (2009) 64-69.
- [51] H.Y. Lee, M.J. Park, U. Chung, H. Young Lee, W.I. Yang, S.H. Cho, et al., Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father-son pairs, *Int.J.Legal Med.* 121 (2007) 128-135.
- [52] A. Gómez-Martín, L.J. Martínez-González, M.J. Alvarez-Cubero, C. Entrala, F. Fernández-Rosado, A. Gonzalez-Liñan, et al., Population data for Y-chromosome

haplotypes and haplogroups defined by 16 STRs (AmpFSTR® Yfiler™ Kit) in Spain, IALM. XXI (2009).

[53] A.M. Bento, M. Carvalho, V. Lopes, A. Serra, H.A. Costa, L. Andrade, et al., Distribution of Y-chromosomal haplotypes in the Central Portuguese population using 17-STRs, *Forensic Science International: Genetics*. 4 (2009) e35-e36.

[54] 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.

[55] N. Haliti, M. Čarapina, M. Mašić, D. Strinović, I.M. Klarić, M. Kubat, Evaluation of population variation at 17 autosomal STR and 16 Y-STR haplotype loci in Croatians, *Forensic Science International: Genetics*. 3 (2009) e137-e138.

[56] 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.

[57] B. Berger, A. Lindinger, H. Niederstätter, P. Grubwieser, W. Parson, Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay, *Int.J.Legal Med.* 119 (2005) 241-246.

[58] M. Goedbloed, M. Vermeulen, R.N. Fang, M. Lembring, A. Wollstein, K. Ballantyne, et al., Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpF ISTR® Yfiler® PCR amplification kit, *Int.J.Legal Med.* (2009) 1-12.

Capítulo IV: Polimorfismos en la Secuencia del ADN Mitocondrial.

Chapter IV: Polymorphisms in mitochondrial DNA sequence.



ABSTRACT

The high variability in mtDNA regions has been used by scientists to identify testing in criminal and missing people cases, but also in anthropology sciences to evidence the migratory movements of populations. The population of Guatemala is composed by 55% Mestizo (Amerindian-Spanish) and 45% of Amerindian populations. The aim of this chapter is to characterize the mtDNA control region from Mestizo population and four aboriginal Guatemalan populations.

Samples of 49 unrelated Kaqchiquel, 47 Qeqchi, 47 Mam, 48 K'iche' and 105 Mestizos living in Guatemala were collected on FTA® cards (Whatman, Florham Park, NJ). Also, samples of 295 unrelated Spanish samples were studied and compared to the Guatemalan samples.

DNA was purified, according to the manufacturer recommendations. DNA was amplified for HV1 (16024-16365) and HV2 (73-340) with specific primers, following previous international recommendations. Sequencing was carried out

in an ABI 310 (Applied Biosystems, Foster City, CA) using Big Dye terminator v1.1 kit.

Haplotype frequencies were calculated by gene counting method. Haplotype diversities were calculated according to Nei using Arlequin v3.5 software. An analysis of molecular variance (AMOVA) was also performed. Previously published data from several populations were compiled, which comprised of South American, Central American and North American populations. Arlequin v3.5 software was used to calculate the distributions of the observed HV1 sequences (16024-16365) in each group, F-statistics, Reynolds and Slatkin's genetic distances. Furthermore, this 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. To represent the distances in an appropriate way, genetic distances between American populations were graphically summarized by nonmetric multidimensional scaling (NM-MDS), using the software SSPS v.15.0. The estimation of individual ancestry proportions was studied with STRUCTURE v2.3.1. In addition, the relationship between the samples was inferred by Haplogrep.

A total of 205 unique haplotypes out of the 296 Guatemalan samples were found. Most of the samples belonged to haplogroup A2 and the second most frequently observed group is B4 in Guatemalan populations groups.

This data shows enough variability of the population from Guatemala for the mitochondrial DNA hypervariable regions and supports the use of mtDNA analysis for identification purposes.

1. INTRODUCCIÓN

Dentro del genoma humano, existen porciones que poseen una herencia uniparental. Estas regiones tienen importantes aplicaciones en una amplia gama de campos como la genética forense [1,2], antropología molecular y la historia evolutiva de las poblaciones [3-5], la genética médica [6-8], la genealogía genética [9,10], como ya se explicó en el capítulo I.

En el primer capítulo de esta memoria, se desarrollaron las características del ADNmt (ADN mitocondrial). A lo largo de esta introducción solamente se realizará un breve resumen de éstas, se destacará como clasificar las mutaciones encontradas, la forma de agrupar las variaciones en haplogrupos, los fenómenos de variabilidad genética que se pueden encontrar en estas poblaciones, así como nuevas tecnologías para el análisis de las secuencias.

1.1 EL ADNmt

El ADNmt es una fracción pequeña del genoma que no reside en el núcleo de la célula, sino en las mitocondrias. Es una molécula circular de doble cadena con una longitud cercana a 16,6 kb, su herencia es por línea materna, carece de recombinación, se encuentra en un alto número de copias y posee una alta tasa de mutación. Esta tasa de cambio en las bases del ADN, queda fijada en muchos de los casos (debido a la ausencia de mecanismos reparadores) y en otros incluso desaparecerá debido a nuevos cambios que retornen su secuencia al estado original, convirtiéndose en un reloj molecular, que permite estudiar la evolución de la especie humana, por esta razón a la tasa de mutaciones fijadas, se le conoce como tasa de evolución [11].

Las posiciones de las bases del ADNmt humano se numeran de la 1 a la 16.569, de acuerdo a la secuencia de referencia revisada de Cambridge (rCRS) [12,13]. A partir de ésta se encuentran polimorfismos (cambios de nucleótido único, SNP,

y la inserción / deleción de bases, indels). Las tasas de mutación encontradas en las diferentes zonas del ADNmt son distintas: la tasa de mutación en la región control (no codificante) comprendida entre las posiciones 16024-576, es aproximadamente 10 veces mayor que la de la región codificante [14,15]. La explicación de esta diferencia en la tasa de mutación, radica en los procesos de selección, filtrando las posibles mutaciones nocivas. Además de estas diferencias entre la región codificante y no codificante, existen diferencias de la tasa de mutación entre las bases [16,17]; en la zona control posiciones como las 146, 150, 152, 195, 16189, 16311, 16362, 16519 son puntos calientes de alta variación, mientras que otras parecen bastante estables, 477, 493, 16108, 16219. De igual modo, esta diferencia entre bases se observa en la zona codificante [14,18-21]. Mientras que la alta variabilidad en la región control es muy interesante desde el punto de vista forense para su uso en identificación, no lo es, para la antropología molecular ya que puede dificultar la asignación de haplogrupos.

1.2 HAPLOGRUPOS EN EL ADNmt

Los haplotipos del ADNmt se pueden rastrear creando un árbol, llegando a un antepasado común por vía materna que vivió hace aproximadamente 200.000 años en África [22-24]. Al carecer de recombinación, la variación de la secuencia del ADNmt es el resultado de la acumulación de mutaciones, por lo que puede ser representada en un árbol que refleja las relaciones filogenéticas de las variantes de ADNmt [4,20,25-29]. En el año 2007, se generó un árbol mundial de la filogenia del ADNmt, MITOMAP [30]. El árbol MITOMAP rápidamente se consideró obsoleto, ya que durante ese año y el siguiente se publicaron muchas secuencias de ADNmt completo, que mejoraron la resolución de la filogenia. Gracias a miles de secuencias completas del ADNmt, en poco más de un año se generó un árbol filogenético actualizado del ADNmt humano [2]. Este árbol, accesible en la web <http://www.phylotree.org/>, está siendo frecuentemente actualizado (última actualización 20 de julio de 2011, visto 1 de septiembre de 2011).

Como ya se desarrolla en el primer capítulo, los haplotipos se pueden agrupar en haplogrupos. Los primeros haplogrupos se denominaron A, B, C y D [31,32], encontrados en indígenas americanos, que a su vez serán los más frecuentes en la población maya. Los posteriores haplogrupos se nombraron con letras sucesivas, hoy en día todas las letras están en uso.

La nomenclatura fue evolucionando y aparecieron nuevos subgrupos para designar nuevos linajes (A2q, H1b, etc.).

1.3 MOVIMIENTOS MIGRATORIOS Y EL ADN_{mt}

Cuando en las poblaciones aparecen eventos de flujo genético, migración y colonización, también lo hacen señales en nuestro genoma que pueden ser rastreadas a lo largo de los años [33]. La colonización puede conducir a una alteración y/o reducción de la diversidad genética a través de efectos fundadores o cuellos de botella [34]. El flujo genético por inmigración, la deriva genética, puede contrarrestar las fuerzas comentadas anteriormente [35]. Las consecuencias de estos movimientos genéticos van a depender del nivel y la estructura de la diversidad genética dentro y entre las poblaciones nativas, tamaño de la población fundadora, y el número individuos que aporta la migración, colonización u otros eventos que han ocurrido [36]. Los movimientos genéticos humanos se han caracterizado por el vertiginoso crecimiento de la población y su expansión con eventos de colonización desde África y posteriormente a otras regiones [37-39]. En este capítulo se estudiarán consecuencias genéticas de las migraciones relativamente recientes, buscando caracteres comunes a otros estudios de poblaciones aisladas por factores culturales y religiosos [40,41], en comunidades insulares [42], o en las migraciones de esclavos africanos [43], y la colonización europea de las Américas [29,44].

1.4 NUEVAS TÉCNICAS

La aparición de la secuenciación de segunda generación, NGS (Next generation sequencing), permite con un coste mucho menor, el análisis del genoma humano, y aún más de una región tan pequeña como es el ADNmt. Estas nuevas tecnologías de alto rendimiento, permiten agrupar por haplogrupos, todos los individuos estudiados en los proyectos de secuenciación de genoma completo [45]. La NGS puede facilitar en gran medida, el estudio de la clasificación poblacional y la asignación de los haplotipos para la genética forense y la antropología molecular [46,47]. Sin embargo, recientes estudios realizados por A. Salas, piden cautela a la hora de usar estas tecnologías en medicina legal, debido a los posibles errores encontrados en la asignación de los haplotipos [48].

2. MATERIAL Y MÉTODOS

2.1 LA MUESTRA

Al abordar el análisis de ADNmt, se realizaron análisis para 105 mestizos y 49 Kaqchiquel, 48 K'iche', 47 Mam y 47 Q'eqchi'. Y 295 individuos españoles [49]. No se pudieron realizar análisis para los 400 individuos donantes de Guatemala, debido a que muchas de las muestras se habían agotado llegado este análisis, por lo que en total, se analizaron 296 individuos.

La toma de la muestra se detalla en el capítulo II.

2.2 ANÁLISIS GENÉTICO MOLECULAR DE LA MUESTRA

Siguiendo las recomendaciones de la ISFG, se realizó la toma de las muestras, los controles, el análisis y su registro.

2.2.1 Extracción

La purificación del ADN, se realizó siguiendo los protocolos de extracción sobre FTA (Whatman, Clifton, NJ, USA).

2.2.2 Amplificación

Se tomó como ADN molde un punch (1.2mm Ø) de papel FTA. En todas las muestras, se amplificó la región control completa (16024-340) mediante los primers A1 y D1, con el kit AmpliTaq Gold® PCR Master Mix (Applied Biosystems, Foster City, CA) [50]. En algunas de las muestras no fue posible, por lo que se tuvo que recurrir a primers internos de la región, B1 y C1. La amplificación se realizó en un termociclador Biometra (T3 Thermocycler).

Los primers utilizados y sus secuencias fueron:

A1- L15996 5'-CTC CAC CAT TAG CAC CCA AAG C-3'

B1- H16401 5'-TGA TTT CAC GGA GGA TGG TG-3'

C1- L48 5'-CTC ACG GGA GCT CTC CAT GC -3'

D1- H408 5'-CTG TTA AAA GTG CAT ACC GCC A -3'

Para la PCR, se emplearon los siguientes reactivos y las condiciones que se pueden ver en la figura IV.1:

Reactivos:

- 12,5 μ l de AmpliTaq Gold® PCR Master Mix (Applied Biosystems Foster City, CA. USA) por cada muestra.
- 0,5 μ l de primer (10 μ M).
- 1 punch de FTA.
- Agua estéril hasta completar un volumen final de reacción de 25 μ l.

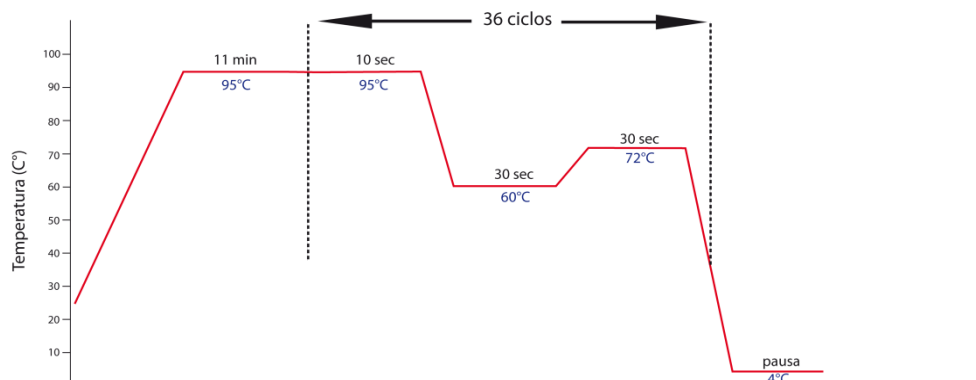


Figura IV.1 Condiciones de amplificación.

Además de las muestras, sistemáticamente se amplificaron dos controles: uno negativo y uno positivo.

Una vez finalizada la reacción de amplificación se evalúa la cantidad y calidad de los productos de amplificación mediante electroforesis en geles de agarosa al 2% (fase móvil=TBE).

2.2.3 Purificación del Producto de PCR

El objeto de esta fase es eliminar los nucleótidos (dNTPs) y primers que no se han utilizado en la reacción de PCR y que podrían interferir en la reacción de secuenciación. Se realizó mediante filtros Microcon[®]YM-30 (Millipore, Billerica, MA. USA) con un tamaño de poro 30 kDa dentro de los cuales se añade todo el producto de PCR. El proceso de filtrado se consigue al centrifugar y lavar la muestra en varias ocasiones.

2.2.4 Secuenciación

La secuenciación tiene por objeto la determinación del orden de los nucleótidos que componen el fragmento de ADNmt a estudiar. En este trabajo se empleó el método de secuenciación cíclica o Método de Sanger, mediante el kit ABI PRISM Big Dye v1.1 (Applied Biosystems.)

Este método se fundamenta en una reacción de elongación mediada por una ADN polimerasa (Amplitaq DNA polimerase, FS) en presencia de deoxinucleótidos (ddATP, ddCTP, ddTTP, ddGTP) marcados con fluorocromos. Se denomina secuenciación cíclica porque emplea sucesivos ciclos de desnaturalización, hibridación de primer, y polimerización para crear grandes cantidades de producto en una única reacción, incrementándose linealmente el producto de ADN. Los primers utilizados son los mismos que en la amplificación pero a una concentración de trabajo de 1 μ M.

La preparación de la reacción de secuenciación se llevó a cabo en un volumen final de 8 μ l, de los cuales:

- 1,5 μ l de kit de secuenciación BigDye v1.1.
- 2 μ l de primer forward (1 μ M) o 3 μ l de primer reverse (1 μ M).
- 2 μ l de producto de PCR purificado.
- 1,5 μ l de agua estéril.
- 1 μ l de buffer 10x BigDye v1.1 & v3.1.

La reacción de secuenciación se realizó según las condiciones que aparecen en la Figura IV.2.

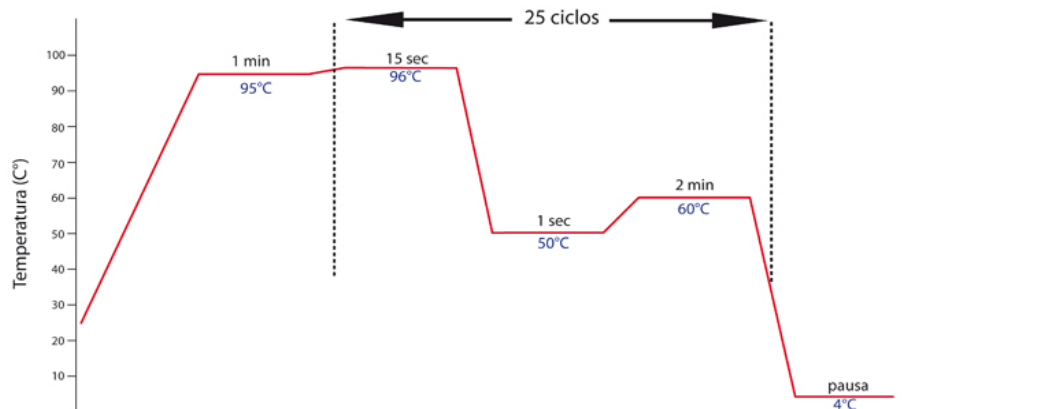


Figura IV.2: Condiciones de la reacción de secuenciación

2.2.5 Purificación del Producto de Secuenciación

Una vez finalizado el proceso, el siguiente paso es la purificación del producto de secuenciación para tratar de eliminar el exceso de primers y los terminadores no incorporados, que de no ser eliminados convenientemente podrían enmascarar la secuencia de ADN durante la electroforesis complicando así la interpretación de los resultados.

El protocolo empleado en este caso fue la purificación por centrifugación mediante columnas DTR o Dye Terminator Removal (Performa DTR Gel Filtration Cartridges, Edge Biosystems).

2.2.6 Electroforesis Capilar

Todas las muestras secuenciadas se han sometido a un proceso de electroforesis capilar mediante el equipo automatizado ABI Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems. Foster City, CA) en el que los diferentes fragmentos de ADN se separan en función del tamaño. Las muestras se prepararon convenientemente antes de introducirlas en el equipo añadiendo 10 μ l formamida Hi-Di (Applied Biosystems. Foster City, CA) al producto recuperado de las columnas anteriores.

El proceso de electroforesis es similar al comentado en los capítulos II y III.

Finalmente los datos generados (raw data) son analizados por el software. En este caso, el software que se emplea es el Sequencing Analysis Software que asignará los nucleótidos en cada posición en la secuencias obtenida.

Para la comparación de los datos analizados con la secuencia de referencia se usó SeqScape® Software for Mutation Profiling v2.6 (Applied Biosystems), a través del alineamiento simple de secuencias. Los resultados se comparan con la regiones HV1 y HV2 de la secuencia de Anderson [12] para individualizar cada polimorfismo.

2.3 ANÁLISIS ESTADÍSTICO DE LA MUESTRA

- El cálculo de las frecuencias de los haplotipos encontrados en cada una de las poblaciones se realizó mediante una hoja de cálculo (Microsoft Office Excel 2007). Del mismo modo se calculó el número de polimorfismos presentes.
- Mediante Arlequin v3.5 [51] se compararon los 6 grupos poblacionales, calculando los índices de diversidad molecular, la distribución *mismatch* (distribución de los emparejamientos erróneos), el número de posiciones polimórficas y el número de haplotipos a partir de la región HV1.
- Los haplogrupos se calcularon mediante distintas aplicaciones, disponibles on-line: Haplogroup Prediction Tool (National Geographic's Project); <http://nmhgtool.nationalgeographic.com/> Aplo v1.0.3, disponible en (http://www.doppiovu3.it/mtdna/aplo/aplo_en.htm) mtDNAManager (http://mtmanager.yonsei.ac.kr/search_sample.php) aplicación on-line Haplogrep (haplogrep.uibk.ac.at/) [52] basado en Phylotree.org [2]; disponible en www.phylotree.org. Los resultados obtenidos por el programa son confirmados mediante el cálculo de presencia o ausencia de las posiciones diagnósticas que darán el haplogrupo definitivo [53].
- El resto de los software utilizados, para calcular las distancias, análisis de la varianza molecular (AMOVA), redes, ancestros comunes y la

estructura de la población fueron ya mencionados en capítulos anteriores: Arlequin v3.5: [51] Network: [54] (DNA Alignment® Fluxus Technology), Phylip v3.66: [55,56] Treeview (Win16 [57]; Estadística v8: StatSoft® (StatSoft ibérica) SPSS v15.0: STRUCTURE 2.3.1 [58-61].

El estudio de distancias con otras poblaciones se realizó con poblaciones cercanas a la población de Guatemala, tomadas de la bibliografía que se muestran en la tabla IV.1.

El análisis de las secuencias se realiza mediante Arlequin v3.5, con este software se pudo realizar el análisis de las distancias genéticas. Posteriormente, las distancias obtenidas son sometidas a un análisis multidimensional mediante el programa SPSS v15.0 para poder visualizar los resultados obtenidos en Arlequin.

Tabla IV.1: Poblaciones, abreviatura, lengua, situación geográfica, número de individuos y autor del estudio publicado.

Población	Código	Lengua	País	Localización	N	Bibliografía
Mestizo	Mestizo	Español	Guatemala	Centroamérica	105	Estudio Actual
Kaqchiquel	Kaqchiquel	Amerindio	Guatemala	Centroamérica	49	Estudio Actual
K'iche'	Kiche	Amerindio	Guatemala	Centroamérica	48	Estudio Actual
Mam	Mam	Amerindio	Guatemala	Centroamérica	47	Estudio Actual
Q'eqchi'	Qeqchi	Amerindio	Guatemala	Centroamérica	47	Estudio Actual
Huetar	Huetar	Chibchan-Paezan	Costa Rica	Centroamérica	27	Santos et al. 1994 [62]
Kuna	Kuna	Chibchan-Paezan	Panamá	Centroamérica	63	Batista et al. 1995 [63]
Ngöbe	Ngobe	Chibchan-Paezan	Panamá	Centroamérica	46	Kolman et al. 1995 [64]
Bella Coola	BellaCo	Amerindio	Canadá	Norteamérica	40	Ward et al.1993 [65]
Nuu-Chah-Nulth	Nuu_Chah	Amerindio	Canadá	Norteamérica	63	Ward et al.1993 [65]
Haida	Haida	Na-Dene	EEUU	Norteamérica	25	Torrioni et al.1993 [31]
Indios Yakima	Yakima	Amerindio	EEUU	Norteamérica	42	Shields et al. 1993 [66]
Circumpolar populations	Circump	Eskimo	Groenlandia	Norteamérica	90	Shields et al. 1993 [66]
Inuit	Inuit	Eskimo	Groenlandia	Norteamérica	82	Saillard at al. 2000 [67]
Na-Dene	Na_Dene	Na-Dene	EEUU	Norteamérica	109	Malhi et al. 2002 [68]
Nativos del Nor-Centro de Mexico	Nt_NCMx	Amerindio	México	Norteamérica	223	Green et al.2000 [69]
Indios Mapuche	Mapuche	Andino	Patagonia Argentina	Sudamérica	39	Ginther et al. 1993 [70]
Este de Brasil	E_Brasil	Amerindio	Brasil	Sudamérica	82	Alves-Silva et al.2000 [71]
Nativos de Brasil	Nt_Brasil	Ecuatorial-Tucano	Brasil	Sudamérica	89	Ward et al. 1996 [72]
Gavião	Gaviao	Ecuatorial-Tucano	Brasil	Sudamérica	27	Ward et al. 1996 [72]
Xavante	Xavante	Ge-Pano-Caribeño	Brasil	Sudamérica	25	Ward et al. 1996 [72]
Yanomama	Yanom	Chibchan-Paezan	Brasil	Sudamérica	24	Torrioni et al. 1993 [31]
Zoró	Zoro	Ecuatorial-Tucano	Brasil	Sudamérica	30	Ward et al. 1996 [72]
Cayapa	Cayapa	Chibchan-Paezan	Ecuador	Sudamérica	204	Rickards et al. 1999 [73]
Tayacaja	Tayac	Andino	Perú	Sudamérica	61	Fuselli et al. 2003 [74]
Arequipa	Areq	Andino	Perú	Sudamérica	22	Fuselli et al. 2003 [74]
San Martín	S_Martin	Andino	Perú	Sudamérica	22	Fuselli et al. 2003 [74]
Piemonte	Piemon	Ecuatorial-Tucano	Bolivia	Sudamérica	46	Corella et al. 2007 [75]
Moxos	Moxos	Ecuatorial-Tucano	Bolivia	Sudamérica	54	Bert et al. 2004 [76]
Gran Chaco	G_Chaco	Ge-Pano-Caribeño	Frontera entre Argentina y Bolivia	Sudamérica	204	Cabana et al. 2006 [77]
Nativos de Chile	Nt_Chile	Andinos	Chile	Sudamérica	74	Moraga et al. 2000 [78]

3. RESULTADOS Y DISCUSIÓN

3.1 ESTUDIO HAPLOTÍPICO

Se caracterizaron 296 individuos procedentes de Guatemala y 295 individuos españoles, obteniendo los haplotipos para cada una de las regiones HV1 y 2. Se encontraron un total de 205 haplotipos únicos en la población de Guatemala. Los resultados se muestran en la tabla IV.2.

Tabla IV.2: Estudio haplotípico de las poblaciones.

POBLACIÓN	N	p, Probabilidad de coincidencia.	Diversidad Haplotípica $(n/n-1)(1-\Sigma p_i^2)$	Capacidad de discriminación $(n^\circ \text{ haplotipos}/n) \times 100$
Españoles	295	0,003-0,033	0,9769	78,31%
Mestizos	105	0,013-0,057	0,9943	71,43%
Kaqchiquel	49	0,020-0,061	0,9928	81,63%
K'iche'	48	0,021-0,083	0,9926	81,25%
Mam	47	0,021-0,084	0,9926	85,11%
Q'eqchi'	47	0,021-0,063	0,9889	78,72%

Mediante Arlequin v3.5 se calcularon las posiciones polimórficas (sustituciones) encontradas en la región HV1 de las 6 poblaciones, distinguiendo entre transversiones y transiciones. También se calcularon los índices de diversidad poblacional (dentro de cada población): H, diversidad haplotípica y π , número de diferencias de emparejamiento (number of pairwise differences). Todos los resultados se muestran en la tabla IV.3.

Tabla IV.3: Estudio haplotípico. Excepto para el número de haplotipos encontrados, los datos obtenidos se refieren a HV1

Población	n	Haplot.	Trans.	Trasv.	Sustit.	Polimorf.	H	π
Kaqchiquel	49	40	41	4	45	44	33±17	5,8±3,1
Mam	47	40	26	2	28	28	21±8	4,8±2,6
K'iche'	48	39	42	5	47	45	24±13	5,5±3,0
Q'eqchi'	47	37	33	5	38	38	48±26	5,5±3,0
Mestizos	105	75	46	6	52	49	44±12	5,9±3,1
Españoles	295	230	96	13	109	102	35±8	4,6±2,5

Los haplotipos encontrados en cada una de las poblaciones de Guatemala se muestran en la siguientes tablas (IV.4, 5, 6, 7 y 8).

Tabla IV.4: Haplotipos de la población mestiza. Nombre de la muestra (Individuo), haplogrupo (Hg), número de individuos que comparten el haplotipo (N) y haplotipo (Polimorfismos encontrados en HV1, HV2).

Individuo	Hg	N	Polimorfismos encontrados en HV1 , HV2
Mest-088-094-103-301-386-499	A2	6	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mest-006-401-451-496	A2	4	16111T, 16223T, 16290T, 16319A, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mest-270-302-307-467	A2	4	16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-003-328-396	B4	3	16183C, 16189C, 16217C, 73G, 263G, 309.1C, 315.1C.
Mest-038-281-428.	A2	3	16111T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mest-039-080-469	A2	3	16111T, 16209A, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mest-070-468-481	B4	3	16092C, 16182C, 16183C, 16189C, 16217C, 73G, 263G, 309.1C, 315.1C.
Mest-344-487-495	A2q	3	16111T, 16187T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-378-470-497	A2q	3	16111T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-014-482	A2	2	16111T, 16187T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 234G, 235G, 263G, 309.2C, 315.1C.
Mest-016-017	A2	2	16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mest-040-425	A2	2	16104T, 16111T, 16223T, 16290T, 16311C, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.2C, 315.1C.
Mest-162-172	A4	2	16223T, 16290T, 16319A, 16362C, 73G, 146C, 235G, 263G, 315.1C.
Mest-426-449	A2	2	16111T, 16223T, 16290T, 16319A, 16360T, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 309.2C, 315.1C.
Mest-439-498	A2	2	16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 263G, 309.1C, 315.1C.
Mest-492-494	R8	2	16182C, 16183C, 16189C, 73G, 143A, 146C, 195C, 263G, 309.2C, 315.1C.
Mest-001	A2	1	16111T, 16189C, 16223T, 16290T, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-010	A2q	1	16111T, 16140C, 16192T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-012	A2	1	16111T, 16223T, 16290T, 16299G, 16319A, 16362C, 73G, 146C, 153G, 197G, 235G, 263G, 309.1C, 315.1C.
Mest-013	A2	1	16111T, 16129A, 16189C, 16223T, 16234T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mest-018	C	1	16129A, 16183C, 16189C, 73G, 195C, 263G, 309.1C, 315.1C.
Mest-041	C1	1	16183C, 16189C, 16223T, 16259A, 16298C, 16325C, 16327T, 73G, 249del, 251A, 263G, 290del, 291del, 309.1C, 315.1C.
Mest-042	A2	1	16111T, 16223T, 16187T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 263G, 309.1C, 315.1C.
Mest-066	H2a2b	1	16291T, 16298C, 195C, 263G, 309.1C, 315.1C.
Mest-109	A2	1	16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 235G, 263G, 309.2C, 315.1C.
Mest-203	B	1	16092C, 16182C, 16183C, 16189C, 73G, 152C, 263G, 309.1C, 315.1C.
Mest-204	B	1	16183C, 16189C, 73G, 195C, 263G, 309.2C, 315.1C.
Mest-213	A2	1	16111T, 16156A, 16183C, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-219	A2	1	16111T, 16189C, 16223T, 16290T, 16299G, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-232	C1	1	16183C, 16189C, 16223T, 16325C, 16327T, 73G, 249del, 263G, 290del, 291del, 309.1C, 315.1C.
Mest-233	H3d	1	73G, 146C, 263G, 315.1C.
Mest-241	A4b	1	16111T, 16129A, 16136C, 16183C, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 235G, 263G, 309.2C, 315.1C.
Mest-245	C1	1	16223T, 16298C, 16325C, 16327T, 73G, 153G, 249del, 263G, 290del, 291del, 309.1C, 315.1C.

Mest-248	A2	1	16111T, 16223T, 16235G, 16290T, 16319A, 16360T, 16362C, 73G, 146C, 152C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-249	A2	1	16111T, 16223T, 16290T, 16319A, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-251	B4	1	16183C, 16189C, 16217C, 16249C, 73G, 263G, 309.1C, 315.1C.
Mest-254	A2	1	16111T, 16223T, 16225T, 16319A, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-269	B4	1	16169T, 16183C, 16189C, 16217C, 16259T, 73G, 103A, 153G, 263G, 315.1C.
Mest-276	L3e1d	1	16176T, 16223T, 16327T, 73G, 150T, 152C, 189G, 194T, 200G, 263G, 315.1C.
Mest-278	A4d	1	16223T, 16240C, 16290T, 16319A, 16362C, 73G, 146C, 151T, 235G, 249G, 263G, 309.1C, 315.1C.
Mest-284	A2	1	16111T, 16223T, 16290T, 16299G, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-288	B4	1	16104T, 16183C, 16189C, 16217C, 16362C, 73G, 204C, 263G, 309.1C, 315.1C.
Mest-299	J1c	1	16069T, 16126C, 16311C, 73G, 185A, 228A, 263G, 295T, 315.1C.
Mest-321	A2	1	16111T, 16187T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 263G, 309.2C, 315.1C.
Mest-325	L3e2a	1	16223T, 16269G, 16320T, 73G, 150T, 195C, 198T, 263G, 315.1C.
Mest-355	A2	1	16111T, 16136C, 16223T, 16290T, 16294T, 16311C, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mest-365	A2	1	16111T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-375	A2	1	16111T, 16223T, 16259T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-377	B2 +16311	1	16183C, 16189C, 16217C, 16311C, 73G, 263G, 275A, 309.1C, 315.1C.
Mest-383	B4	1	16129A, 16182C, 16183C, 16189C, 16217C, 16261T, 73G, 263G, *
Mest-394	A2	1	16111T, 16223T, 16290T, 16319A, 16360T, 16362C, 73G, 153G, , 235G, , 263G, , 309.2C, , 315.1C.
Mest-398	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 235G, 263G, 309.2C, 315.1C.
Mest-399	A2d1a	1	16111T, 16136C, 16223T, 16274A, 16290T, 16319A, 16362C, 73G, 153G, 235G, 263G, 315.1C.
Mest-400	A2	1	16111T, 16223T, 16290T, 16319A, 16360T, 16362C, 73G, 146C, 153G, 263G, 309.2C, 315.1C.
Mest-424	A2	1	16111T, 16223T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 164T, 309.1C, 315.1C.
Mest-430	C1b2	1	16104T, 16223T, 16274A, 16298C, 16325C, 16327T, 73G, 152C, 214G, 249del, 290del, 291del, 309.1C, 315.2C.
Mest-434	B	1	16124C, 16182C, 16183C, 16189C, 16311C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mest-440	B4	1	16182C, 16183C, 16189C, 16217C, 73G, 263G, 309.2C, 315.1C.
Mest-441	A2	1	16111T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 291G, 309.1C, 315.1C.
Mest-452	B4	1	16092C, 16182C, 16183C, 16189C, 16217C, 16218T, 16362C, 73G, 263G, 315.1C.
Mest-454	N	1	16223T, 16311C, 16325C, 73G, 150T, 185A, 189G, 263G, 315.1C.
Mest-458	A2	1	16111T, 16129A, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mest-463	C1	1	16129A, 16189C, 16209C, 16223T, 16298C, 16325C, 16327T, 73G, 249del, 263G, 290del, 291del, 309.1C, 315.1C.
Mest-471	A2	1	16111T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 204C, 235G, 263G, 315.1C.
Mest-472	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 200G, 235G, 263G, 309.1C, 315.1C.
Mest-478	A2	1	16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.2C, 315.1C.
Mest-479	A2	1	16111T, 16175G, 16223T, 16290T, 16300G, 16319A, 16362C, 73G, 146C, 153G, 194T, 235G, 263G, 309.1C, 315.1C.
Mest-480	B4	1	16129A, 16182C, 16183C, 16189C, 16217C, 16249C, 73G, 263G, 309.2C, 315.2C.

Mest-483	A2q	1	16111T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 234G, 235G, 263G, 309.1C, 315.1C.
Mest-484	J2a2	1	16069T, 16126C, 16241G, 73G, 150T, 195C, 263G, 295T, 315.1C.
Mest-485	B4	1	16104A, 16183C, 16189C, 16217C, 73G, 146C, 263G, 309.2C, 315.1C.
Mest-486	T2	1	16126C, 16294T, 16296T, 73G, 146C, 263G, 309.1C, 315.1C.
Mest-488	C1	1	16223T, 16256T, 16274A, 16298C, 16325C, 16327T, 73G, 204C, 214G, 249del, 263G, 290del, 291del, 309.2C, 315.1C.
Mest-491	HV2	1	16129A, 16183C, 16189C, 16217C, 16352C, 73G, 195C, 263G, 309.1C, 315.1C.
Mest-493	B2a	1	16111T, 16182C, 16183C, 16189C, 16217C, 73G, 152C, 228A, 263G, 315.1C.

Tabla IV.5: Haplotipos de la población Kaqchiquel.

Individuo	Hg	n	Polimorfismos encontrados en HV1 , HV2
Kaq-134, 183, 477	A2q	3	16111T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-069, 117, 131	A2	2	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Kaq-118, 179	A2	2	16111T, 16223T, 16290T, 16300G, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-126, 133	A2	2	16223T, 16240G, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-135, 168	A2	2	16111T, 16140C, 16223T, 16319A, 73G, 146C, 153G, 185A, 235G, 263G, 309.1C, 315.1C.
Kaq-184, 195	C1	2	16223T, 16256T, 16298C, 16325C, 16327T, 73G, 249d, 263G, 290d, 291d, 315.1C.
Kaq-475, 476	A2	2	16111T, 16223T, 16290T, 16319A, 73G, 146C, 153G, 235G, 263G, 315.1C.
Kaq-026	B	1	16092C, 16182C, 16183C, 16189C, 16289G, 73G, 146C, 183G, 263G, 310C, 314d, 315d.
Kaq-111	A2	1	16111T, 16187T, 16223T, 16270T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-115	C1	1	16093C, 16223T, 16298C, 16325C, 16327T, 73G, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Kaq-116	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 235G, 263G, 309.1C, 315.1C.
Kaq-119	A2	1	16111T, 16223T, 16290T, 16300G, 16319A, 16362C, 73G, 146C, 153G, 194T, 235G, 263G, 309.1C, 315.1C.
Kaq-120	P4a	1	16092C, 16182C, 16183C, 16189C, 16127C, 16319A, 16362C, 73G, 244G, 263G, 315.1C.
Kaq-122	C1	1	16092C, 16223T, 16298C, 16325C, 16327T, 73G, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Kaq-123	A2	1	16111T, 16166G, 16223T, 16290T, 16299G, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-124	A2	1	16223T, 16240G, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Kaq-125	D1	1	16129A, 16223T, 16325C, 16362C, 73G, 263G, 315.1C.
Kaq-127	C1	1	16189C, 16221T, 16223T, 16325C, 16327T, 73G, 207A, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Kaq-128	C1	1	16126C, 16223T, 16298C, 16325C, 16327T, 73G, 222T, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Kaq-129	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-130	A2	1	16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-132	D1	1	16223T, 16291T, 16311C, 16325C, 16362C, 73G, 214G, 263G, 267C, 315.1C.
Kaq-164	A2	1	16092C, 16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-165	A2	1	16111T, 16129A, 16223T, 16290T, 16311C, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-167	A2	1	16223T, 16240G, 16278T, 16290T, 16299G, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-169	A2	1	16223T, 16240G, 16290T, 16319A, 16362C, 73G, 114T, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-170	B4	1	16129A, 16182C, 16183C, 16189C, 16127C, 16217C, 16249C, 73G, 193G, 263G, 315.1C.
Kaq-173	B2+16311	1	16182C, 16183C, 16189C, 16217C, 16311C, 73G, 263G, 315.1C.

Kaq-174	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 150T, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-176	B4	1	16124C, 16183C, 16189C, 16217C, 73G, 263G, 315.1C.
Kaq-177	A2	1	16093C, 16111T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 195C, 235G, 263G, 309.1C, 315.1C
Kaq-181	A2	1	16111T, 16223T, 16261T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 235G, 263G, 309.1C, 315.1C.
Kaq-182	B4	1	16183C, 16189C, 16217C, 16224C, 16259T, 16357C, 73G, 146C, 263G, 309.1C, 315.1C.
Kaq-191	A2	1	16111T, 16213A, 16223T, 16290T, 16319A, 73G, 146C, 153G, 195C, 235G, 263G, 309.1C, 315.1C.
Kaq-192	B4	1	16183C, 16189C, 16217C, 73G, 263G, 309.1C, 315.1C.
Kaq-194	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 150T, 153G, 235G, 263G, 315.1C.
Kaq-196	A	1	16111T, 16136C, 16186T, 16223T, 16290T, 16319A, 16355T, 73G, 146C, 235G, 263G, 309.1C, 315.1C.
Kaq-227	A2r	1	16111T, 16223T, 16239T, 16290T, 16319A, 16362C, 73G, 143A, 146C, 152C, 153G, 235G, 263G, 309.1C, 315.1C
Kaq-473	A2	1	16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Kaq-474	A2q	1	16111T, 16140C, 16192T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235R, 263G, 309.1C, 315.1C

Tabla IV.6: Haplotipos de la población K'iche'.

Individuo	Hg	n	Polimorfismos encontrados en HV1 , HV2
Kic-137, 140, 148, 226	A2	4	16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-049, 056	A2	2	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 214G, 235G, 263G, 315.1C.
Kic-082, 200	A2q	2	16111T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Kic-086, 228	A2	2	16111T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-087, 139	A2	2	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 214G, 235G, 263G, 309.1C, 315.1C.
Kic-171, 209	A2	2	16111T, 16189C, 16223T, 16290T, 16319A, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-187, 218	A2	2	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Kic-085	A2	1	16092C, 16111T, 16183C, 16189C, 16223T, 16259T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-136	A2	1	16111T, 16147T, 16187T, 16223T, 16270T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-141	A2	1	16223T, 16290T, 16319A, 73G, 143A, 146C, 152C, 153G, 235G, 263G, 315.1C.
Kic-145	B4	1	16183C, 16184T, 16185T, 16189C, 16217C, 73G, 152C, 263G, 309.1C, 315.1C.
Kic-146	B4	1	16092C, 16182C, 16183C, 16189C, 16217C, 73G, 228T, 263G, 315.1C.
Kic-151	A2	1	16111Y, 16223T, 16290Y, 16319A, 73G, 146C, 153G, 235G, 263G, 315.1C.
Kic-152	A2q	1	16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-153	A2	1	16111T, 16172C, 16186T, 16223T, 16290T, 16319A, 16360T, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Kic-163	A2	1	16095T, 16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-166	C1	1	16223T, 16274A, 16298C, 16325C, 73G, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Kic-172	A2	1	16129A, 16168T, 16223T, 16240G, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-185	A2	1	16111T, 16223T, 16290T, 16319A, 73G, 146C, 153G, 183G, 195C, 235G, 263G, 283C, 284C, 290d, 291d, 309.1C, 315.1C
Kic-186	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-190	C1d	1	16051G, 16142T, 16192T, 16223T, 16292T, 16298C, 16325C, 16327T, 73G, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Kic-193	C1d+194	1	16051G, 16223T, 16260T, 16265G, 16298C, 16325C, 16327T, 73G, 194T, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Kic-197	A2q	1	16187T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Kic-198	A4	1	16223T, 16290T, 16311C, 16319A, 16362C, 73G, 146C, 182T, 235G, 263G, 315.1C.

Kic-202	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 195C, 235G, 263G, 315.1C.
Kic-207	A2b1	1	16111T, 16223T, 16265G, 16290T, 16319A, 16362C, 73G, 146C, 235G, 263G, 315.1C.
Kic-212	B4	1	16183C, 16189C, 16217C, 16259T, 16357C, 73G, 263G, 315.1C.
Kic-214	B	1	16182C, 16183C, 16189C, 73G, 263G, 309.1C, 315.1C.
Kic-215	A2	1	16111T, 16136C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-217	B2+16311	1	16173T, 16183C, 16189C, 16217C, 16311C, 16334A, 16352C, 73G, 195C, 263G, 309.1C, 315.1C.
Kic-221	A2	1	16111T, 16223T, 16272G, 16290T, 16319A, 16360T, 16362C, 73G, 146C, 150T, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-222	A2	1	16111T, 16189C, 16223T, 16290T, 16319A, 73G, 146C, 153G, 235G, 263G.
Kic-224	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 200G, 235G, 263G, 309.1C, 315.1C.
Kic-229	A2	1	16111T, 16265C, 16290T, 16293G, 16311C, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-237	A2	1	16223T, 16240G, 16290T, 16319A, 16362C, 73G, 146C, 153G, 195C, 235G, 263G, 309.1C, 315.1C.
Kic-243	B4 +16261 +16129	1	16129A, 16182C, 16183C, 16189C, 16217C, 16261T, 73G, 146C, 153G, 263G, 309.1C, 315.1C.
Kic-247	A2q	1	16111T, 16187T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 195C, 235G, 263G, 309.1C, 315.1C.
Kic-256	A2	1	16093C, 16223T, 16290T, 16293G, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Kic-258	C	1	16129A, 16182C, 16183C, 16189C, 16249C, 73G, 263G, 315.1C.

Tabla IV.7: Haplotipos de la población Mam.

Individuo	Hg	n	Polimorfismos encontrados en HV1 , HV2
Mam-236, 360, 361	A2	3	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mam-240, 372	A2	2	16111T, 16223T, 16259T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mam-362, 414	A2	2	16111T, 16223T, 16290T, 16319A, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mam-367, 368	A2	2	16111T, 16223T, 16235G, 16290T, 16319A, 16362C, 73G, 146C, 182T, 235G, 263G, 315.1C.
Mam-395, 412	B4	2	16092C, 16182C, 16183C, 16189C, 16217C, 73G, 263G, 309.1C, 315.1C.
Mam-402, 410	C1c+16319	2	16223T, 16274A, 16298C, 16319A, 16325C, 16327T, 73G, 214G, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Mam-363	A2	1	16111T, 16223T, 16290T, 16319A, 73G, 146C, 153G, 183G, 195C, 235G, 263G, 315.1C.
Mam-364	A2	1	16111T, 16223T, 16259T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mam-369	A2	1	16111T, 16223T, 16290T, 16319A, 97A, 106d, 107d, 108d, 109d, 110d, 111d, 146C, 153G, 235G, 263G, 315.1C.
Mam-370	A2	1	16111T, 16213A, 16223T, 16290T, 16319A, 73G, 146C, 153G, 195C, 235G, 263G, 309.1C, 315.1C.
Mam-371	A2	1	16111T, 16223T, 16259T, 16290T, 16319A, 16362C, 73G, 97A, 146C, 153G, 263G, 315.1C.
Mam-373	A2	1	16111T, 16187T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 263G.
Mam-374	C1	1	16186T, 16293G, 16298C, 16325C, 16327T, 73G, 146C, 174T, 199C, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Mam-376	C1c+16319	1	16086C, 16223T, 16319A, 73G, 194T, 249d, 263G, 290d, 291d.
Mam-379	C	1	16086C, 16223T, 16319A, 73G, 114T, 146C, 153G, 182T, 235G, 263G, 315.1C.
Mam-380	A4b	1	16183C, 16189C, 16264T, 16290T, 16319A, 16362C, 73G, 263G.
Mam-381	A2	1	16111T, 16223T, 16264T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mam-382	A2	1	16111T, 16223T, 16290T, 16319A, 73G, 146C, 153G, 263G, 309.1C, 315.1C.
Mam-384	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 226C, 235G, 263G, 309.1C, 315.1C.
Mam-385	A2	1	16111T, 16223T, 16290T, 16319A, 73G, 146C, 153G, 183G, 195C, 235G, 263G, 309.1C, 315.1C.
Mam-388	A2q	1	16111T, 16129A, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 263G.,
Mam-389	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 214G, 235G, 263G, 309.1C, 315.1C.
Mam-390	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 111G, 146C, 153G, 195C, 235G, 263G, 309.1C, 315.1C.
Mam-391	A2	1	16111T, 16223T, 16290T, 16295T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mam-392	C1	1	16223T, 16274A, 16298C, 16325C, 16327T, 16355T, 73G, 249d, 263G, 290d, 291d,

			309.1C, 315.1C.
Mam-393	A2	1	16111T, 16183C, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mam-397	A2	1	16111T, 16223T, 16274A, 16290T, 16319A, 16355T, 16362C, 73G, 146C, 152C, 153G, 214G, 235G, 263G, 309.1C, 315.1C.
Mam-403	C1	1	16186T, 16223T, 16293G, 16298C, 16325C, 16327T, 73G, 146C, 174T, 199C, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Mam-405	C1	1	16223T, 16298C, 16325C, 16327T, 73G, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Mam-407	C1	1	16223T, 16298C, 16325C, 73G, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Mam-408	A2	1	16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G.
Mam-409	A2	1	16093C, 16223T, 16290T, 16293G, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G.
Mam-411	A2	1	16111T, 16187T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mam-415	A2	1	16111T, 16129A, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Mam-416	C1	1	16223T, 16293G, 16298C, 16325C, 16327T, 73G, 174T, 199C, 249d, 263G, 290d, 291d, 315.1C.
Mam-417	A2q	1	16111T, 16129A, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G.
Mam-418	A2	1	16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mam-419	A2	1	16111T, 16187T, 16189C, 16223T, 16290T, 16319A, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mam-420	A2	1	16111T, 16223T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Mam-423	A2	1	16111T, 16223T, 16287T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.

Tabla IV.8: Haplotipos de la población Q'eqchi'.

Individuo	Hg	n	Polimorfismos encontrados en HV1 , HV2
Qe-266, 272, 273	A2q	3	16111T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-283, 306, 312	A2	3	16111T, 16223T, 16266T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-259, 262	A2	2	16223T, 16290T, 16319A, 16352C, 16362C, 73G, 146C, 153G, 182T, 235G, 263G, 315.1C.
Qe-260, 319	A4	2	16223T, 16240G, 16274A, 16287T, 16290T, 16295T, 16319A, 16362C, 73G, 146C, 235G, 263G, 309.1C, 315.1C.
Qe-282, 345	A2	2	16111T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Qe-285, 294	C1	2	16223T, 16274A, 16298C, 16325C, 16327T, 16355T, 73G, 207A, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Qe-286, 348	A2	2	16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-318, 358	C1	2	16223T, 16298C, 16325C, 16327T, 73G, 249d, 263G, 290d, 291d, 309.1C, 315.1C.
Qe-263	A2	1	16111T, 16223T, 16290T, 16319A, 73G, 146C, 153G, 226C, 235G, 263G, 315.1C.
Qe-264	D4h3a5	1	16223T, 16241G, 16300.1A, 16342C, 16346A, 16362C, 73G, 152C, 263G, 315.1C.
Qe-265	B4	1	16183C, 16189C, 16217C, 16258C, 73G, 152C, 210G, 263G, 290d, 291d.
Qe-271	A4b	1	16189C, 16223T, 16290T, 16319A, 16352C, 16362C, 73G, 146C, 153G, 182T, 235G, 263G, 309.1C, 315.1C.
Qe-274	A2	1	16111T, 16223T, 16294T, 16362C, 73G, 146C, 150T, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-275	A2	1	16104T, 16111T, 16223T, 16290T, 16311C, 16319A, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-279	C1	1	16223T, 16298C, 16325C, 16327T, 73G, 249d, 263G, 288G, 290d, 291d, 309.1C, 315.1C.
Qe-280	A2	1	16111T, 16223T, 16266T, 16290T, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-287	A2	1	16111T, 16223T, 16219T, 16319A, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-289	A2q	1	16111T, 16187T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-290	A2q	1	16111T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Qe-291	A2q	1	16111T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 150T, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-292	H2a1	1	16108T, 16189C, 16354T, 263G, 315.1C.
Qe-295	B2a	1	16111T, 16145A, 16183C, 16189C, 16217C, 16258C, 73G, 143A, 146C, 153G, 235G, 263G, 291G, 309.1C, 315.1C, 338T.

Qe-300	A2	1	16223T, 16290T, 16319A, 16362C, 73G, 146C, 150T, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-303	A2	1	16111T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-304	A2	1	16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 182T, 184A, 204C, 235G, 263G, 309.1C, 315.1C.
Qe-308	A2	1	16111T, 16223T, 16274A, 16290T, 16319A, 16362C, 73G, 146C, 150T, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-309	B4	1	16182C, 16183C, 16189C, 16217C, 73G, 263G, 315.1C.
Qe-310	B4	1	16095T, 16183C, 16189C, 16217C, 73G, 263G, 309.1C, 315.1C.
Qe-311	A2	1	16111T, 16183C, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Qe-313	A4d	1	16223T, 16240G, 16290T, 16319A, 16362C, 73G, 146C, 151T, 235G, 249G, 263G, 309.1C, 315.1C.
Qe-314	A2	1	16094Y, 16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-315	A2b1	1	16111T, 16129A, 16192T, 16223T, 16265G, 16290T, 16311C, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.
Qe-316	A2	1	16111T, 16189C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-320	A2	1	16092C, 16111T, 16187T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-330	A2q	1	16111T, 16209C, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-343	A2	1	16111T, 16129A, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.
Qe-347	A2	1	16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 226C, 235G, 263G, 315.1C.

Mediante Arlequin v3.5 también se estudiará la estructura de cada población realizando un análisis mismatch. La distribución de mismatch tiende a ser multimodal en poblaciones en equilibrio demográfico, en cambio es unimodal en aquellas que han experimentado una expansión demográfica [79]. Como se puede ver en las figuras IV.3, 4, 5 y 6, las poblaciones mayas se encuentran en equilibrio demográfico según este análisis; resultado que responde a las características de estas poblaciones, muy antiguas, aisladas y con flujos poblacionales muy limitados. En la población mestiza se observa una distribución similar a las poblaciones mayas, como se ve en la figura IV.7. En cambio, en la población española, figura IV.8, se aprecia una distribución unimodal, lo que indica que está en expansión. Después de los resultados obtenidos en los capítulos anteriores, puede parecer extraño que la población mestiza se encuentre en equilibrio, y no como la española, en expansión. Como ya se verá más adelante, los resultados obtenidos en ADNmt difieren de los obtenidos en aSTRs y Y-STRs, ya que el ADNmt conserva una composición muy similar a la encontrada en la población Maya.

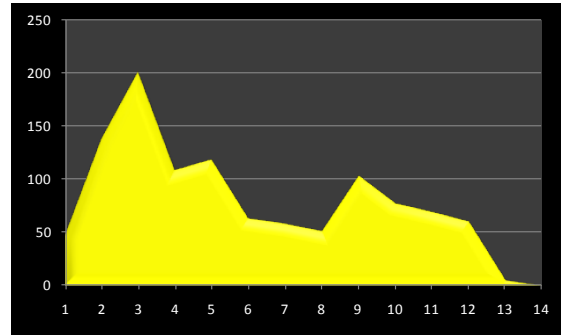
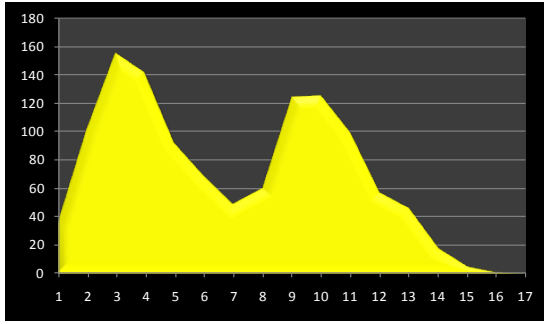


Figura IV.3 y 4 Distribución mismatch de la población K'iche' (derecha) y Mam (izquierda).

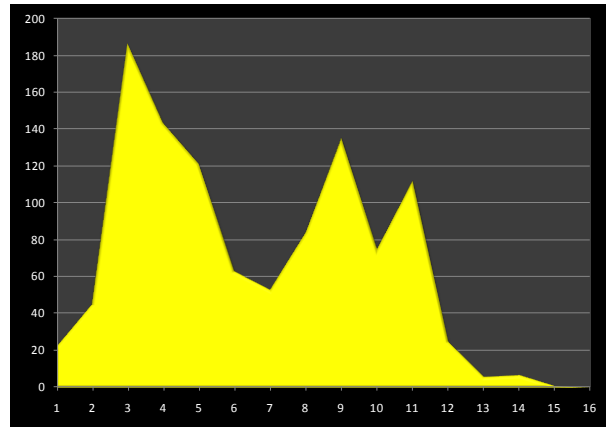
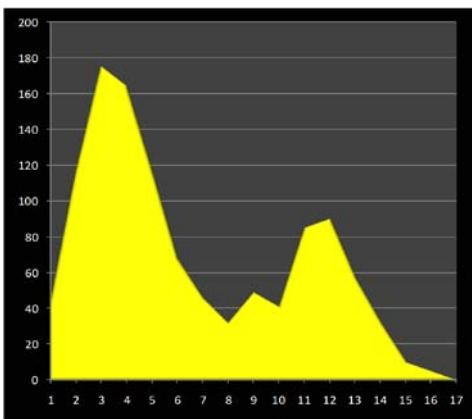


Figura IV.5 y 6 Distribución mismatch de la población K'iche' (derecha) y Q'eqchi' (izquierda).

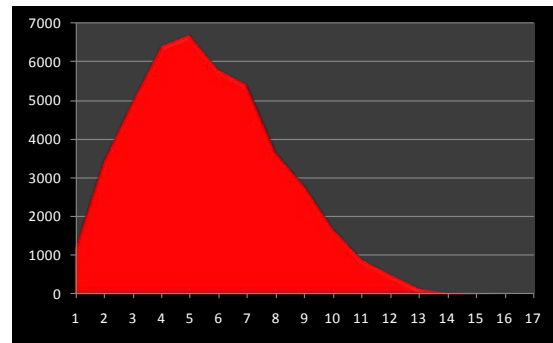
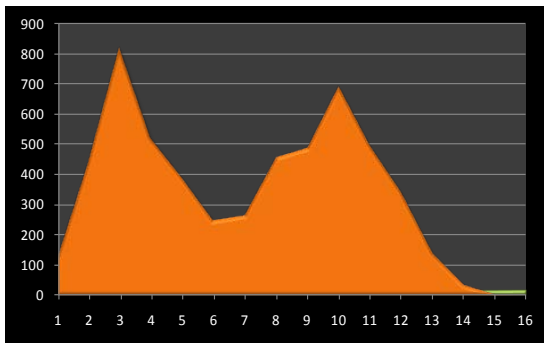


Figura IV.7 y 8 Distribución mismatch de la población mestiza (derecha) y española (izquierda).

3.2 CÁLCULO DE HAPLOGRUPOS

Los haplogrupos determinados por los distintos programas son los que se presentan en la tabla IV.9 y en las figuras IV.10 y IV.11.

Tabla IV.9: Haplogrupos encontrados en las poblaciones estudiadas.

Haplogrupo / Población		Kaqchiquel	K'iche'	Mam	Q'eqchi'	Mestizos	Españoles	
A	A2	A2b1	0,0000	0,0210	0,0000	0,0210	0,0000	0,0000
		A2d1a	0,0000	0,0000	0,0000	0,0000	0,0100	0,0000
		A2q	0,0610	0,1040	0,0430	0,1490	0,0760	0,0000
		A2r	0,0200	0,0000	0,0000	0,0000	0,0000	0,0000
			0,5920	0,6460	0,6810	0,5110	0,5240	0,0030
	A4	A4b	0,0000	0,0000	0,0210	0,0210	0,0100	0,0000
		A4d	0,0000	0,0000	0,0000	0,0210	0,0100	0,0000
			0,0000	0,0210	0,0000	0,0210	0,0000	0,0000
		0,0200	0,0000	0,0000	0,0000	0,0000	0,0000	
B	B2	B2a	0,0000	0,0000	0,0000	0,0210	0,0100	0,0000
				0,0200	0,0210	0,0000	0,0000	0,0100
	B4	B4a3	0,0000	0,0000	0,0210	0,0000	0,0000	0,0000
				0,0820	0,0830	0,0210	0,0640	0,1430
C	C1	C1b2	0,0000	0,0000	0,0000	0,0000	0,0100	0,0000
		C1c	0,0000	0,0000	0,0640	0,0000	0,0000	0,0000
		C1d	0,0000	0,0420	0,0000	0,0000	0,0000	0,0030
			0,1220	0,0210	0,1280	0,1060	0,0480	0,0000
D	D1		0,0410	0,0000	0,0000	0,0000	0,0000	0,0000
	D4	D4h3	0,0000	0,0000	0,0000	0,0210	0,0000	0,0000
H	H2	H2a1	0,0000	0,0000	0,0000	0,0210	0,0000	0,0000
		H2a2	0,0000	0,0000	0,0000	0,0000	0,0100	0,1210
	H3	H3d	0,0000	0,0000	0,0000	0,0000	0,0100	0,0630
				0,0000	0,0000	0,0000	0,0000	0,0000
HV	HV0		0,0000	0,0000	0,0000	0,0000	0,0100	0,0410
J	J1	J1c	0,0000	0,0000	0,0000	0,0000	0,0100	0,0380
		J2	J2a2	0,0000	0,0000	0,0000	0,0000	0,0190
L	L3	L3e	0,0000	0,0000	0,0000	0,0000	0,0190	0,0130
M	M1	M1a	0,0000	0,0210	0,0000	0,0000	0,0100	0,0030
	M32		0,0000	0,0000	0,0210	0,0000	0,0000	0,0000
N			0,0000	0,0000	0,0000	0,0000	0,0100	0,0220
U			0,0000	0,0000	0,0000	0,0000	0,0000	0,1750
Otros			0,0200	0,0210	0,0000	0,0000	0,0580	0,1870

El haplogrupo A es el mayoritario en las cuatro poblaciones maya. Este haplogrupo, nace en Asia y se extiende al continente americano aunque también aparece con frecuencia en pueblos siberianos. Muchos de los haplogrupos derivados de él, como el A2, son frecuentes en pueblos indígenas americanos, o subgrupos como el A2q que son también típicos de estos [80].

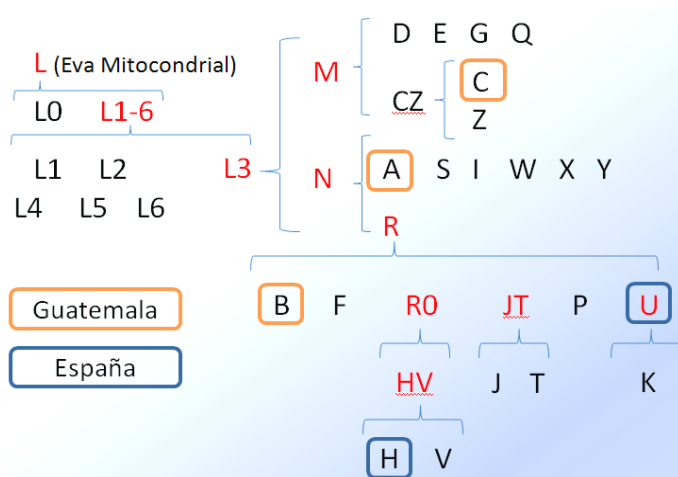


Figura IV.9 Filogenética de los haplogrupos mitocondriales, en rojo los haplogrupos ancestrales. Enmarcados en naranja los haplogrupos mayoritarios de la población Maya y Mestiza de Guatemala y en azul los mayoritarios de la población española.

A parte de este haplogrupos mayoritario en las poblaciones maya y mestiza de Guatemala, se encuentran otros que también son frecuentes en estas poblaciones indígenas [81], como son los haplogrupos C1 y el B4. Por otro lado, aparecen algunos individuos M, ancestros de los haplogrupos C y D. En algunas de las poblaciones mayas aparece el haplogrupos D, frecuente en indígenas del norte y sur de América, pero menos frecuente en las poblaciones de Centroamérica [81].

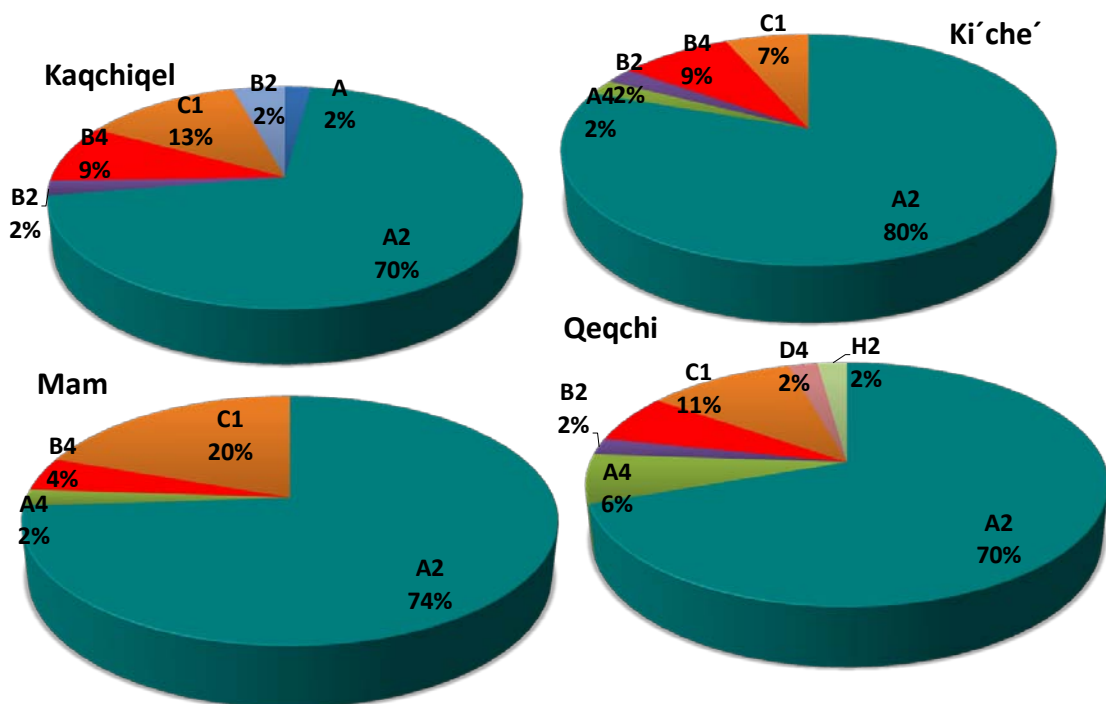


Figura IV.10: Gráficos de sectores de haplogrupos en las poblaciones Maya. Cada gráfico circular representa los haplogrupos obtenidos en un grupo poblacional.

En la población española en cambio, aparecen haplogrupos, de los conocidos como modernos, derivados del haplogrupo R. En un 40% de la población aparecen haplogrupos derivados del HV, H2, H3 y HV0. Otro de los haplogrupos más presentes son el U y, en menor medida, el T.

El haplogrupo H, es el haplogrupo más frecuente en Europa, con todas sus subclases. Este grupo es uno de los mejor definidos y a su vez refleja un mayor número de individuos y suele representar un 40-50% en las poblaciones caucásicas. Al estudiar este haplogrupo, se busca su aparición en la población mestiza, como haplogrupo mayoritario en las poblaciones europeas. Viendo los resultados se observa la pequeña proporción con la que ha contribuido la población europea a la población mestiza de Guatemala.

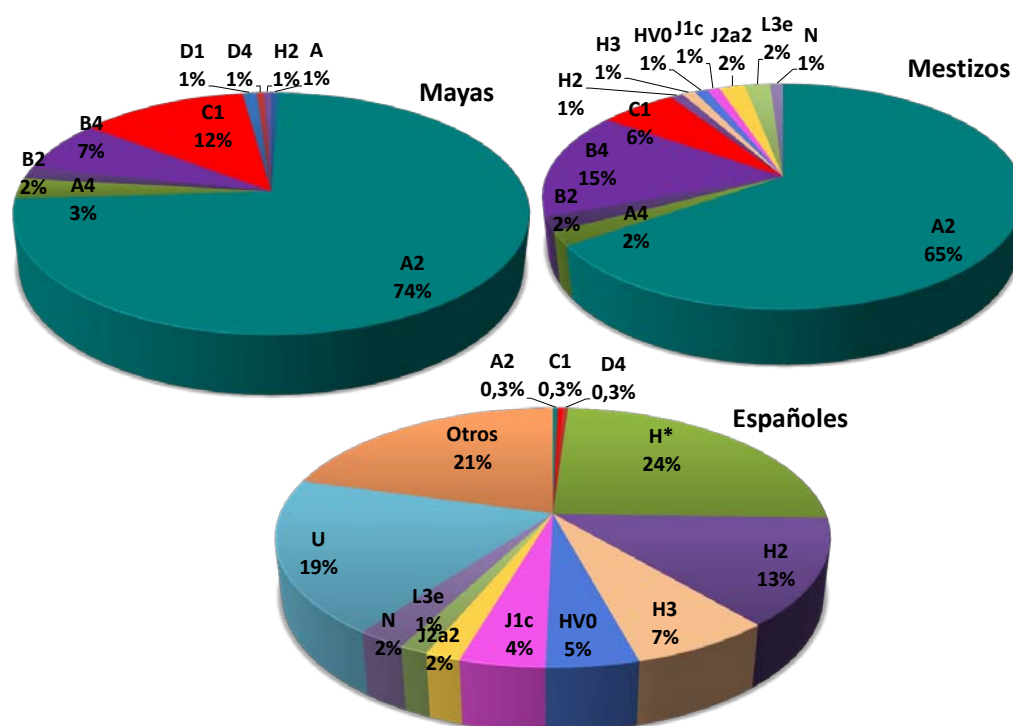


Figura IV.11: Gráficos de sectores de haplogrupos en las poblaciones estudiadas. El gráfico de la población maya es el resultado del conjunto de los cuatro subgrupos poblacionales.

Para poder comprender mejor como se distribuyen los individuos de estas poblaciones, se construyen árboles filogenéticos, basándonos en *PhyloTree*,

indicando, los porcentajes de cada haplogrupo que se han encontrado en esta población. Si se observan los árboles filogenéticos (fig. IV.12), se puede comprobar que la población maya, está formada por el haplogrupo más antiguo de los típicamente americanos. En la población mestiza, aunque aparecen nuevos haplogrupos fruto de las migraciones, se ve como el resultado es muy similar al que se encuentra en cada una de las poblaciones mayas.

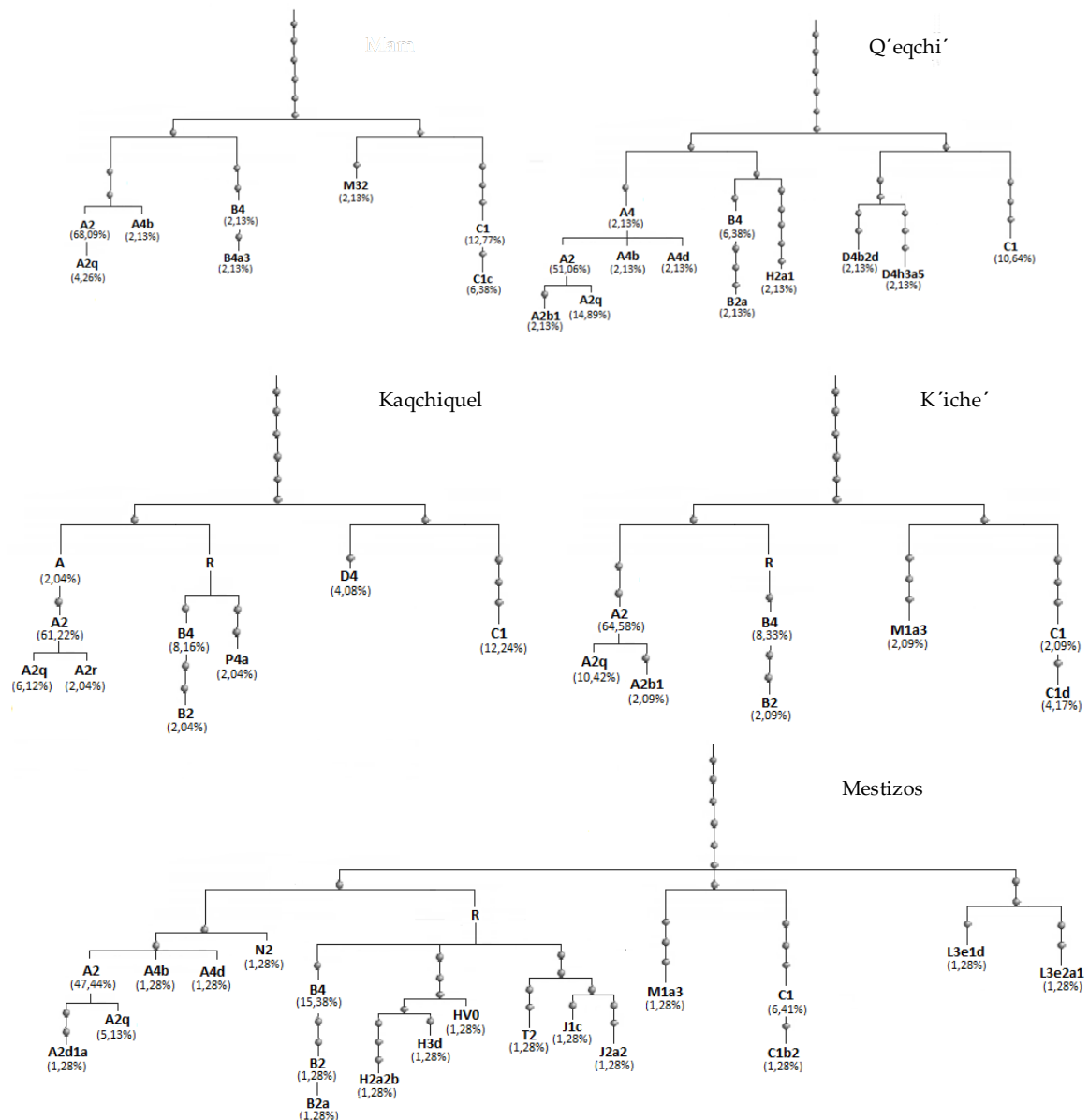


Figura VI.12: Árboles filogenéticos y proporción de los haplogrupos encontrados en las poblaciones mayas y en la población mestiza de Guatemala.

3.2.1 Estudio de Redes en la Población de Guatemala

Mediante un estudio Networks se caracterizaron a los distintos haplotipos encontrados. Este programa, selecciona cuales son los haplogrupos de origen de la red y va uniendo el resto de los haplotipos por un cúmulo de variaciones con respecto al haplotipo predecesor, primero, los más similares y así sucesivamente. En primer lugar, se verán los estudios de cada uno de los pueblos mayas (fig. IV.13 a 16), así como de la población mestiza (fig. IV.17). Por último, se estudiará cómo se distribuyen los individuos de toda Guatemala que pertenecen al haplogrupo A (fig. IV.18), haplogrupo que va a dar mayor información sobre la historia y evolución del pueblo de Guatemala, por ser el mayoritario y el ancestro más antiguo. A partir de éste, se podrá calcular el ancestro común de los actuales habitantes de Guatemala. Para las representaciones en imágenes (fig. 13 a 18) se eliminan las redes, tomando siempre el árbol más probable, para facilitar su visualización.

- Estudio de redes en la población Kaqchiquel

La población Kaqchiquel presenta un estudio de redes que diferencia los 4 grupos en cuatro ramas. El más alejado del resto es el B, con un mayor cúmulo de variaciones, con respecto al cálculo del nodo central que se situaría en el haplogrupo A. Desde la rama que une el haplogrupo A con el B, parte una rama, desde la que se formarán los haplogrupos C y D. Esta disposición es la esperada si se observa la filogenia de los haplogrupos mitocondriales (fig. IV.9).

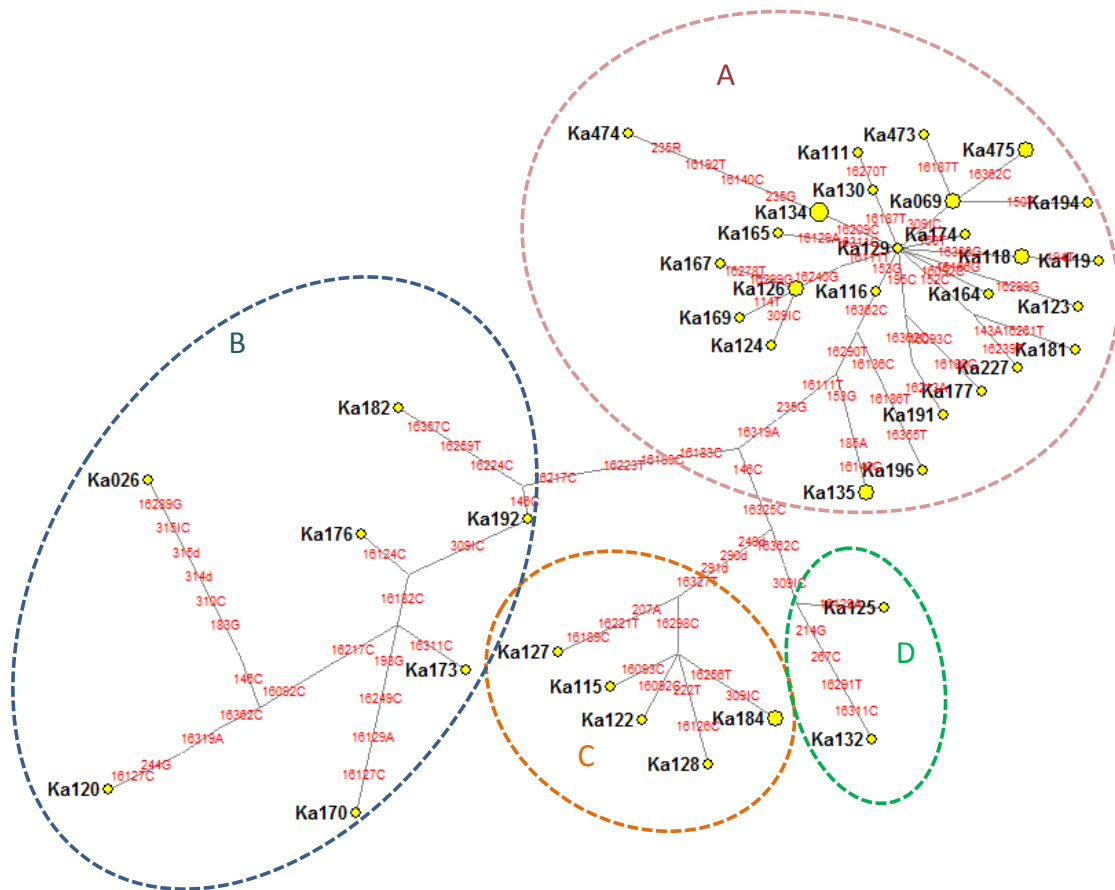


Figura IV.13: Gráfico de redes calculado mediante Network tomando como base los individuos de la población Kaqchiquel.

El haplogrupo A tiene un haplotipo central (a la derecha de la figura IV.13) corresponde al individuo Ka-129 (A2, 16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C). A partir de éste se ve como aparecen los nodos desde los que partirán distintas ramas y diez individuos, seis de ellos a partir de una única mutación. Los individuos del haplogrupo A más alejados de esta zona central son el 135, 196 y 474 (este último, A2q). Se puede ver que han acumulado variaciones, alejándose de los que el programa considera como haplotipos originarios.

- Estudio de redes en la población K'iche'

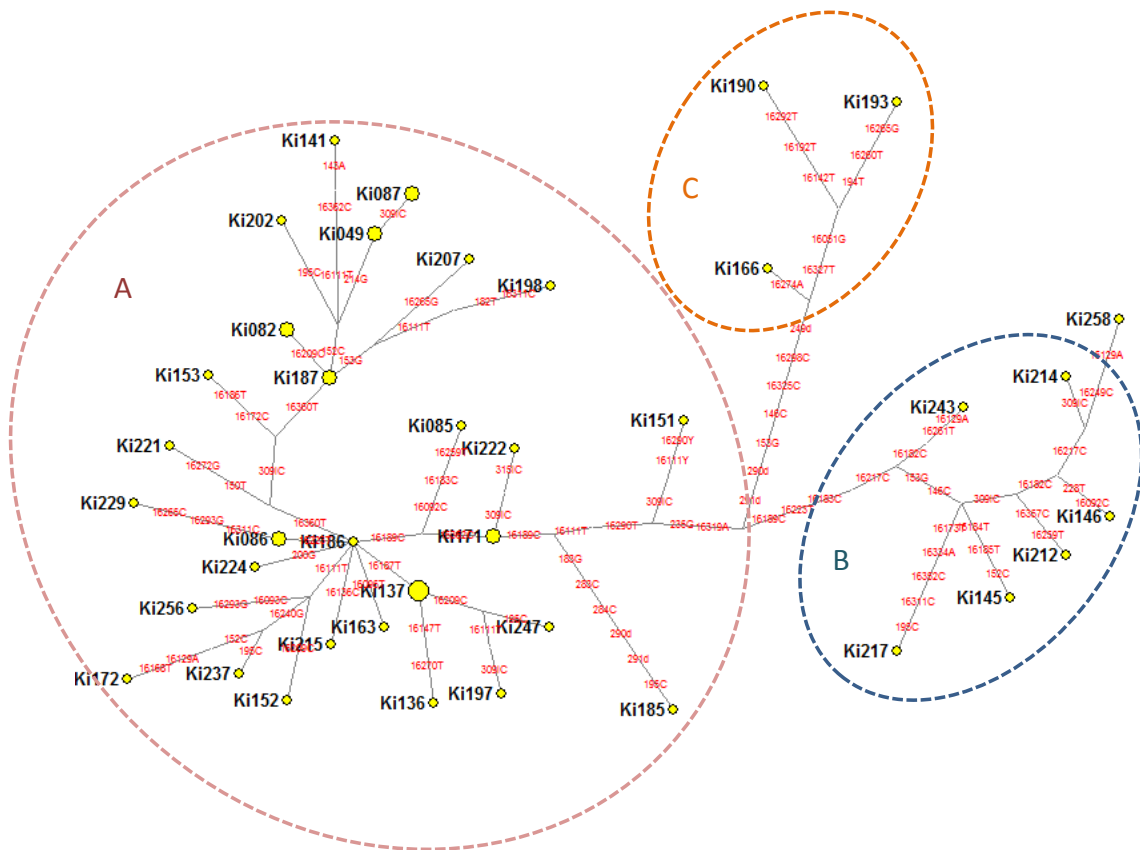


Figura IV.14: Gráfico de redes calculado mediante Network tomando como base los individuos de la población K'iche'.

En el grupo K'iche' se distinguen también en tres zonas distintas de la red, los haplogrupos A, B y C. El mayoritario es el A donde nace el árbol; a partir de éste se forma una red central, formada por once individuos, donde un haplotipo está formado por cuatro individuos, Ki137, y tres haplotipos están formados por dos individuos. Cuando se observa el haplogrupo A, se ve que el haplotipo central, del que se originan más individuos, es el Ki186. A partir de éste y con una sola mutación, se encuentran seis haplotipos y dos nodos. Éste haplotipo central además coincide con el encontrado en la población Kaqchiquel (Ka129). El haplotipo mayoritario, Ki137, se origina a partir del Ki186, acumulando una mutación, la 16187T. El haplotipo más alejado de la red central es el Ki185, donde aparecen las mutaciones 283C, 284C y la delección en 290-291, también desaparece el SNPs 16362C, que sí aparece en Ki186.

- Estudio de redes en la población Mam

Cuando se realiza el estudio de redes de la población Mam, se observa como desde el haplogrupo A se van alejando haplotipos, los más alejados, los que forman parte del haplogrupo C y B (éste último formado por un único haplotipo compartido por dos individuos).

En el inicio de la red del haplogrupo A el haplotipo Mam236, compartido por tres individuos, Mam-236, 360, 361; es el que se encuentra en la posición más central de la red. De este haplotipo, 16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C, aparecen cinco distintos haplotipos acumulando una sola mutación, y dos nodos. De los cinco haplotipos, dos de ellos, Mam362 y Mam240; poseen dos individuos cada uno.

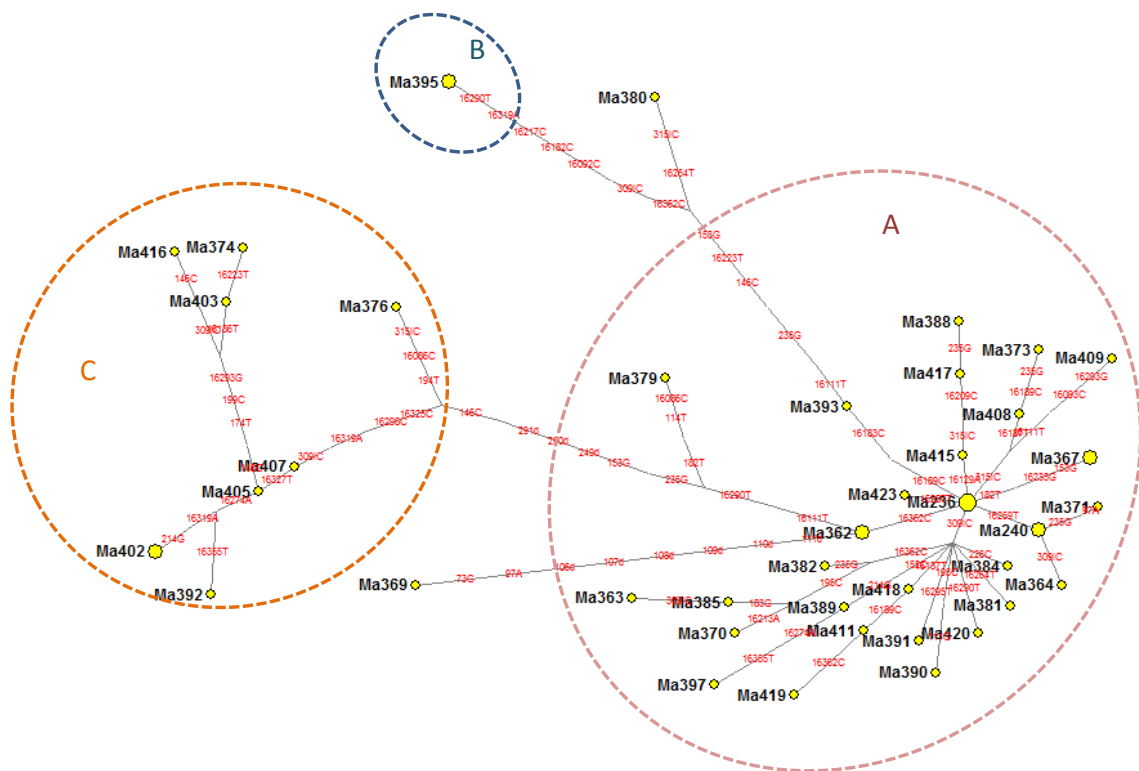


Figura IV.15: Gráfico de redes calculado mediante Network tomando como base los individuos de la población Mam.

Al igual que los haplotipos que forman parte de los haplogrupos C y B, alejados del inicio de la formación de la red; se encuentran los individuos 369 y 379 (ambos A2) y el 380 que corresponde al subgrupo A4b.

- Estudio de redes en la población Q'eqchi'

Como en los estudios anteriores, el árbol formado a partir del estudio de redes parte desde los haplotipos con haplogrupo A, a partir de un nodo del que no hay representante en la población, el nodo mv1, 16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C. Éste corresponde con los haplotipos centrales de las poblaciones Kaqchiquel y K'iche'.

Desde esta posición central, aparecen tres haplotipos A, con la diferencia en la posición 16209 C, el haplotipo Qe266, 4 individuos, y el Qe286, dos, y Qe266 con tres individuos cada uno. Del haplogrupo A, los haplotipos más alejados son los Qe274 y Qe260, con más de 8 sustituciones con respecto al nodo central.

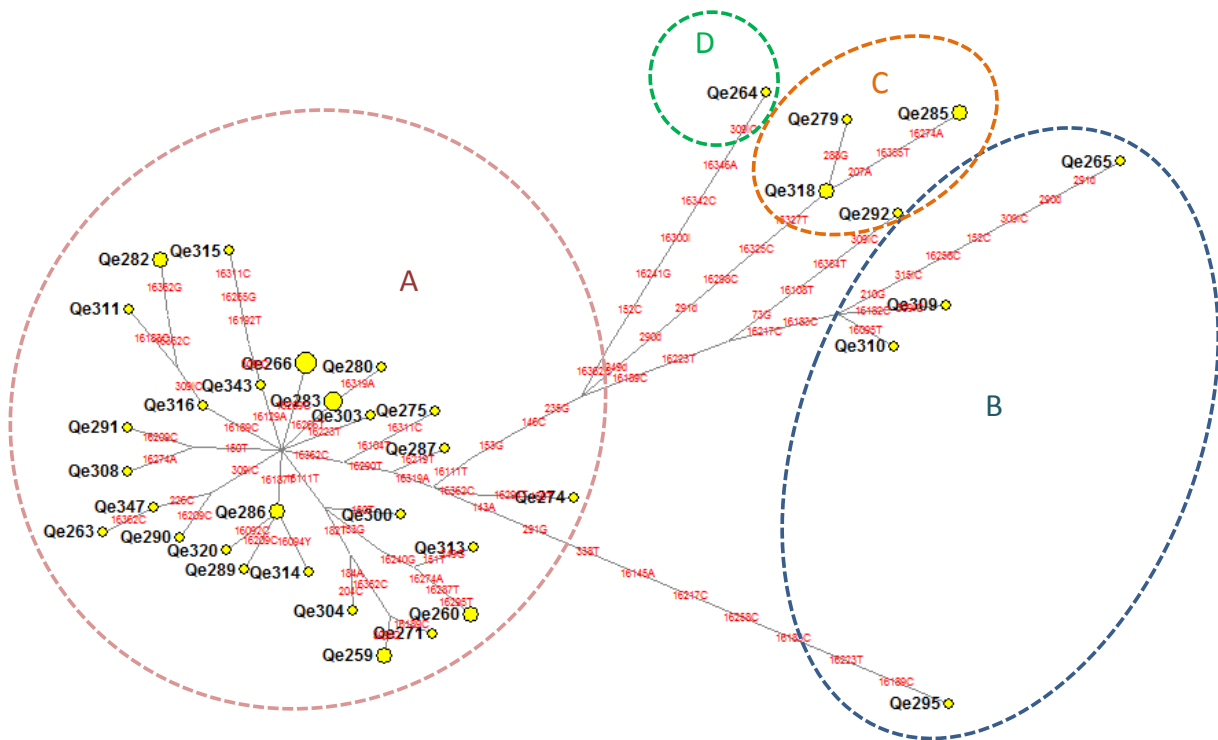


Figura IV.16: Gráfico de redes calculado mediante Network tomando como base los individuos de la población Q'eqchi'.

- Haplogrupo A en la población Mestiza

La red del grupo mestizo es la más compleja, no sólo por tener un mayor número de individuos, sino por la disposición de éstos.

En la parte superior de la fig. IV.17, se encuentran los haplotipos que forman parte del haplogrupo A, el más numeroso. En la parte inferior, se ven los individuos que pertenecen al haplogrupo B y C. Entre estos últimos y los haplotipos A se encuentran diez haplotipos, alejados unos de otros, que son los asignados como haplogrupos B, C, H3 o L3.

En la zona central del haplogrupo A, se ve el haplotipo Mest249, *16111T 16223T 16290T 16319A 73G 146C 153G 235G 263G 309.1C 315.1C*. De este individuo parten tres redes que lo conectan con un haplotipo representado por cuatro individuos, Mest006, y otras dos que lo conectan con otras redes. Otra de las posiciones centrales más interesantes es un nodo que carece de haplotipo, desde donde parten el mayor número de redes dentro del haplogrupo A, se une a 10 haplotipos y un nodo. De estos 10 haplotipos, dos de ellos son de los más numerosos, el Mest088 con seis individuos y Mest270 con cuatro individuos.

Los individuos más alejados de la zona central dentro del haplogrupo A son los individuos Mest471 (A2) y Mest278 (A4d).

- Haplogrupo A en la población de Guatemala

Finalmente, después de estudiar cada una de las poblaciones, se representaron los individuos con haplogrupo A de la población de Guatemala. Para esto se tomaron todos los componentes con haplogrupo A y se hizo un nuevo análisis. Para poder diferenciarlos, tras realizar el análisis cada una de las poblaciones estudiadas se representaron en un color, como se puede ver en la figura IV.18.

Una vez realizado el análisis, se ve que la red central está formada por dos haplotipos de los que parten la gran mayoría de los haplotipos encontrados, Ka129 y Ka069. Desde el que parte mayor número de redes es desde el Ka129, una de ellas hasta el Ka069, del que sólo se diferencia en la inserción 309.1C. En total, surgen 11 haplotipos con una única diferencia y un nodo (39 individuos con una única sustitución). Destacar entre estos haplotipos el Ka130, que se diferencia con el anterior en la posición 16187 y está formado por doce individuos; y el Mest378 que lo comparten siete individuos, con la sustitución en 16111. Como ya se ha mencionado, desde el Ka129 surge también el haplotipo Ka069. Este haplotipo lo comparten trece individuos, y de él parten 9 haplotipos (25 individuos conectados directamente). De estos nueve haplotipos, el Ka475 es el más numeroso, lo comparten siete individuos con una única variación en la posición 16362.

Los haplogrupos más alejados son los que pertenecen a haplogrupos A4b.

Como se puede ver en el resultado, no se encuentra una distribución en ramas de cada una de las poblaciones, por lo que todos los haplotipos de las distintas poblaciones, incluida la mestiza se encuentran entremezclados. Tampoco se ve una población predominante en la zona central. Lo que indica que entre las poblaciones no hay signos de que alguna de ellas sea más antigua o que de origen al resto, sino más bien que poseen un origen común y que la cultura y lengua común hacen que los distintos pueblos mayas posean un flujo que haga que sean muy similares.

3.3 DISTANCIAS GENÉTICAS

Mediante el programa Arlequin v3.5, se realizó el análisis de las distancias entre las poblaciones. Para comprobar la distancia genética entre las poblaciones se realizó un cálculo F_{st} [82] y de las distancias de Reynolds [82] y Slatkin [33]. Se tomaron todas las secuencias obtenidas y se calculó la distancia entre estas.

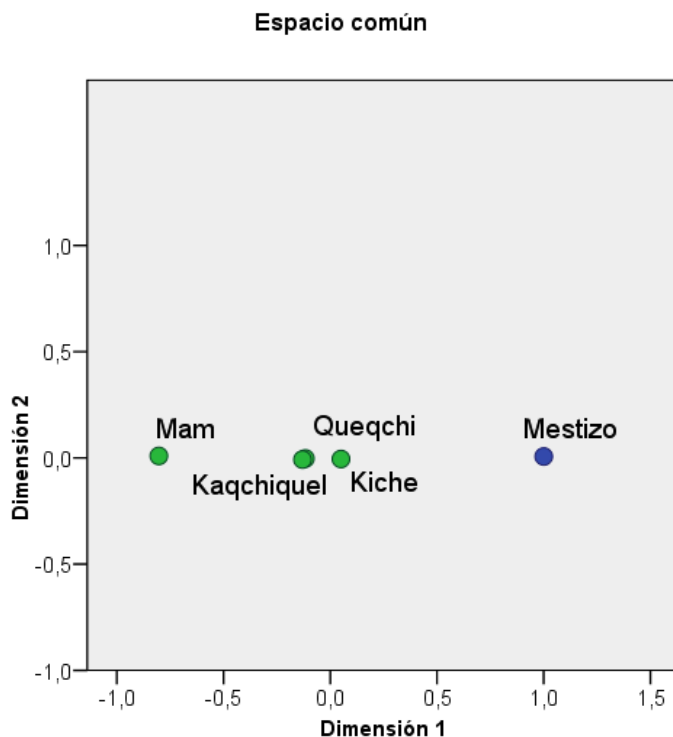


Figura IV.19: Gráfico de distancias genéticas de Reynolds, obtenidas a partir de las cinco poblaciones de Guatemala. Para su representación se realizó un cálculo MDS bidimensional. En color verde las poblaciones mayas, en azul la mestiza

Las distancias obtenidas a partir de las secuencias destacan la gran similitud entre las poblaciones mayas, algunas de ellas se superponen en la representación, ver figura IV.19. Cuando se representan todas las poblaciones de Guatemala se ve que la distancia con la población mestiza es algo mayor que las distancias entre las poblaciones mayas.

Los resultados corroboran los obtenidos mediante el estudio de las distancias F_{st} y las distancias de Reynolds y Slatkin. Queda reflejado el pequeño porcentaje de variación cuando se agrupan los mayas, un 0,08%, y como aumenta levemente al incluir las mestizas, un 0,92%, y haciéndose siempre mucho mayor cuando la comparación incluye a la población española (entre el 13,37 y 29,30%).

Otro de los análisis realizados para comprobar las distancias entre las poblaciones estudiadas, fue un análisis de correspondencias realizado mediante las frecuencias de los haplogrupos encontrados. Posteriormente, se representó en un gráfico tridimensional, figura IV.20. En la primera dimensión se observa como la población española se aleja y se puede ver como la población mestiza se acerca levemente a ésta. La disposición de la población española se debe a la frecuencia del haplogrupo H. En la segunda dimensión, se puede ver como la población Kaqchiquel se aleja levemente del resto de las mayas, debido a la frecuencia de los haplotipos D1.

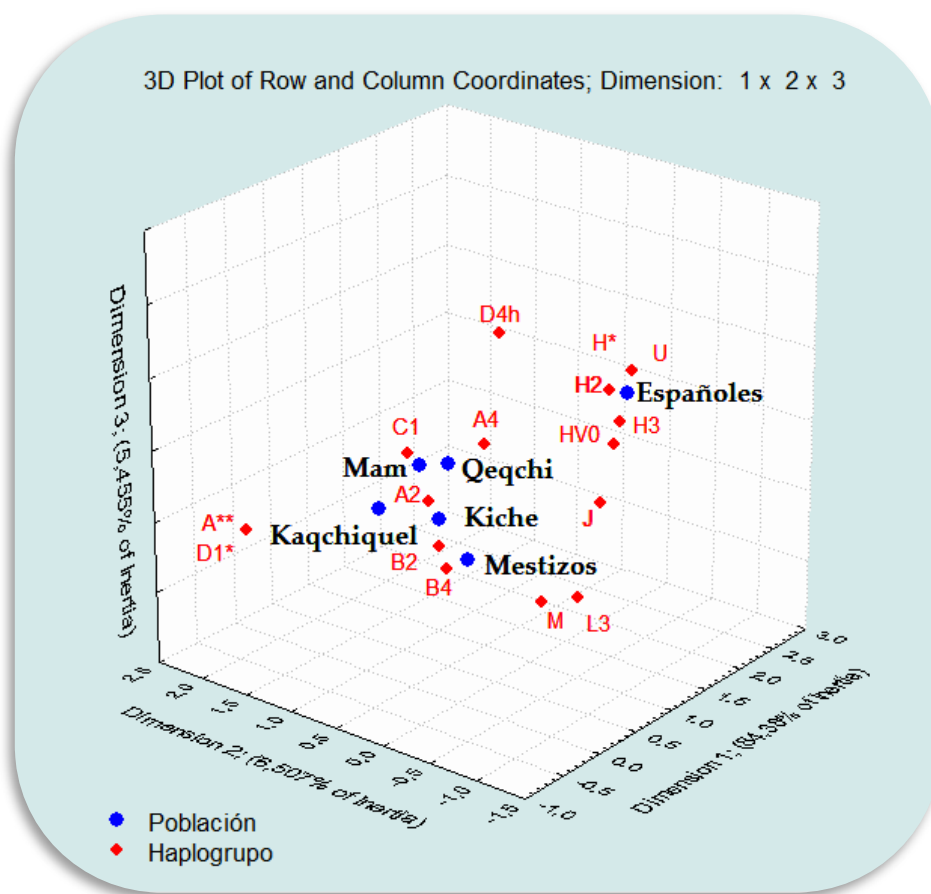


Figura IV.20: Gráfico de distancias genéticas obtenidas a partir de las frecuencias de 129 posiciones del ADNmt, obtenidas a partir de las poblaciones española y mayas y mestizos de Guatemala. Para su representación se realizó un cálculo de análisis de correspondencias y una representación tridimensional.

3.4 DISTANCIAS CON OTRAS POBLACIONES

Una vez calculados los haplotipos de estas poblaciones, al igual que en capítulos anteriores, se realizaron comparaciones con otras poblaciones. Este estudio se realizó de nuevo con el software Arlequin v3.5 y los resultados de las

matrices de distancias de Reynolds [82] se representaron mediante NM-MDS realizados con SPSS v15.

En primer lugar se realizaron comparaciones con poblaciones africanas, europeas y asiáticas: africanas del norte (Argelia [83], Libia [84] y Mauritania [85]), y centro (Angola [86], Etiopía [87]), europeas del norte (Finlandia [88], Islandia [89], Suecia [90]), del centro (Francia [91], Inglaterra [92] y Alemania [93]) del sur (Portugal [94], Italia [49,95] y España [49]) y Rusia; y asiáticas del norte (Mongolia [96] y Japón [97]) y del centro (China [98], Pakistán [99] y Vietnam [100]).

Al realizar la comparación con estos grupos poblacionales, figura IV.21, se puede ver como las poblaciones más cercanas a las poblaciones de este estudio son las asiáticas, en primer lugar las asiáticas del norte y después las del sur. De las poblaciones europeas, las más cercanas a la población española son las europeas del sur y después de éstas las de centro Europa. Muy alejadas a las poblaciones de Guatemala, así como a las asiáticas y europeas se encuentran las poblaciones Africanas.

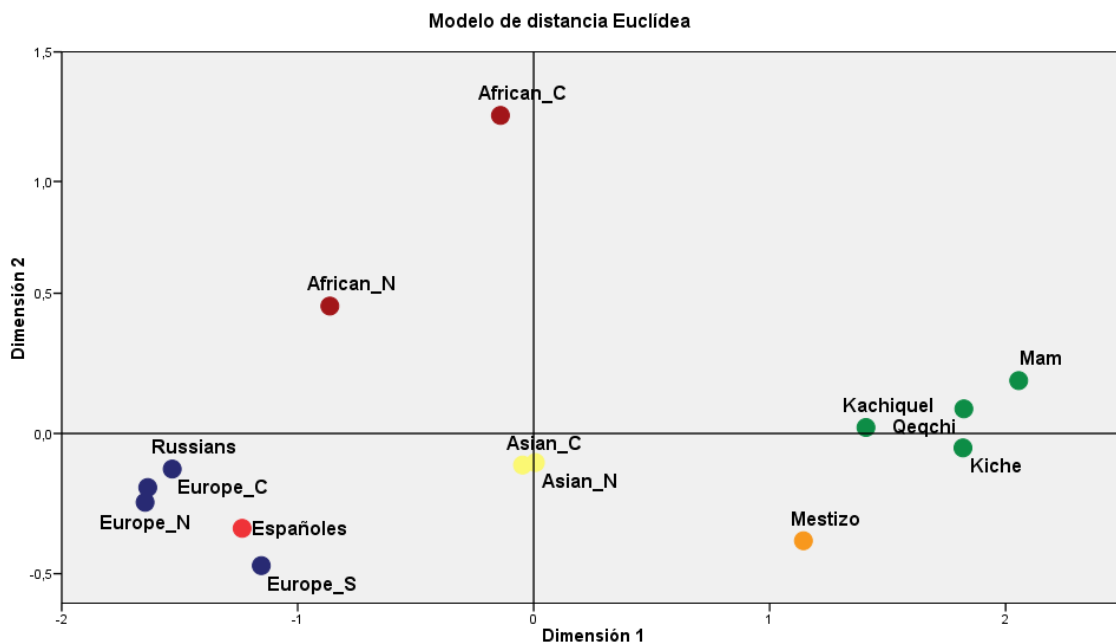


Figura IV.21: NM-MDS de las distancias de Reynolds obtenidas al comparar las poblaciones maya de Guatemala (en color verde), mestiza de Guatemala (naranja), española (rojo), europeas (azul), africanas (marrón) y asiáticas (amarillo).

Las distancias obtenidas en este análisis ponen de manifiesto que los Mestizos se encuentran muy próximos a los mayas, y por tanto también a las poblaciones asiáticas. La población mestiza muestra una mayor proximidad a la población española y sur europea, en comparación con las poblaciones mayas. Las poblaciones africanas se encuentran muy alejadas a las poblaciones estudiadas. La norteafricana se acerca a las poblaciones europeas, pero las poblaciones centro-africanas se alejan del resto, teniendo las mayores diferencias con todas las poblaciones excepto con la norte africana. Este hecho refleja la poca interacción que han tenido los pueblos africanos en la población mestiza de Guatemala, así como en el resto de las poblaciones.

Después de este primer estudio, donde se obtuvieron los resultados esperados, se realizó un segundo análisis, para poder conocer mejor los fenómenos evolutivos en América. Para este estudio se tomaron secuencias de otro análisis en poblaciones nativas americanas, tabla VI.1. Con este trabajo se puede

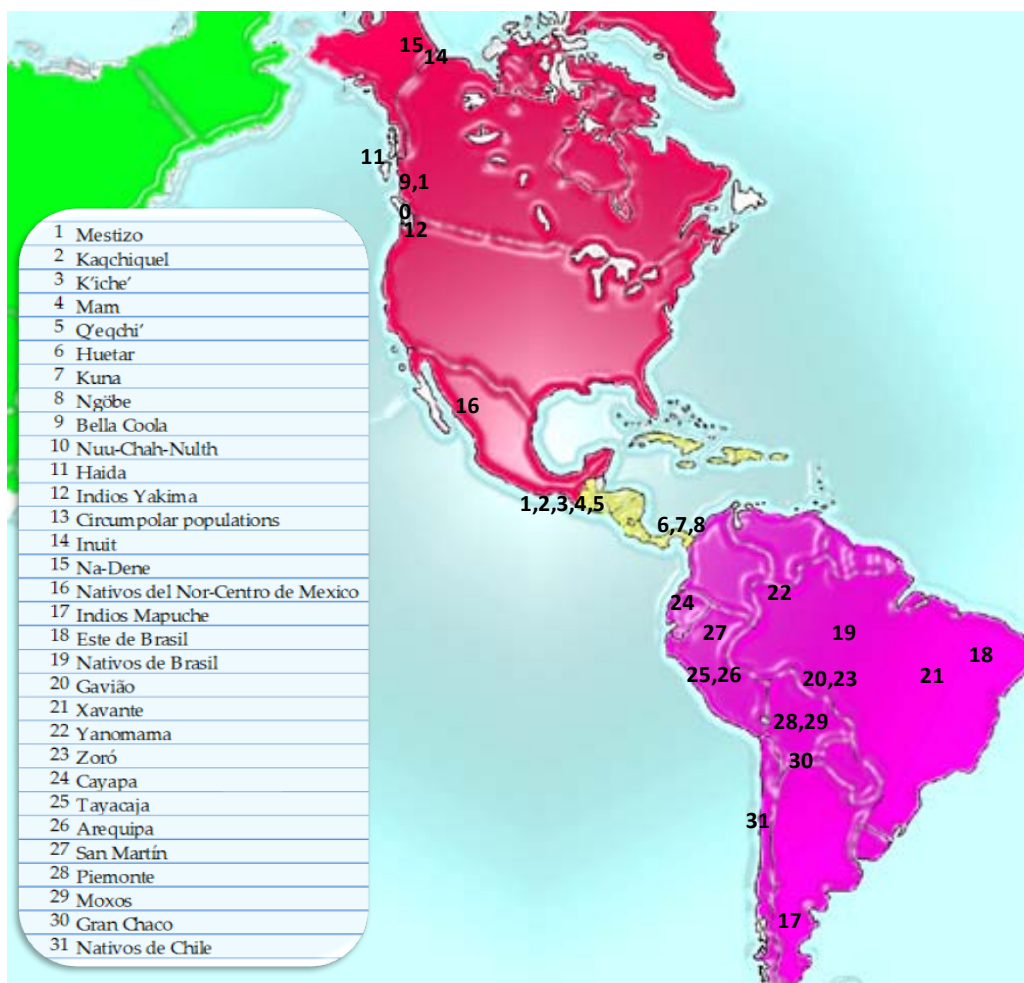


Figura IV.22: Situación de las poblaciones nativas americanas estudiadas para los análisis de ADNmt. Ocho en Norteamérica, ocho en Centroamérica y quince de Sudamérica.

explicar la distancia genética entre poblaciones nativas que se encuentran aisladas lingüística, cultural y geográficamente. La disposición geográfica de estas poblaciones se muestra en la figura VI.22.

Los resultados del análisis de distancias se reflejan en la figura IV.23. En esta figura, se puede ver como las poblaciones más alejadas, son las de Norte América y Groenlandia, Inuit y circumpolares.

Las poblaciones de Guatemala se encuentran cercanas a otras poblaciones nativas centro-americanas, como son las Chibchan (Ngòbe y Huetar), pero también a poblaciones que habitan en Canadá, como lo Bella Coola. La población Kuna, también centroamericana, se encuentra más alejada del resto de las poblaciones, situándose en la representación (fig. IV.23), en la zona central más baja.

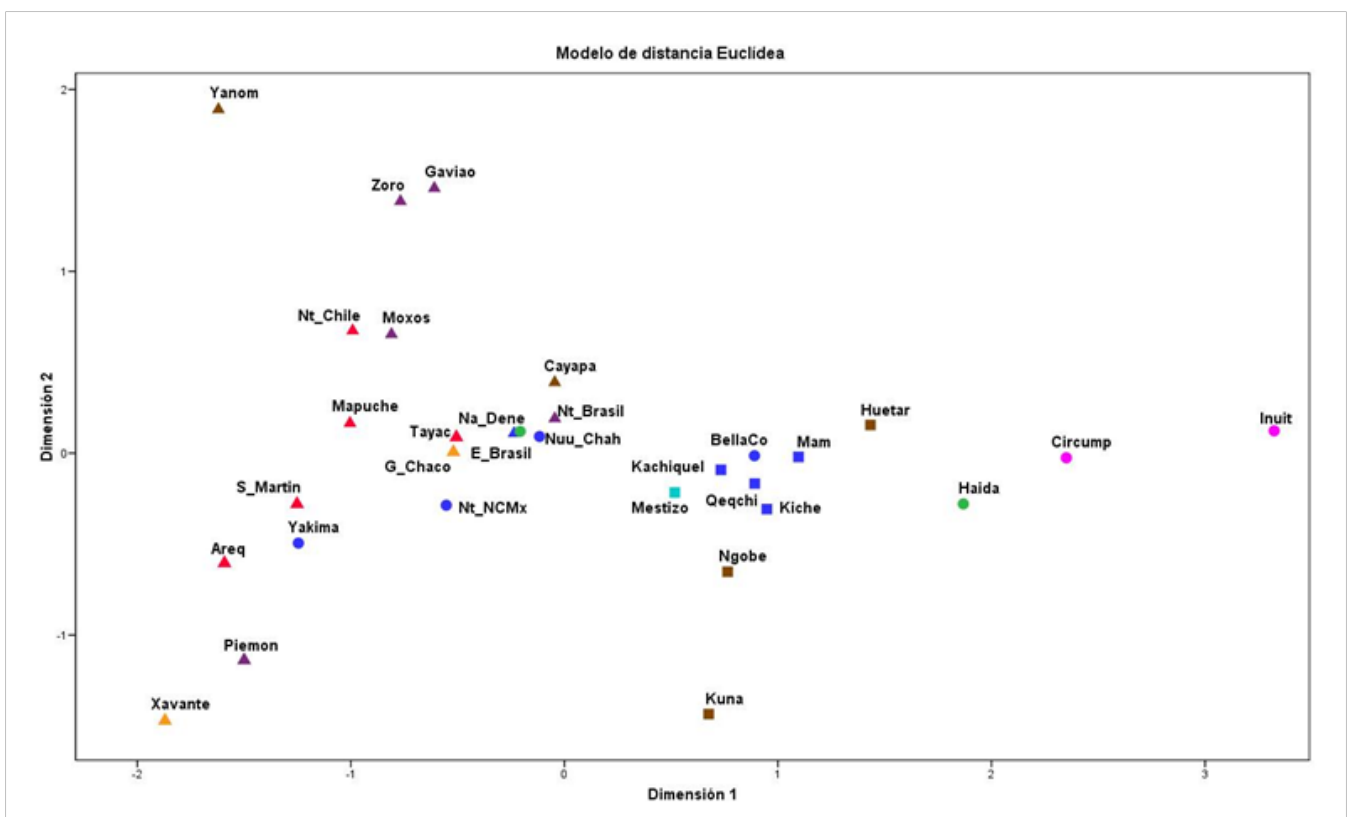


Figura IV.23: Plot de las distancias euclídeas calculado a partir de las distancias de Reynolds entre las poblaciones nativas americanas y relacionadas con éstas. En el análisis se comparan: poblaciones americanas del Norte (contorno circular; incluyendo Alaska, Canadá, EEUU y Méjico), Centro americanas (contorno cuadrado) y Sudamericanas (contorno circular). El color de cada uno de los puntos se debe a su familia lingüística. Stress=0,2374

En este mismo análisis se puede ver como muchas de las poblaciones nativas de Sudamérica: Mapuches, Tayacajas, Cayapa, Gran Chaco, etc. se encuentran muy cercanas a los grupos poblacionales norteamericanos. Tres son los grupos norteamericanos que se entremezclan con los sudamericanos, los Na-Dene, los Nativos de la zona Norte de Méjico y Nuu-Chah-Nulth, mientras que otros grupos como los Haida (también con lengua Na-Dene), se acercan a los grupos de Alaska.

La disposición que se encuentra al estudiar las poblaciones refleja como las distancias geográficas (se puede observar sobre todo en las norteamericanas) o la lengua (muy destacado en el Chibchan-Paezan o en los Na-Dene) no influyen tanto, en la medida que lo hacen los acontecimientos microevolutivos que ocurren en cada población a la hora de formar su pool genético.

Tras este análisis de distancias, se comprobó si la fuente de variación en las distancias genéticas obtenidas se debía a su disposición geográfica, o a la historia y cultura de las mismas. Para esto se diseñó un análisis de la varianza molecular (AMOVA), teniendo en cuenta por un lado, el origen de su lengua y en un segundo estudio, su disposición geográfica.

Para este estudio, AMOVA, se dividieron las poblaciones en siete grupos lingüísticos: Amerindio, Chibchan-Paezan, Na_Dene, Andino, Ecuatorial-Tucano y Ge-Pano-Caribeño. Y tras esto, en seis grupos geográficos: Alaska y Canadá (grupo 1); EEUU y Méjico (grupo 2); Centroamérica (grupo 3); Norte de Sudamérica (grupo 4); Brasil (grupo 5) y Sur de Sudamérica (grupo 6).

Los resultados de este estudio AMOVA se muestran en la tabla IV.10. En éstos se observa que, al hacer el estudio agrupando por el origen de su lengua, se obtiene una mayor diferencia dentro de cada grupo (intragrupal) que la que existe entre los grupos (intergrupal); con valores de $14,19 > 6,73\%$ respectivamente. Cuando los grupos se realizan según su origen geográfico, las distancias intergrupales e intragrupalas son similares (10,55% y 10,90%).

Tabla IV.10: Estudio AMOVA de las poblaciones nativas americanas, teniendo en cuenta su lengua y su situación geográfica.

Fuente de variación	d.f.	s.Cuadrados	Varianza entre componentes	Porcentaje de variación	Índices de fijación
Grupos con mismo origen lingüísticos					
1. Amerindio: Kaqchiquel, Mam, K'iche', Q'eqchi', Nuu-Chah-Nulth, Bella Coola, Yakima, Nativos del norte, centro de Méjico y Nativos del este de Brasil.					
2. Chibchan-Paezan: Ngobè, Kuna, Huetar, Yanomami y Cayapa.					
3. Na-Dene: Na_Dene y Haida					
4. Eskimo: Inuit y Circumpolares					
5. Andino: Mapuche, Tayacaja, Arequipa, SanMartin y Nativos de Chile.					
6. Ecuatorial-Tucano: Nativos de brasil, Gavia, Zoro, Piedmonte y Moxos.					
7. Ge-Pano-Caribeño: Xavante y GranChaco.					
Entre los grupos	6	565.364	0,22053 Va	6,73%	FSC:0.15211
Dentro de cada grupo	23	587.674	0,46465 Vb	14,19%	FST:0.20920 FCT:0.06733
Grupos con situación geográfica próxima					
1. Alaska y Canadá: Inuit, Circumpolares, Nuu-Chah-Nulth, Bella Coola y Haida.					
2. EEUU y Mejico: Na-Dene, Yakima y Nativos del norte, centro de Méjico.					
3. CentroAmerica: Kaqchiquel, Mam, K'iche', Q'eqchi', Ngobè, Kuna y Huetar.					
4. Norte de Sudamérica: Cayapa, SanMartin, Tayacaja y Arequipa					
5. Brasil: Yanomami, Nativos del este de Brasil, Nativos de Brasil, Gavia, Zoro y Xavante.					
6. Sur de Sudamérica: Mapuche, Nativos de Brasil, Piedmonte, Moxos y GranChaco.					
Entre los grupos	5	630.661	0,34796 Va	10,55%	FSC:0.12182
Dentro de cada grupo	24	522.377	0,35930 Vb	10,90%	FST:0.21450 FCT:0.10553

Aunque las diferencias encontradas en este estudio no son muy grandes, los resultados, en los análisis AMOVA, muestran que, aún teniendo un mismo origen lingüístico, las variaciones que se han generado tras la división de estos grupos, es mayor que las que se conservan de su origen común. Los resultados casi idénticos, cuando se agrupan las poblaciones según su disposición geográfica, confirman que estas poblaciones se encuentran aisladas y no existe un gran flujo genético entre las poblaciones estudiadas y sus pueblos vecinos. Este estudio, por tanto, confirma que la fuente de variación, se encuentra en los cambios evolutivos intrapoblacionales, es decir, los cambios que se producen en el seno de cada población.

Como se detalla en el primer capítulo, la aparición de los pueblos nativos americanos, a partir de poblaciones asiáticas, se puede demostrar fácilmente teniendo en cuenta los datos genéticos. Pero así como este efecto fundador es

claro, es muy difícil poder demostrar si este proceso se originó en una sola etapa o en varias. Los resultados, podrían discutirse tanto con las teorías que apoyan la aparición de los pueblos americanos con un único movimiento migratorio y un origen asiático común [101], como por varios orígenes, porque las distancias genéticas no se corresponden con la formación de cada una de las poblaciones. Son mucho mayores las distancias, fruto de la evolución de cada uno de los pueblos, por tanto es difícil describir el paso del reloj evolutivo en las poblaciones nativas americanas.

3.5 ESTUDIO DEL MESTIZAJE

Como en capítulos anteriores, se calculó la aportación de las poblaciones parentales mayas y europeas a los mestizos de Guatemala. En el caso del ADNmt, se puede diferenciar como ha sido la aportación por vía materna de éstas.

Para realizar una aproximación, se realizó un análisis de distancias genéticas con otras poblaciones nativas cercanas (Chibchan) [63,64], una población mestiza del Salvador [102] y poblaciones del centro y sur de Europa.

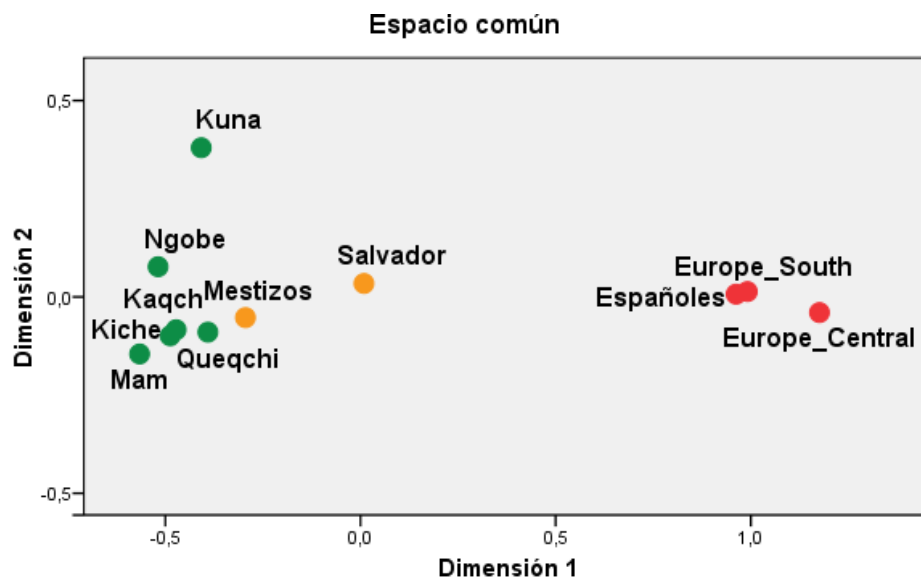


Figura IV.24: Representación ND-MDS del cálculo a partir de las distancias de Reynolds entre las poblaciones Europeas y Americanas, nativas y mestizas. Stress=0.0087

Si se observa la figura IV.24, se ve como la población mestiza se sitúa mucho más cercana a las poblaciones nativas centro americanas que a las poblaciones europeas. Por tanto, la contribución europea, en el caso del estudio del ADNmt, es mucho menor que la maya. Si se comparan los resultados con los obtenidos en capítulos anteriores, se ve como en el caso de la línea materna, la población mestiza posee un aporte mucho más marcado de origen maya.

Si se estudian las frecuencias de las posiciones polimórficas encontradas en el ADNmt (sin tener en cuenta indels) de la población mestiza usando como poblaciones parentales las frecuencias obtenidas en las poblaciones maya y española y mediante los programas Admix Pas y Admix 95, se obtienen los resultados que se muestran en la tabla IV.11.

Tabla IV.11: Contribución maya y española a la población mestiza de Guatemala.

	Admix_Long (51 SNPs)		Admix2_Chakraborty (47 SNPs)	
	m	s.e.	m	s.e.
Maya	0.8022	0.0050	0.8093	0.0890
Española	0.1978	0.0050	0.1907	0.0890

Los resultados en ambos casos (el método de identidad genética [103] y el de los mínimos cuadrados [104]) son similares. El aporte de la población maya es superior al 80% en ambos casos y la contribución española es siempre menor al 20%.

Estos resultados ponen de manifiesto el gran aporte de la población maya en la estructura genética de la población mestiza de Guatemala a nivel de su ADNmt. Como era de esperar tras ver los haplotipos y haplogrupos encontrados en la población mestiza, es mucho más similar a la población maya que a la española.

3.6 ESTRUCTURA DE LA POBLACIÓN

Mediante el programa STRUCTURE 2.3.1 [59,60], se intentaron buscar componentes genéticas en las poblaciones mayas, mestizas y españolas; a partir de 173 polimorfismos encontrados en HV1 y HV2.

Para hacer este estudio se transforma la secuencia de ADN por un código numérico y se eliminan todas las posiciones que no son informativas. Los polimorfismos en la secuencia del ADNmt pondrán de manifiesto la probabilidad de encontrar distintos componentes en cada una de las poblaciones, así como la distribución de los componentes. Se asumieron hasta 25 componentes distintos, a la hora de estudiar conjuntos de poblaciones y hasta 15 al estudiar grupos poblacionales de forma individual.

Como parámetros para el uso del programa, se definió un número de 50.000 interacciones y un periodo "*burn-in*" de 10.000 interacciones adicionales antes de cada uno de los análisis. Todos los análisis se realizaron por triplicado, para comprobar su reproducibilidad.

Si se estudia la población de Guatemala, incluyendo tanto mayas como mestizos, se ve que la probabilidad va aumentando según se aumenta el número de poblaciones asumidas (K) hasta llegar a 6 (asumiendo un modelo "sin mezcla") y 7 (si se presupone que el modelo, es una población producto de mezclas). Si se observan los gráficos de barras, como se puede ver en la figura IV.25, se comprueba que, al distribuirse los componentes a lo largo del grupo muestral Guatemala (4 mayas y mestizos), los componentes se distribuyen de una forma homóloga en cada uno de los grupos poblacionales.

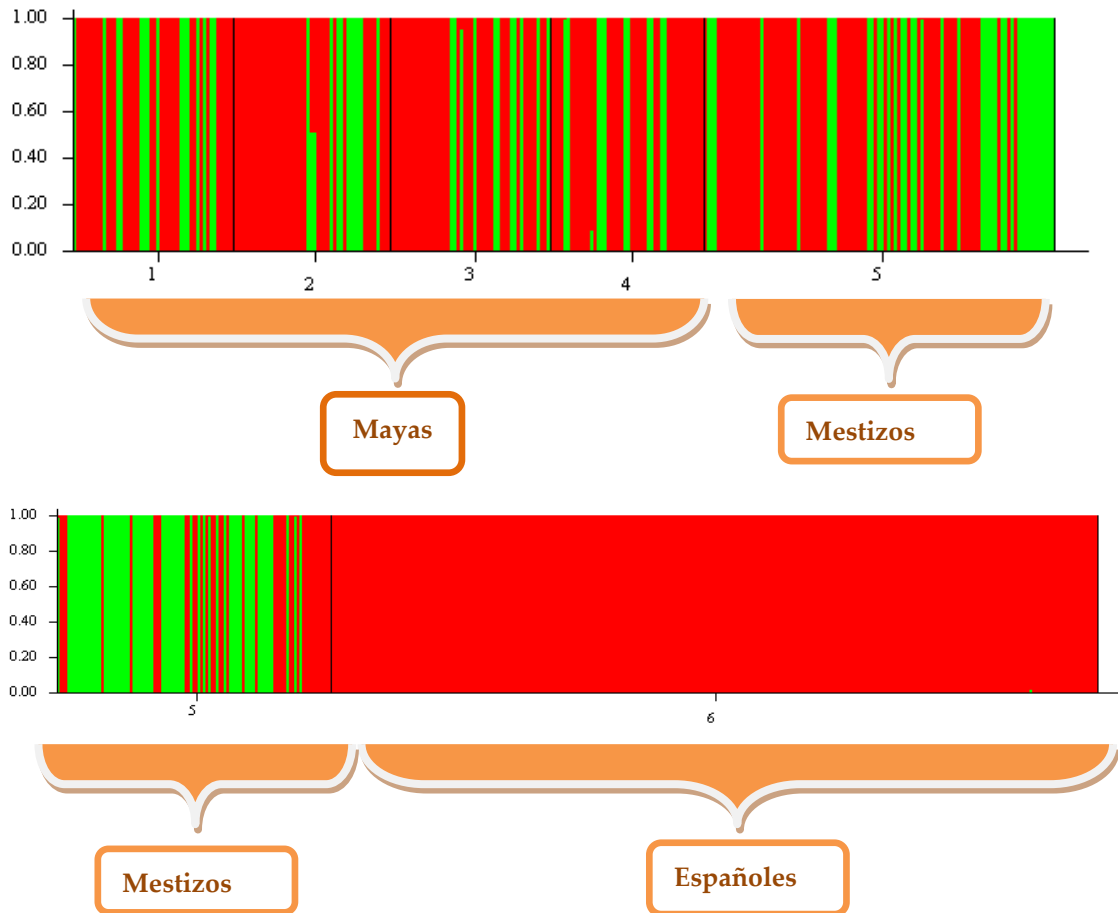


Figura IV.25 y 26: Imágenes obtenidas en Structure. Tomando los cinco grupos poblacionales guatemaltecos, en la figura IV.25 $K=2$. En la figura IV.26 Tomando el grupo poblacional mestizo y español, $K=2$.

Si se observa la figura IV.26, representación del estudio conjunto de la población mestiza de Guatemala y la española, se ve un comportamiento diferente al anterior. En el análisis con dos componente, $K=2$, el programa busca dos componentes en los individuos. Se puede ver como todos los individuos españoles aparecen con la componente representada por el color rojo, mientras que en la población mestiza de Guatemala, la gran mayoría muestra la componente representada en verde y algunos la componente roja.

Si se estudia al mismo tiempo las 6 poblaciones ($n=591$) la probabilidad va aumentando según se va asumiendo mayor número de poblaciones. Pero al observar cómo se distribuyen éstas en los individuos de cada una de las poblaciones se ve perfectamente cómo se pueden diferenciar dos grupos. Desde

$K=2$ en adelante, se puede diferenciar siempre la población española de la guatemalteca. Como se puede ver en la figura IV.27 ($K=3$), los tres grandes grupos, mayas, mestizos y españoles, se diferencian perfectamente; por un lado los dos grupos guatemaltecos (5 poblaciones) y por otro lado el grupo español. La componente representada en azul solamente se puede observar en la población maya mientras que la componente roja aparece mayoritariamente en la población española.

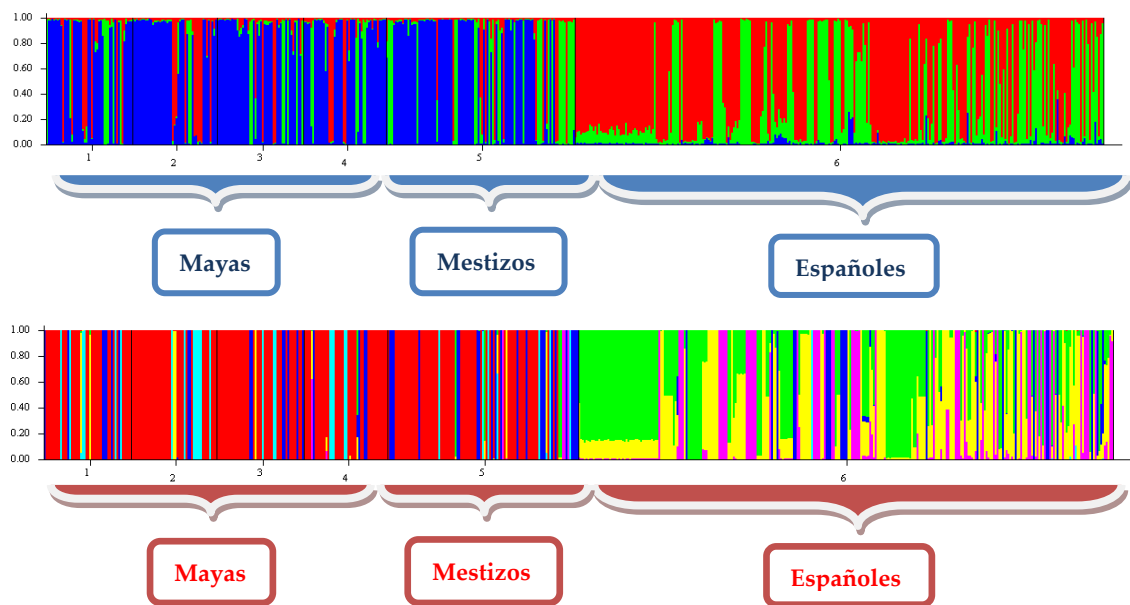


Figura IV.27 y 28: Imágenes obtenidas en Structure. Tomando los cinco grupos poblacionales guatemaltecos y el español. En la figura IV.27 $K=3$. En la figura IV.28 $K=6$

Cuando se aumenta el número de poblaciones asumidas, la diferencia entre guatemaltecos y españoles se hace más patente. Si se observa la figura IV.28, se ve cómo se diferencian los dos grupos, españoles y guatemaltecos. Con $K=6$ se distribuyen seis componentes, pero en vez de diferenciar las poblaciones mayas entre sí o las mayas con la mestiza de Guatemala, lo que siguen mostrando es la diferencia entre los individuos españoles y guatemaltecos.

CONCLUSIONS

Among the 296 individuals of Guatemala, there are 205 unique haplotypes. Among the 295 Spaniards there are 205 unique haplotypes as well. The discrimination power for identification is between a 71% in Mestizo and 85% in Mam. The discrimination power is much lower in mtDNA comparing to a-STRs and Y-STRs, but the features in ADNmt makes it quite interesting for forensic genetics and more for molecular anthropology.

The Guatemalan, Mestiza and Mayan populations, have a mismatch bimodal distribution. However, the Spanish population is unimodal.

Most of the haplogroups in Mayan and Mestizo populations of Guatemala are A and B haplogroups, while in Spanish population the most prevalent is H haplogroup.

Genetic distances among Mayan populations are very small. The distance between Mayan and Mestizo population is also small, but it's a bit bigger than distances among Mayan populations.

When comparing these populations with others Native American ones, a vast variability is found. This is the result of evolutionary processes that have occurred a big divergence among Native American populations.

The crossbreeding in the Mestizo population of Guatemala, when studying DNAm, is not highlighted. More than 80% of the contribution in the DNAm in Guatemalan Mestizo population is from Mayan parental. However, European populations are always far from Mestizo populations. When the structure of the population in Guatemala is studied, either differences between Maya and Mestizo populations nor possible different ancestors can be found. In DNAm, there is no significant differences between Mayan and Mestizo populations that allow us to distinguish them in different population groups.

4. BIBLIOGRAFÍA

- [1] M. Kayser, Uni-parental markers in human identity testing including forensic DNA analysis, *BioTechniques*. 43 (2007) S16-S21.
- [2] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum.Mutat.* 30 (2009) E386-394.
- [3] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: An evolutionary marker comes of age, *Nature Reviews Genetics*. 4 (2003) 598-612.
- [4] A. Torroni, A. Achilli, V. Macaulay, M. Richards, H.-. Bandelt, Harvesting the fruit of the human mtDNA tree, *Trends in Genetics*. 22 (2006) 339-345.
- [5] P.A. Underhill, T. Kivisild, Use of y chromosome and mitochondrial DNA population structure in tracing human migrations, *Annual Review of Genetics*. 41 (2007) 539-564.
- [6] C. Krausz, L. Quintana-Murci, G. Forti, Y chromosome polymorphisms in medicine, *Ann.Med.* 36 (2004) 573-583.
- [7] C. Krausz, E. Guarducci, L. Becherini, S. Degl'Innocenti, L. Gerace, G. Balercia, et al., The clinical significance of the POLG gene polymorphism in male infertility, *J.Clin.Endocrinol.Metab.* 89 (2004) 4292-4297.
- [8] R.W. Taylor, D.M. Turnbull, Mitochondrial DNA mutations in human disease, *Nature Reviews Genetics*. 6 (2005) 389-402.
- [9] B.D. Blansit, Resources for genetic genealogy, *Journal of Electronic Resources in Medical Libraries*. 3 (2006) 23-35.
- [10] M.D. Shriver, R.A. Kittles, Genetic ancestry and the search for personalized genetic histories, *Nature Reviews Genetics*. 5 (2004) 611-618.
- [11] G. Biondi, O. Rickards, *Umani Da Sei Milioni Di Anni: L'evoluzione Della Nostra Specie*, 1st ed., Carocci, Roma, 2009.
- [12] S. Anderson, A.T. Bankier, B.G. Barrell, Sequence and organization of the human mitochondrial genome, *Nature*. 290 (1981) 457-465.
- [13] 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.

- [14] N. Howell, J.L. Elson, C. Howell, D.M. Turnbull, Relatives rates of evolution in the coding and control regions of African mtDNAs. *Mol Biol Evol.* 24(10) (2007) 2213-2221.
- [15] B. Pakendorf, M. Stoneking, Mitochondrial DNA and human evolution, *Annual Review of Genomics and Human Genetics.* 6 (2005) 165-183.
- [16] M. Hasegawa, A. Di Rienzo, T.D. Kocher, A.C. Wilson, Toward a more accurate time scale for the human mitochondrial DNA tree, *J.Mol.Evol.* 37 (1993) 347-354.
- [17] S. Meyer, G. Weiss, A. Von Haeseler, Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA, *Genetics.* 152 (1999) 1103-1110.
- [18] N. Galtier, D. Enard, Y. Radondy, E. Bazin, K. Belkhir, Mutation hot spots in mammalian mitochondrial DNA, *Genome Res.* 16 (2006) 215-222.
- [19] M. Ingman, U. Gyllensten, Rate variation between mitochondrial domains and adaptive evolution in humans, *Hum.Mol.Genet.* 16 (2007) 2281-2287.
- [20] T. Kivisild, M. Metspalu, H.J. Bandelt, M. Richards, R. Villems, The world mtDNA phylogeny, in: Bandelt H.J., Macaulay V., Richards M. (Eds.), *Human Mitochondrial DNA and the Evolution of Homo sapiens*, Springer, 2006, pp. 149-179.
- [21] B.A. Malyarchuk, I.B. Rogozin, V.B. Berikov, M.V. Derenko, Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region, *Hum.Genet.* 111 (2002) 46-53.
- [22] 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.
- [23] D. Mishmar, E. Ruiz-Pesini, P. Golik, V. Macaulay, A.G. Clark, S. Hosseini, et al., Natural selection shaped regional mtDNA variation in humans, *Proc.Natl.Acad.Sci.U.S.A.* 100 (2003) 171-176.
- [24] 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.* 308 (2005) 1034-1036.
- [25] H.-. Bandelt, A. Achilli, Q.-. Kong, A. Salas, S. Lutz-Bonengel, C. Sun, et al., Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies, *Biochem.Biophys.Res.Commun.* 333 (2005) 122-130.
- [26] H.J. Bandelt, M. Richards, V. Macaulay, *Human mitochondrial DNA and the evolution of Homo sapiens*, Heidelberg: Springer. (2006).

- [27] Q.P. Kong, H.J. Bandelt, C. Sun, Y.G. Yao, A. Salas, A. Achilli, et al., Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations, *Hum.Mol.Genet.* 15 (2006) 2076-2086.
- [28] M. Richards, The mitochondrial DNA tree and forensic science, *International Congress Series.* 1261 (2004) 91-93.
- [29] A. Salas, H.-. Bandelt, V. Macaulay, M.B. Richards, Phylogeographic investigations: The role of trees in forensic genetics, *Forensic Sci.Int.* 168 (2007) 1-13.
- [30] 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.
- [31] 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.
- [32] A. Torroni, R.I. Sukernik, T.G. Schurr, Y.B. Starikovskaya, M.F. Cabell, M.H. Crawford, et al., mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans, *Am.J.Hum.Genet.* 53 (1993) 591-608.
- [33] M. Slatkin, Linkage disequilibrium in growing and stable populations, *Genetics.* 137 (1994) 331-336.
- [34] M. Nei, A.K. Roychoudhury, Sampling Variances of Heterozygosity and Genetic Distance, *Genetics.* 76 (1974) 379-390.
- [35] S.J. Novak, R.N. Mack, Genetic bottlenecks in alien plant species: Influence of mating systems and introduction dynamics, *Species Invasions: Insights into Ecology, Evolution, and Biogeography.* (2005) 201-228.
- [36] M.C. Davis, S.J. Novak, G. Hampikian, Mitochondrial DNA analysis of an immigrant basque population: Loss of diversity due to founder effects, *Am.J.Phys.Anthropol.* (2011).
- [37] M.A. Jobling, P. Gill, Encoded evidence: DNA in forensic analysis, *Nature Reviews Genetics.* 5 (2004) 739-751.
- [38] 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.
- [39] S. Ramachandran, O. Deshpande, C.C. Roseman, N.A. Rosenberg, M.W. Feldman, L.L. Cavalli-Sforza, Support from the relationship of genetic and geographic in human populations for a serial founder effect originating in Africa, *Proc.Natl.Acad.Sci.U.S.A.* 102 (2005) 15942-15947.

- [40] 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, *The American Journal of Human Genetics*. 83 (2008) 725-736.
- [41] E.G. Puffenberger, Genetic heritage of the Old Order Mennonites of southeastern Pennsylvania, *American Journal of Medical Genetics - Seminars in Medical Genetics*. 121 C (2003) 18-31.
- [42] M. Carvalho, M. Anjos, L. Andrade, V. Lopes, M.V. Santos, J. Gamero, et al., Y-chromosome STR haplotypes in two population samples: Azores Islands and Central Portugal, *Forensic Sci.Int.* 134 (2003) 29-35.
- [43] A. Salas, M. Richards, M.V. Lareu, R. Scozzari, A. Coppa, A. Torroni, et al., The African diaspora: mitochondrial DNA and the Atlantic slave trade, *Am.J.Hum.Genet.* 74 (2004) 454-465.
- [44] G. Bedoya, P. Montoya, J. García, I. Soto, S. Bourgeois, L. Carvajal, et al., Admixture dynamics in Hispanics: A shift in the nuclear genetic ancestry of a South American population isolate, *Proc.Natl.Acad.Sci.U.S.A.* 103 (2006) 7234-7239.
- [45] M. Via, C. Gignoux, E.G. Burchard, The 1000 Genomes Project: new opportunities for research and social challenges, *Genome medicine*. 2 (2010) 1-3.
- [46] J.A. Irwin, J.L. Saunier, H. Niederstätter, K.M. Strouss, K.A. Sturk, T.M. Diegoli, et al., Investigation of heteroplasmy in the human mitochondrial DNA control region: A synthesis of observations from more than 5000 global population samples, *J.Mol.Evol.* 68 (2009) 516-527.
- [47] J.A. Irwin, W. Parson, M.D. Coble, R.S. Just, mtGenome reference population databases and the future of forensic mtDNA analysis, *Forensic Science International: Genetics*. 5 (2011) 222-225.
- [48] H.-. Bandelt, A. Salas, Current Next Generation Sequencing technology may not meet forensic standards, *Forensic Science International: Genetics*. (2011).
- [49] 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.
- [50] J.J. Mulero, C.W. Chang, L.M. Calandro, R.L. . Green, Li Y, C.L. Johnson, et al., Development and Validation of the Amp[®]STR[®] Yfiler[™] PCR Amplification Kit: A Male Specific, Single Amplification 17 Y-STR Multiplex System, *J.Forensic Sci.* 5 (2006) 64-75.

- [51] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: An integrated software package for population genetics data analysis, *Evolutionary Bioinformatics Online*. 1 (2005) 47.
- [52] 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.
- [53] H.Y. Lee, I. Song, E. Ha, S.B. Cho, W.I. Yang, K.J. Shin, mtDNAmanager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences, *BMC Bioinformatics*. 9 (2008) 483.
- [54] W. Ludwig, O. Strunk, R. Westram, L. Richter, H. Meier, A. Yadhukumar, et al., ARB: A software environment for sequence data, *Nucleic Acids Res.* 32 (2004) 1363-1371.
- [55] J. Felsenstein, An alternating least squares approach to inferring phylogenies from pairwise distances, *Syst.Biol.* 46 (1997) 101-111.
- [56] J. Felsenstein, Comparative methods with sampling error and within-species variation: Contrasts revisited and revised, *Am.Nat.* 171 (2008) 713-725.
- [57] R.D. Page, Visualizing phylogenetic trees using TreeView. *Current protocols in bioinformatics / editorial board, Andreas D.Baxevanis ...[et al.]*. Chapter 6 (2002).
- [58] M. Stephens, P. Donnelly, *Inference in molecular population genetics*, *R.Stat.Soc.Ser.B.* (2000).
- [59] J.K. Pritchard, M. Stephens, N.A. Rosenberg, P. Donnelly, Association mapping in structured populations, *Am.J.Hum.Genet.* 67 (2000) 170-181.
- [60] M.J. Hubisz, D. Falush, M. Stephens, J.K. Pritchard, Inferring weak population structure with the assistance of sample group information, *Molecular Ecology Resources*. 9 (2009) 1322-1332.
- [61] 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.
- [62] M. Santos, R.H. Ward, R. Barrantes, mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica, *Hum.Biol.* 66 (1994) 963-977.
- [63] O. Batista, C.J. Kolman, E. Bermingham, Mitochondrial DNA diversity in the Kuna Amerinds of Panama, *Hum.Mol.Genet.* 4 (1995) 921-929.

[64] C.J. Kolman, E. Bermingham, R. Cooke, R. Ward, T.D. Arias, F. Guionneau-Sinclair, Reduced mtDNA diversity in the Ngobe Amerinds of Panama, *Genetics*. 140 (1995) 275.

[65] R.H. Ward, A. Redd, D. Valencia, B. Frazier, S. Pääbo, Genetic and linguistic differentiation in the Americas, *Proc.Natl.Acad.Sci.U.S.A.* 90 (1993) 10663.

[66] G.F. Shields, A.M. Schmiechen, B.L. Frazier, A. Redd, M.I. Voevoda, J.K. Reed, et al., mtDNA sequences suggest a recent evolutionary divergence for Beringian and Northern North American populations, *Am.J.Hum.Genet.* 53 (1993) 549-562.

[67] J. Saillard, P. Forster, N. Lynnerup, H.-. Bandelt, S. Norby, mtDNA variation among Greenland Eskimos: The edge of the Beringian expansion, *Am.J.Hum.Genet.* 67 (2000) 718-726.

[68] R.S. Malhi, J.A. Eshleman, J.A. Greenberg, D.A. Weiss, B.A.S. Shook, F.A. Kaestle, et al., The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America, *The American Journal of Human Genetics*. 70 (2002) 905-919.

[69] L.D. Green, J.N. Derr, A. Knight, mtDNA affinities of the peoples of north-central Mexico, *Am.J.Hum.Genet.* 66 (2000) 989-998.

[70] C. Ginther, D. Corach, G.A. Penacino, J.A. Rey, F.R. Carnese, M.H. Hutz, et al., Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. *EXS.* 67 (1993) 211-219.

[71] J. Alves-Silva, M. Da Silva Santos, P.E.M. Guimaraes, A.C.S. Ferreira, H. Bandelt, S.D.J. Pena, et al., The ancestry of Brazilian mtDNA lineages, *Am.J.Hum.Genet.* 67 (2000) 444-461.

[72] R.H. Ward, F.M. Salzano, S.L. Bonatto, M.H. Hutz, C.E.A. Coimbra Jr., R.V. Santos, Mitochondrial DNA polymorphism in three Brazilian Indian tribes, *Am.J.Hum.Biol.* 8 (1996) 317-323.

[73] O. Rickards, C. Martínez-Labarga, J.K. Lum, G.F. De Stefano, R.L. Cann, mtDNA history of the Cayapa Amerinds of Ecuador: Detection of additional founding lineages for the native American populations, *Am.J.Hum.Genet.* 65 (1999) 519-530.

[74] S. Fuselli, E. Tarazona-Santos, I. Dupanloup, A. Soto, D. Luiselli, D. Pettener, Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders, *Mol.Biol.Evol.* 20 (2003) 1682.

[75] A. Corella, F. Bert, A. Pérez-Pérez, M. Gené, D. Turbón, Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Mosesten, Aymara and Quechua, *Ann.Hum.Biol.* 34 (2007) 34-55.

- [76] F. Bert, A. Corella, M. Gene, A. Pérez-Pérez, D. Turbón, Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands, *Ann.Hum.Biol.* 31 (2004) 9-28.
- [77] G.S. Cabana, D.A. Merriwether, K. Hunley, D.A. Demarchi, Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation, *Am.J.Phys.Anthropol.* 131 (2006) 108-119.
- [78] M. Moraga, E. Aspillaga, P. Carvallo, F. Rothhammer, Analyses of mitochondrial DNA polymorphisms in skeletal remains and extant populations of Northern Chile, *Chungara.* 32 (2000) XXXXV-XXXXVI.
- [79] A.R. Rogers, H. Harpending, Population growth makes waves in the distribution of pairwise genetic differences, *Mol.Biol.Evol.* 9 (1992) 552-569.
- [80] N.J.R. Fagundes, R. Kanitz, S.L. Bonatto, A reevaluation of the native American MtDNA genome diversity and its bearing on the models of early colonization of Beringia, *PLoS ONE.* 3 (2008).
- [81] E. Tamm, T. Kivisild, M. Reidla, M. Metspalu, D.G. Smith, C.J. Mulligan, et al., Beringian standstill and spread of native American founders, *PLoS ONE.* 2 (2007).
- [82] 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.
- [83] R. Ivanova, A. Astrinidis, S. Djoulah, V. Lepage, E. Wijnen, J. Hors, et al., Mitochondrial DNA polymorphisms of a west Algerian population (Oran region), *Biomedicine and Pharmacotherapy.* 53 (1999) 386-392.
- [84] K. Fadhlou-Zid, L. Rodríguez-Botigué, N. Naoui, A. Benammar-Elgaaied, F. Calafell, D. Comas, Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley, *Am.J.Phys.Anthropol.* 145 (2011) 107-117.
- [85] A.M. González, V.M. Cabrera, J.M. Larruga, A. Tounkara, G. Noumsi, B.N. Thomas, et al., Mitochondrial DNA variation in mauritania and mali and their genetic relationship to other Western Africa populations, *Ann.Hum.Genet.* 70 (2006) 631-657.
- [86] S. Plaza, A. Salas, F. Calafell, F. Corte-Real, J. Bertranpetit, Á. Carracedo, et al., Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola, *Hum.Genet.* 115 (2004) 439-447.
- [87] T. Kivisild, M. Reidla, E. Metspalu, A. Rosa, A. Brehm, E. Pennarun, et al., Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears, *Am.J.Hum.Genet.* 75 (2004) 752-770.

- [88] M. Hedman, A. Brandstatter, V. Pimenoff, P. Sistonen, J.U. Palo, W. Parson, et al., Finnish mitochondrial DNA HVS-I and HVS-II population data, *Forensic Sci.Int.* 172 (2007) 171-178.
- [89] H. Wittig, A. Aschenbrenner, M. Poetsch, D. Krause, E. Lignitz, Variability of mitochondrial DNA in a population sample from Iceland, *Leg.Med.* 5 (2003) S173-S176.
- [90] A.O. Tillmar, M.D. Coble, T. Wallerström, G. Holmlund, Homogeneity in mitochondrial DNA control region sequences in Swedish subpopulations, *Int.J.Legal Med.* 124 (2010) 91-98.
- [91] F. Rousselet, P. Mangin, Mitochondrial DNA polymorphisms: A study of 50 French Caucasian individuals and application to forensic casework, *Int.J.Legal Med.* 111 (1998) 292-298.
- [92] R. Piercy, K.M. Sullivan, N. Benson, P. Gill, The application of mitochondrial DNA typing to the study of white Caucasian genetic identification, *Int.J.Legal Med.* 106 (1993) 85-90.
- [93] S. Tetzlaff, A. Brandstatter, R. Wegener, W. Parson, V. Weirich, Mitochondrial DNA population data of HVS-I and HVS-II sequences from a northeast German sample, *Forensic Sci.Int.* 172 (2007) 218-224.
- [94] L. Prieto, B. Zimmermann, A. Goios, A. Rodriguez-Monge, G.G. Paneto, C. Alves, et al., The GHEP-EMPOP collaboration on mtDNA population data—A new resource for forensic casework, *Forensic Science International: Genetics*. In Press, Corrected Proof (2011).
- [95] 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.
- [96] C.J. Kolman, N. Sambuughin, E. Bermingham, Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders, *Genetics*. 142 (1996) 1321.
- [97] M. Tanaka, V.M. Cabrera, A.M. Gonzalez, J.M. Larruga, T. Takeyasu, N. Fuku, et al., Mitochondrial genome variation in eastern Asia and the peopling of Japan, *Genome Res.* 14 (2004) 1832-1850.
- [98] Q.P. Kong, Y.G. Yao, M. Liu, S.P. Shen, C. Chen, C.L. Zhu, et al., Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China, *Hum.Genet.* 113 (2003) 391-405.
- [99] L. Quintana-Murci, R. Chaix, R.S. Wells, D.M. Behar, H. Sayar, R. Scozzari, et al., Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor, *Am.J.Hum.Genet.* 74 (2004) 827-845.

[100] H. Li, X. Cai, E.R. Winograd-Cort, B. Wen, X. Cheng, Z. Qin, et al., Mitochondrial DNA diversity and population differentiation in southern East Asia, *Am.J.Phys.Anthropol.* 134 (2007) 481-488.

[101] T. Goebel, M.R. Waters, D.H. O'Rourke, The Late Pleistocene dispersal of modern humans in the Americas, *Science.* 319 (2008) 1497-1502.

[102] J.A. Lorente, J.C. Monterrose, L.J. Martinez-Gonzalez, E.A. Martinez-Espin J.C., C. Entrala, Fernandez-Rosado, F.J. Morales de Monterrosa, J.A., Mitochondrial DNA Population Data From El Salvador (Central America), *Proceedings of the American Academy of Forensic Sciences.* 13 (2007).

[103] R. Chakraborty, Estimation of race admixture. A new method, *Am.J.Phys.Anthropol.* 42 (1975) 507-511.

[104] J.C. Long, The genetic structure of admixed populations, *Genetics.* 127 (1991) 417-428.

Capítulo V: Problemas de Identificación en Poblaciones Pequeñas y Aisladas

Chapter V: Possible Troubles in DNA Identification of Isolated Populations



ABSTRACT

Some problems that can appear in forensic practice have been detected though the parameters obtained in the study of the Guatemalan population proves that the selected markers are extremely informative and the performed analyses give these markers a high discrimination power.

The aim of this study is to analyze 13 autosomal STRs and determine some statistical forensic parameters, such as, Observed Heterozygosity (H_o) and Power of Discrimination (PD), in a sample from isolated and non-isolated populations. Studies about these types of populations disclose that the increase of homozygosity and the decrease of genotypic variability [1]. So the power of discrimination, typical paternity index and power of exclusion may be lower too. It is also shown in some cases; where the effects of this kind of population will make it difficult when trying to make a correct identification of first degree relatives. Our results may contribute to alert the scientific community, and offer a change for the common protocol to be used in case of isolated populations. This protocol will try to avoid uncertain results. In other cases like big mass disasters, identification paternity tests or other forensic case work, new procedures should be assumed to collect and process the samples.

1. INTRODUCCIÓN

Dentro de las poblaciones de Guatemala se encuentran pequeños grupos poblacionales como Mam, Kaqchiquel, K'iche' y Q'eqchi'. Aunque entre ellos no existen grandes barreras geográficas, sí que se pueden encontrar signos de aislamiento al estudiar su estructura genética. Al igual que ocurre en Guatemala, se pueden encontrar poblaciones aisladas en muchas regiones de todo el mundo. Normalmente estas poblaciones tienen una identidad muy marcada y su aislamiento puede deberse a factores culturales, lingüísticos o geográficos. Desde hace mucho tiempo se ha descrito que en grupos aislados se pueden encontrar fenómenos característicos, desde el punto de vista genético, debido al fenómeno de endogamia [2]. Esto puede suponer un grave problema al abordar muchos casos en identificación genética.

Generalmente los problemas de endogamia en una comunidad suelen aparecer debido a las relaciones dentro de la misma, sobretodo, si es pequeña. Esto puede ocurrir en comunidades aborígenes que quieren preservar su identidad cultural, donde prefieren, al establecerse nuevas familias, que sean fruto de la unión de miembros de su misma comunidad o de comunidades con características culturales muy similares, por lo que usualmente tienen un mismo origen. Normalmente, esta preferencia de uniones consanguíneas, viene siendo una pauta generalizada en el pasado y una tradición que se conservará en el futuro en muchas de estas etnias. Es por esto, que en el presente, aún muchas comunidades, buscan preservar su identidad, sus características culturales; ya sea por razones religiosas, o por conservar su tradición lingüística, artística, etc. [3]

Desde que en Medicina Legal se comenzaron a utilizar marcadores genéticos para realizar identificaciones, y aun más con el uso de los STRs que se aplican diariamente a millones de casos forenses, el uso de estos ha ido siempre acompañado del estudio de la población donde se va a utilizar [4,5]. El cálculo

de las frecuencias de los marcadores en la población, así como su validez, depende de la discriminación que ofrezca para un tamaño determinado de población. Gracias a los estudios poblacionales, se ha podido observar la gran diferencia interpoblacional en las frecuencias de cada una de las posibilidades alélicas que pueden aparecer, variando más cuanto más distantes son las poblaciones. De la misma forma, estos mismos marcadores son frecuentemente usados para hacer estudios de variabilidad interpoblacional, estudios de mestizaje o para conocer a fondo la estructura genética de una población [6-8].

1.1 PROBLEMAS DE IDENTIFICACIÓN

Aunque los parámetros obtenidos en nuestro estudio poblacional de Guatemala, demuestran que los marcadores estudiados son muy informativos, y los análisis realizados otorgan a los marcadores estudiados un alto poder de discriminación (ver tabla V.1), se han detectado problemas que pueden aparecer en la práctica forense, si no se toman las suficientes medidas de precaución.

Tabla V.1: Tipo de marcador, población, poder de discriminación calculado y diversidad haplotípica en la población de Guatemala.

Marcador	Población	PD	Diversidad haplotípica
a_STRs	Mestizos	0,9999999999999996	1
	Kaqchiquel	0,99999999999996924	1
	K'iche'	0,9999999999999597	1
	Mam	0,9999999999999322	1
	Q'eqchi'	0,9999999999999683	1
Y_STRs	Mestizos	0,999999989	1
	Mayas	0,999999403	1
ADNmt	Mestizos	0,7143	0,9769
	Kaqchiquel	0,8163	0,9943
	K'iche'	0,8125	0,9928
	Mam	0,8511	0,9926
	Q'eqchi'	0,7872	0,9926

En el trabajo diario con poblaciones pequeñas y aisladas es frecuente que se presenten algunos problemas a la hora de realizar identificaciones cuando sólo se dispone de un familiar cercano, incluso en el caso de que el familiar que se

encuentre sea un familiar de primer grado; ya sea madre o padre del individuo estudiado.

En el laboratorio se han detectado algunos casos en individuos aborígenes de Guatemala, que alertan de los problemas que se pueden encontrar en este tipo de poblaciones.

Se presentan dos casos de identificación, donde en un primer momento sólo se contaba con las muestras del padre y del hijo. Al realizar el análisis de los marcadores con el kit AmpFlSTR® Identifiler® de Applied Biosystems, Foster City, CA [9] aparecían marcadores donde no existía compatibilidad, no obstante podían ser explicados por la tasa de mutación de los mismos. (Ver tabla V.2)

Tabla V.2: Perfiles genéticos de dos casos de identificación en comunidades mayas de Guatemala. (Caso 1: menor y supuesto padre; Caso 2: menor y supuesto padre)

Muestra	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH0	D13S31	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Caso 1 Menor	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
Caso 1 Supuesto Padre	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
Caso 2 Menor	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
Caso 2 Supuesto Padre	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

El índice y la probabilidad de paternidad (IP y P en adelante) en ambos casos eran muy bajos, pero en ninguno de los dos casos se puede determinar claramente la exclusión. Al tener en cuenta las tasas de mutación los valores fueron $IP=7.4343 \cdot 10^{-5}$ y $P=27.01\%$ en el primer caso y en el segundo fueron $IP=6.8335 \cdot 10^{-5}$ y $P=6.40\%$.

Es frecuente que en estas poblaciones los índices de paternidad no sean muy altos debido a la elevada endogamia. Hay muchos marcadores con repeticiones muy prevalentes en la población lo que hace más difícil poder asumir una paternidad en estas poblaciones.

Tras encontrar estos datos inconcluyentes, se pasó a solicitar muestras de un segundo familiar, para poder confirmar la exclusión de éste. Gracias a la segunda muestra se confirmó, con facilidad, la exclusión de la paternidad, ver tabla V.3 donde se puede ver como aumentan los marcadores que excluyen la paternidad, en comparación con los resultados obtenidos en la tabla V.2.

Tabla V.3: Perfiles genéticos de dos casos de identificación en comunidades maya de Guatemala. (Caso 1: madre, menor y supuesto padre; caso 2: madre, menor y supuesto padre).

Muestra	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH0	D13S31	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Caso 1 Madre	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
Caso 1 Menor	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
Caso 1 Supuesto Padre	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
Caso 2 Madre	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
Caso 2 Menor	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
Caso 2 Supuesto Padre	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

Conocer la problemática de la población de la cual procedían estas muestras, llevó a estudiar con más detalle poblaciones similares, pequeñas y aisladas, comparadas con otras poblaciones. El estudio de los parámetros estadísticos más descriptivos dentro de la identificación genética, dará información sobre el comportamiento de las poblaciones pequeñas con respecto a otro tipo de poblaciones no aisladas. El estudio de los loci más utilizados en identificación permitirá ver si existen diferencias significativas entre unas y otras poblaciones. También estudiar este tipo de poblaciones, así como casos particulares de identificación genética ayudará a determinar nuevos protocolos de actuación para proceder en casos similares a los que se describen en el presente capítulo.

1.2 PODER DE DISCRIMINACIÓN, EXCLUSIÓN Y HOMOCIGOSIDAD

Aunque en el capítulo I se describió el poder de discriminación y de exclusión, hay que recordar, en este momento, que la discriminación en identificación genética es una medida relativa de la eficacia del sistema o sistemas analizados (loci STRs). Al realizar la identificación de una muestra con un determinado patrón genético, se obtendrá distinta probabilidad de discriminación en función del laboratorio donde se realiza, dependiendo básicamente de dos variables, los marcadores STRs escogidos y la frecuencia de estos en la población. Una probabilidad de discriminación cercana a 1 es lo más favorable para realizar una identificación. La exclusión e inclusión, al comparar dos muestras al azar,

viene determinada igualmente por estos dos parámetros, los loci estudiados y la frecuencia de sus variables en la población.

La heterocigosidad observada en una población es el número de individuos heterocigóticos que se encuentran, con respecto al número de individuos estudiados.

$$H_o = \frac{\text{nº individuos heterocigotos}}{\text{nº total de individuos}}$$

La heterocigosidad esperada (heterocigosidad media o diversidad genética media) (H_e) [10] en una población es la que se espera encontrar en un locus si el apareamiento fuese al azar.

$$H_e = \frac{\sum_{j=1}^n H_j}{n} = \frac{\sum_{j=1}^n \left(1 - \sum_{i=1}^m p_{ij}^2 \right)}{n}$$

H_j : Heterocigosidad esperada en un locus j
 P_{ij} : Frecuencia del alelo i del locus j
 m : Número de alelos en el locus j
 n : Número de loci considerados

La heterocigosidad así como la homocigosidad, son parámetros que indican el polimorfismo y eficacia de un marcador genético.

En una población que se encuentra en equilibrio Hardy-Weinberg es posible estimar la heterocigosidad media examinando la proporción de heterocigotos.

Normalmente las poblaciones pequeñas pueden presentar algunas desviaciones de estos parámetros. La consanguinidad generalmente disminuirá la variabilidad y aumentará el número de individuos homocigóticos.

1.3 PROBLEMAS EN BASES DE DATOS DE IDENTIFICACIÓN CIVIL

Bases de datos forenses civiles: su único fin es la identificación de personas (niños y adultos) desaparecidos, lo cual se hace comparando el ADN de la personas no identificadas con el de los familiares. Por sus características humanitarias, es requerimiento legal y ético que los familiares colaboren voluntariamente y firmen un consentimiento informado (esto las diferencia de las bases de datos criminales).

Que las bases de datos de identificación civil se fundamenten en los datos genéticos de familiares, hace que sea muy importante el conocimiento del origen de los familiares, así como las características de la población. En poblaciones aisladas, el alto grado de endogamia así como la disminución del poder de discriminación de los marcadores aSTRs hace que para la inclusión de individuos de estas poblaciones en la base de datos se haya de tener en cuenta un mayor número de marcadores que permitan discriminar mejor entre individuos.

Dentro del Laboratorio de Identificación Genética, del Departamento de Medicina Legal y Toxicología de la Universidad de Granada se implementan diariamente distintas bases de datos de identificación civil dentro del Programa Fénix, y el Proyecto DNA-PROKIDS, los problemas en los que se centra este capítulo pueden aparecer, y son de mayor aplicación en el segundo proyecto, pero se hará una pequeña reseña de ambas:

1.3.1 Personas Desaparecidas (Programa FÉNIX)

En 1998 se creó en la universidad de Granada una base de datos genética civil. Su finalidad era prestar ayuda a los familiares de las personas desaparecidas, (Programa FENIX). Este programa fue pionero en el mundo entero ya que anteriormente no existía ni existe ninguna base de datos con estos fines. Esta iniciativa de un grupo de investigación de la Universidad de Granada fue apoyada por la Guardia Civil, lo que permitió la creación de un programa Nacional de identificación de cadáveres y de restos de personas desaparecidas [11]. El programa FENIX contiene dos bases de datos independientes donde se pueden comparar automáticamente secuencias de ADN para identificar perfiles relacionados entre sí. Una se conoce como Base de Datos de Referencia que contiene secuencias de ADN mitocondrial de parientes que provienen de la línea materna así como información de microsatélites de ADN nuclear. Debido a que esta base de datos es una base de datos no criminal, sólo se encuentra en ella aquellos individuos que donan su ADN de forma voluntaria y siempre tras firmar un documento de consentimiento. La otra base de datos se denomina Base de Datos Dubitada y contiene todos los perfiles de restos encontrados que no han podido ser identificados por métodos clásicos, identificación de huellas dactilares, parámetros

antropológicos, odontología, rayos X, etc. Actualmente, 862 familias de personas desaparecidas participan en el programa FENIX y se han logrado 319 identificaciones de restos encontrados [11].

1.3.2 Tráfico de Seres Humanos (Iniciativa DNA-PROKIDS)

DNA-PPROKIDS es un proyecto internacional centrado en la lucha contra el tráfico de seres humanos a través de la identificación genética de las víctimas y sus familiares, especialmente en niños. Su meta no se centra únicamente en la resolución de crímenes domésticos sino que se pretenden extender los resultados a nivel mundial con el objetivo de estimular la lucha contra el tráfico de seres humanos y ayudar a prevenir la misma, <http://www.dna-prokids.org/> [12].

La principal misión de DNA-PROKIDS es la identificación de las víctimas y la reunión de éstas con sus familias. Por otro lado, el proyecto DNA-PROKIDS pretende estudiar y aplicar mejoras para ayudar a la policía y al sistema judicial a luchar contra el tráfico de seres humanos de una manera más efectiva. Además, se pretende analizar y sugerir una legislación común para solucionar el problema del tráfico humano. Otro de los objetivos es solventar los problemas sociales y de comunicación que dificultan la lucha contra el tráfico de seres humanos. Por último, se pretende preparar y coordinar a expertos en identificación genética de diferentes países.

Hasta la fecha, los países participantes en la iniciativa DNA-PROKIDS han analizado más de 2.500 casos (principalmente de México, Guatemala, El Salvador, Paraguay, Perú, Bolivia en Latino América; y Filipinas, Tailandia, Indonesia e India en Asia). Los análisis de ADN y la posterior aplicación de los datos que acompañan han ayudado a la identificación de más de 330 niños perdidos que han sido reunidos con sus familias.

2. MATERIALES Y MÉTODOS

2.1 POBLACIONES Y MUESTRAS

Para realizar este estudio se han examinado las diferencias entre los análisis realizados con STRs autosómicos (aSTRs) en distintas poblaciones; las poblaciones estudiadas, poblaciones analizadas en el Laboratorio de Identificación Genética de la Universidad de Granada y otras tomadas de datos publicados por otros autores. Tanto las poblaciones analizadas en el laboratorio como aquellas tomadas de otros estudios, han sido estudiadas para 15 aSTRs. Al realizar los análisis con distintos kits los loci estudiados no coinciden en todas ellas, por lo que para comparar unas poblaciones con otras se ha reducido el estudio al uso de 13 marcadores STRs (aquellos incluidos en el CODIS).

Las poblaciones se dividieron según sus características en tres grupos distintos: poblaciones pequeñas y aisladas, poblaciones medianas, que aún teniendo un mayor número de individuos, habitan en un lugar no muy grande; y finalmente poblaciones de gran tamaño distribuidas por un amplio territorio.

- Pequeñas: Kaqchiquel, K'iche', Mam, Q'eqchi' (todas ellas de Guatemala), Otomi Ixmiquilpan (Méjico), Huastecos (Méjico), Chakma (Bangladesh), Čakovec (Croacia), Shiite Muslim (India), Adi Panggi y Adi Komkar (ambas de India).
- Medianas: Granada (España), Liaoning (China), Kuala Lumpur, Dobruja (Este de Rumania y Bulgaria), Norte de Polonia, Karamoja (Noreste de Uganda), Rio de Janeiro (Brasil), Este de Timor (Indonesia) y Ovambo (norte de Namibia y el sur de Angola)
- Grandes: Mestizo de Guatemala, España, Chihuahua (Méjico), Paraguay, Macedonia, Marruecos, Somalia, Costa Rica, Brasil y Jalisco (Méjico).

Nota: Las poblaciones subrayadas han sido analizadas en el Laboratorio de Identificación Genética de la Universidad de Granada.

Las poblaciones incluidas en el primer grupo, poblaciones pequeñas, han sido identificadas como poblaciones que por alguna razón han quedado aisladas ya sea por la orogenia de su geografía, por factores lingüísticos o culturales.

Las poblaciones Kaqchiquel, K'iche', Mam, Q'eqchi' y Huastecos tienen como origen la población maya. Aunque cada una de ellas posee su lengua y unas características diferentes, las cuatro comparten una serie de rasgos debido a su origen común. En este momento, los grupos mayas o descendientes son más de una cincuenta, lo que da una señal de cuan aislados se encuentran estos grupos mayas. Aunque todos estos grupos poseen un lenguaje con un origen común, comparten ramas muy alejadas como puede ser el caso de los Huastecos, que se separarían del resto hace unos 4000 años, o los Mam que se separarían de los Kaqchiquel, K'iche' y Q'eqchi' hace unos 2500 años. Estos últimos grupos se sitúan en los altiplanos de Guatemala, mientras que los Huastecos se encuentran en una pequeña zona aislada de los altiplanos y valles cercanos al golfo de Méjico en los estados de Veracruz, Hidalgo y Tamaulipas. Otro grupo, en este caso, no maya pero también afincado en Méjico es el Otomi Ixmiquilpan. En realidad, se trata de una serie de pequeños grupos poblacionales aislados, denominados otomís. Estos grupos mantienen aún un antiguo lenguaje denominado Hñahñu, muy alejado de las lenguas mayas anteriores, dentro del grupo de las lenguas Oto-manguanas.

Los romaníes europeos son un ejemplo de una población errante. Aunque todos ellos poseen un origen indio común, que se ha fragmentado en numerosos subgrupos endogámicos a causa de las migraciones, estas poblaciones poseen patrones complejos en su estructura socio-cultural [13]. Las minorías étnicas romaníes se establecieron en el noroeste de Croacia durante el siglo XIX después de la abolición de la esclavitud en Rumanía. En los últimos tiempos abandonaron algunas de las costumbres y oficios tradicionales pero mantienen un dialecto arcaico del rumano (Limba d 'Bjas)

como lengua predominante. Uno de estos grupos romaníes es el que se denomina Čakovec por encontrarse en esta ciudad en este momento [14].

El grupo musulmán, Shiite Muslim, es una rama de los chiíes. Los chiíes constituyen hoy entre un 10 y un 15% de los musulmanes. Son mayoritarios en Irán, Azerbaiyán, Iraq, Bahreín y el sur del Líbano y existen minorías chiíes en otros lugares, especialmente en Siria, Afganistán y Pakistán. El chiismo septimano existe en la India, Pakistán, Siria y Yemen. El grupo que se estudia se encuentra aislado en Tamilnadu, situado en el extremo sudeste de la India [15]. Aunque hablan la lengua local, tienen su propia lengua llamada *Lisānu l-Dā'wat*. La lengua del Dā'wat" [16]. En esta zona la religión principal es el hinduismo, por lo que esta minoría se encuentra muy aislada por razones culturales y religiosas.

Otras poblaciones al igual que las anteriores, conservan su identidad aborigen, su lenguaje, sus costumbres, religión...pero se mantienen aisladas a consecuencia de grandes barreras geográficas, normalmente, alta montaña. Los Chakmas, un grupo étnico de Bangladesh, siguen hablando Chagma y se mantienen aislados de sus poblaciones vecinas [17]. Viven cerca de las estribaciones del Himalaya, en las altas montañas laterales de la cordillera. También en el Himalaya conviven varios grupos con un mismo origen, que comparten una antigua lengua, Adi, dentro de la familia lingüística Tibeto-Burmana, distribuidos por la India, China y Bután [18], los Adi Panggi y los Adi Komkar. Estas dos tribus se originaron en los siglos V-VII a.c. en el Tíbet y están formadas aproximadamente por 2000-3000 individuos aislados en zonas montañosas remotas. Estos tres últimos grupos viven por encima de los 1700 m de altura [17,18].

Para tener más información sobre los estudios de estas poblaciones, se puede consultar la tabla V.4.

Tabla V.4: Información sobre los tres grupos de poblaciones estudiados. Población código, número de muestras (N), número de loci estudiados (a-STRs) y referencias.

Poblaciones aisladas					
Población	Código	País/Localización	N	a-STRs	Referencias
Kaqchiquel	KaqGU	GUATEMALA	50	15	Martínez-González et al. 2010 Datos no publicados
Kiche	KicGU	GUATEMALA	50	15	Martínez-González et al. 2010 Datos no publicados
Mam	MamGU	GUATEMALA	50	15	Martínez-González et al. 2010 Datos no publicados
Q'eqchi'	QeqGU	GUATEMALA	50	15	Martínez-González et al. 2010 Datos no publicados
Chakma	ChaBA	BANGLADESH	113	15	Gazi et al. 2009 [17]
Čakovec	ČakCR	CROACIA	100	15	Novokmet et al. 2007 [14]
Shiite Muslim	ShiIN	Tamilnadu INDIA	62	15	Eaaswarkhanth et al. 2007 [15]
Otomi Ixmiquilpan	OtiMX	MEXICO	83	15	Barrot et al. 2005 [19]
Huastecos	HuaMX	MEXICO	135	15	Barrot et al. 2005 [19]
Adi Panggi	PanIN	Arunachal Pradesh INDIA	110	15	Krithika et al. 2007 [18]
Adi Komkar	KomIN	Arunachal Pradesh INDIA	63	15	Krithika et al. 2007 [18]
Poblaciones de tamaño medio					
Liaoning	LiaCH	CHINA	358	15	Wang et al. 2008 [20]
Kuala Lumpur	KuLMA	MALASIA	110	15	Maruyama et al. 2008 [21]
Dobruja	DobCR	CROACIA	569	15	Stanciu et al. 2009 [22]
Northern Poland	NPIPO	POLONIA	145	15	Szczerkowska et al. 2004 [23]
Karamoja	KarUG	UGANDA	218	15	Gomes et al. 2009 [24]
Rio de Janeiro	RioJBR	BRASIL	300	15	Góes et al. 2004 [25]
East Timor	ETimET	EAST TIMOR	186	15	Souto et al. 2005 [26]
Ovambo	OvaNA	NAMIBIA	195	15	Muro et al. 2008 [27]
Granada	GraSP	ESPAÑA	49	15	Martínez-González et al. 2010 Datos no publicados
Poblaciones grandes					
Macedonia	MacMA	Macedonia	100	15	Havaš et al. 2007 [28]
Morocco	MorMO	MARRUECOS	425	15	Bouabdellah et al. 2008 [29]
Somalia	SomSO	SOMALIA	404	15	Tillmar et al. 2009 [30]
Mestizos	MesGU	GUATEMALA	200	15	Martínez-Espin et al. 2006 [31]
Spain	SpaSP	ESPAÑA	200	15	Martínez-González et al. 2010 Datos no publicados
Chihuahua	ChiMX	MEXICO	161	15	Martínez-González et al. 2005 [32]
Paraguay	ParPA	PARAGUAY	168	13	Martínez-Espin et al. 2003 [33]
Costa Rica	CosCR	COSTA RICA	260	13	Melendez et al. 2004 [34]
Brazil	BraBR	BRASIL	100	13	Figueiredo et al. 2004 [35]
Jalisco	JalMX	MEXICO	309	15	Rubi-Castellanos et al. 2009 [36]

2.2 ANÁLISIS GENÉTICO

Las muestras estudiadas fueron en su mayoría analizadas mediante el kit AmpFlSTR® Identifiler® de Applied Biosystems, Foster City, CA [9]. Algunas de las poblaciones como es el caso de Paraguay, Costa Rica y Brasil se amplificaron con el kit PowerPlex16® (Promega) [37]. El uso de estos dos kits para el genotipado de las muestras, hace necesario reducir los análisis a los 13 marcadores comunes presentes en el CODIS: D3S1358, TH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1PO, vWA, D8S1179, TPOX y FGA.

En todos los casos se utilizó el analizador genético ABI 310 (Applied Biosystems®) para la electroforesis de los STRs, y los fragmentos fueron analizados por distintas versiones del software GeneScan (Applied Biosystems®).

2.3 ANÁLISIS ESTADÍSTICO

En todos los estudios se calcularon una serie de parámetros que en este momento se recalculan para poder comparar las distintas poblaciones. Se obtuvieron las frecuencias de cada uno de los marcadores, gracias al uso de la hoja de cálculo PowerStats [38] también los parámetros forenses. La heterocigosidad se calculó mediante Arlequin v3.1 [39] y todos los parámetros se compararon mediante SPSS v15.0. Así se obtuvieron el poder de discriminación máximo, mínimo y medio en cada uno de los marcadores; y la media, máximos y mínimos para la heterocigosidad. En el caso del poder de discriminación combinado, para visualizar las diferencias, se usó como recurso el logaritmo decimal, calculando: $-\log_{10} (1-PD \text{ combinado})$; poniendo de manifiesto las diferencias entre las poblaciones. Aunque todos los cálculos se realizaron con SPSS v15.0, se escogió M.O. Excel 2007 para realizar las representaciones.

3. RESULTADOS Y DISCUSIÓN

Se estimaron distintos estadísticos forenses para poder comparar los tres grupos de poblaciones. Con respecto al poder de discriminación medio, se obtuvieron valores muy altos en poblaciones grandes (0,91465177) y de tamaño mediano (0,91629333), en cambio en poblaciones pequeñas y aisladas el valor fue sensiblemente menor (0,84700892).

De igual forma, se observa una disminución en los valores de heterocigosidad (H_o). En los tres grupos se distribuyeron de la siguiente forma: poblaciones grandes un 78,27%, poblaciones medianas 78,56% y 76,62% en poblaciones pequeñas.

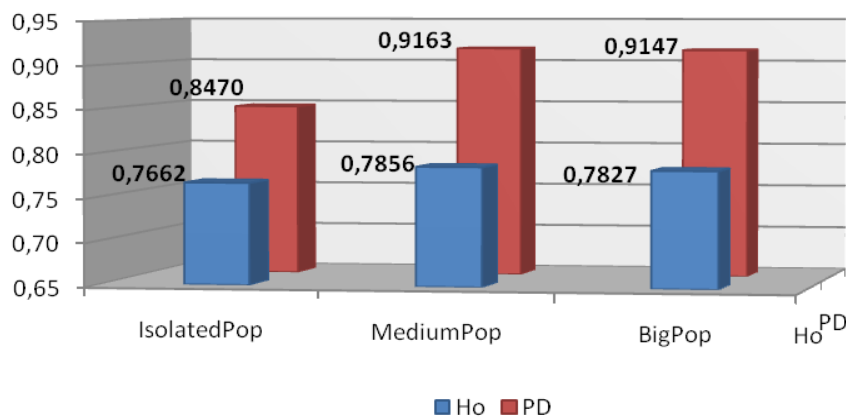


Figura V.1: Histograma de los valores de H_o y PD en los tres tipos de poblaciones estudiadas.

Como se puede ver en la figura V.1, los resultados en poblaciones de tamaño grande y medio, son muy similares, pero, tanto en el caso del poder de discriminación como en la heterocigosidad, los valores son mayores en poblaciones medianas. Esto se debe a que al hacer el estudio de una gran región o gran país, normalmente están incluidas, tanto ciudades grandes y medianas, con gran variabilidad, como pequeños pueblos. Estos últimos en muchas ocasiones se comportan como poblaciones pequeñas y aisladas, lo que se

traduce en un descenso, en una pequeña proporción, de los valores en los estadísticos que se muestran en esta figura. Los valores máximos y mínimos encontrados en las poblaciones, tanto en la heterocigosidad, como en el poder de discriminación medio, son los que registran mayores diferencias. Las poblaciones pequeñas o aisladas muestran un mayor rango entre los valores mínimos y máximos. Se encuentran máximos más cercanos a poblaciones mayores, pero los mínimos nunca se aproximarían a los de las poblaciones medianas o grandes (ver figura V.2 y V.3).

Los estudios realizados en poblaciones aisladas describen frecuentemente, en un gran número de los marcadores estudiados un aumento de la homocigosidad y una disminución del poder de discriminación, esto se hace aún más explícito al compararlos con poblaciones grandes.

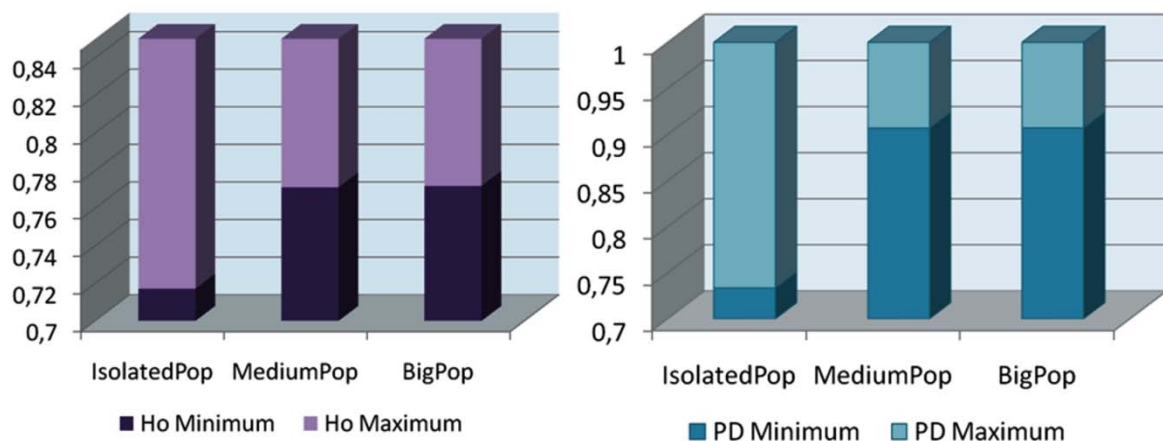


Figura V.2: Representación del Ho máximo y mínimo; figura V.3 Valores de PD máximo y mínimo de cada grupo poblacional.

Se observa que, al estudiar el poder de discriminación y la heterocigosidad, los marcadores varían dependiendo de la distribución. Si se buscan los marcadores con menor y mayor PD, destacan TH01 como el de menor valor y FGA como el de mayor valor en el caso de poblaciones aisladas. En poblaciones medianas destaca D5S818 con valores menores frente a D18S51 que es el mayor; y por último el menos informativo sería TPOX contra D18S51 con mayor poder en las poblaciones más grandes. En el caso de la de la homocigosidad se ven efectos

similares, se comprueba que la heterocigosidad disminuye en marcadores como el TH01 y TPOX, debido evidentemente al menor número de variables en estos, pero no de igual modo en todas las poblaciones estudiadas. En general, se observa que en las poblaciones pequeñas y aisladas existe una pérdida o reducción en el número de variables alélicas.

Locus	PD Med			Ho Med		
	Isolated	Med. size	Big size	Isolated	Med. size	Big size
D8S1179	0,8787682	0,9332667	0,9231352	0,8006645	0,8072556	0,7997732
D21S11	0,90067	0,9487778	0,9418146	0,83881	0,8570111	0,8265023
D7S820	0,8250927	0,9162222	0,9203785	0,7511555	0,7898667	0,7741705
CSF1PO	0,8135882	0,8814333	0,8807923	0,7342555	0,7305	0,7203932
D3S1358	0,7793991	0,8929444	0,9057691	0,6950236	0,7388556	0,7421005
TH01	0,7304964	0,8895778	0,9069241	0,6421973	0,7447	0,7701269
D13S317	0,8855482	0,8972	0,9244729	0,8278782	0,7411222	0,7875423
D16S539	0,8523464	0,9160889	0,9159253	0,7392509	0,7744111	0,7754396
D2S1338	0,91424	0,9582222	0,95175	0,8257827	0,8394556	0,8448413
D19S433	0,86727	0,9292111	0,9154943	0,8260282	0,8090889	0,812984
vWA	0,8431655	0,9307222	0,921301	0,7715509	0,8025111	0,7903342
TPOX	0,7727491	0,8438667	0,8312785	0,6581655	0,6838444	0,6839583
D18S51	0,8956255	0,9615778	0,9639857	0,8029727	0,8695778	0,871587
D5S818	0,8178218	0,8847	0,8659268	0,71836	0,7401111	0,7126286
FGA	0,9283529	0,9605889	0,9625657	0,8613864	0,8555667	0,8587925

Tabla V. 5. Valores medios del PD y Ho por marcador en cada uno de los grupos poblacionales. Donde Isolated, son los valores encontrados en pequeñas poblaciones, Med. Size, poblaciones medianas y Big size, grandes.

El PD combinado representa la capacidad de un sistema de marcadores genéticos de discriminar entre distintos individuos y el resultado siempre debe ser muy próximo a 1. Como se ve en la tabla V.5, el poder combinado disminuye de forma drástica en la mayoría de las poblaciones aisladas estudiadas.

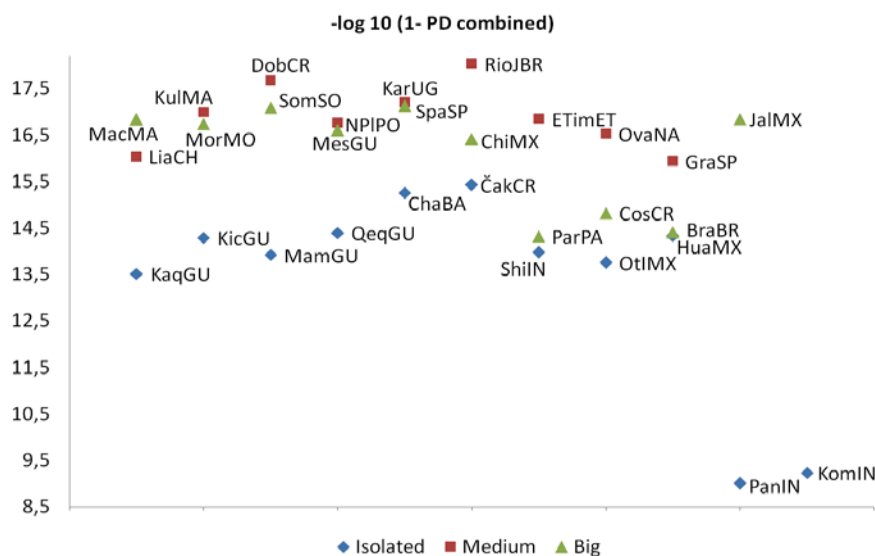


Figura V.4: Representación del logaritmo decimal del PD combinado.

Como se ve en la figura V.4 el poder combinado de los kits de identificación disminuye en las poblaciones aisladas. La representación de la nube de puntos azules corresponde con el grupo de aisladas. Esta nube se haya tres puntos logarítmicos por debajo del resto de las poblaciones, lo que genera una disminución drástica de la probabilidad de identificación. Podría pensarse que es debido al tamaño de población estudiada, pero por esta misma razón se debería observar una disminución en poblaciones de tamaño similar como la población de Granada (n=49) o la de Kuala Lumpur (n=100).

Es importante destacar el aumento del poder de discriminación cuando se incluyen otro tipo de marcadores forenses como pueden ser los STR de cromosoma Y o el ADN mitocondrial. Por esto, es muy importante incluir estos nuevos marcadores en las bases de datos de identificación civil, en algunos de estos programas, como ocurre en el caso del proyecto DNA-PROKIDS, en muchas ocasiones, no se conoce la procedencia de las muestras problema por lo que se debe caracterizar perfectamente los familiares que donen las muestras teniendo en cuenta la población de origen (certificar la caracterización de la población), para así, poder tener una mayor certeza en la identificación de los individuos.

La disminución de la variabilidad observada, debida a la disminución en el PD y en la heterocigosidad, se debe a que las poblaciones quedan aisladas o residen en zonas con barreras geográficas y rara vez interactúan con otras por lo que el pool genético es limitado. En estos grupos aislados se pueden observar los fenómenos característicos de las poblaciones endogámicas.

Las poblaciones aborígenes poseen una carga genética muy conservada, ya que van perdiendo las variables menos numerosas y conservando las más numerosas, lo que produce el aumento de la homocigosidad y pérdida de PD. Se confirma con este estudio la disminución en la presencia de alelos causada, no sólo por un menor tamaño de la población, sino también por el proceso de endogamia.

De igual forma los estudios realizados en la población de Guatemala, confirman que el aumento en la homocigosidad y la caída de PD en cromosomas autosómicos no van ligados a un número limitado de haplotipos en ADNmt y cromosoma Y. La herencia limitada de los marcadores presentes no potencia los efectos de la endogamia.

La obtención de muestras de referencia es muy importante para conseguir una identificación humana eficiente. En contraste con otras, en el estudio de poblaciones endogámicas, se necesitan muestras de referencia apropiadas (uno o varios familiares directos, padres o hijos) y/o las referencias directas, como efectos personales, muestras biológicas (biopsias o cualquier fluido biológico o tejido archivado). Incluso, en algunos casos complejos de identificación de restos degradados, quizás si sólo se posee un familiar directo se debe tener cuidado antes de dar un falso positivo como resultado, tomando familiares que permitan hacer análisis de cromosomas sexuales o ADNmt; o aumentando notablemente el número de marcadores estudiados.

Es necesario generar bases de datos independientes para todas las poblaciones haciendo hincapié en estos pequeños grupos aislados. Aumentando el número de marcadores e incluyendo marcadores de con herencia limitada por el sexo. Es necesario adaptar nuestros protocolos de actuación, tanto en la toma de muestras como en la emisión de resultados, para este tipo de poblaciones.

CONCLUSIONS

The decrease of variability, the PD and the heterozygosity is due to the limited genetic pool of isolated populations. Characteristic phenomena of endogamous populations are noticed in these groups.

Aboriginal populations have a highly conserved genetic burden, lose the less numerous variables and preserve the most common, causing the rise of homozygosity and loss of PD.

Studies performed in the Guatemalan population confirm that the rise of homozygosity and the fall of PD in autosomal chromosomes are not linked to a limited number of DNAmT haplotypes and Y chromosome. The limited inheritance of the markers does not strengthen the effects of endogamy.

The securing of reference samples is very important for an efficient human identification. When endogamous populations are studied it is crucial to have appropriate reference samples (one or more direct relatives, parents or children) and/or direct samples, as personal belongings or biological samples (biopsies or biological fluid or tissue saved).

It is necessary to generate independent databases for all the populations emphasizing isolated small groups as far as raising the number of markers and including markers with sex-linked heredity. It is necessary to adapt our action protocols both in the taking of samples and the emission of results in these populations.

4. BIBLIOGRAFÍA

- [1] E. Mayr, Systematics and the origin of species from the viewpoint of a zoologist, Columbia Biol.Ser. 13 (1942) 334.
- [2] J.D. Thomas, M.M. Douchette, D.C. Thomas, J.D. Stoeckle, Disease, lifestyle, and consanguinity in 58 American Gypsies, Lancet. 2 (1987) 377-379.
- [3] A. Deruyttere, Pueblos indígenas, globalización y desarrollo con identidad: algunas reflexiones de estrategia, InterAmerican Bank for Development: IADB. (2001).
- [4] B. Devlin, Forensic inference from genetic markers. Stat.Methods Med.Res. 2 (1993) 241-262.
- [5] W.K. Fung, Statistical issues in comparing random DNA samples [1], Forensic Sci.Int. 78 (1996) 231-232.
- [6] E. Meyer, P. Wiegand, B. Brinkmann, Phenotype differences of STRs in 7 human populations, Int.J.Legal Med. 107 (1995) 314-322.
- [7] E. Bosch, F. Calafell, A. Pérez-Lezaun, J. Clarimón, D. Comas, E. Mateu, et al., Genetic structure of north-west Africa revealed by STR analysis, European Journal of Human Genetics. 8 (2000) 360-366.
- [8] E. Bosch, F. Calafell, A. Pérez-Lezaun, D. Comas, H. Izaabel, O. Akhayat, et al., Y chromosome STR haplotypes in four populations from northwest Africa, Int.J.Legal Med. 114 (2000) 36-40.
- [9] P.J. Collins, L.K. Hennessy, C.S. Leibelt, R.K. Roby, D.J. Reeder, P.A. Foxall, 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 (2004) 1265-1277.
- [10] M. Nei, A.K. Roychoudhury, Sampling Variances of Heterozygosity and Genetic Distance, Genetics. 76 (1974) 379-390.
- [11] J.K. Pritchard, M. Stephens, N.A. Rosenberg, P. Donnelly, Association mapping in structured populations, Am.J.Hum.Genet. 67 (2000) 170-181.
- [12] D.G. Monckton, R. Neumann, T. Guram, N. Fretwell, K. Tamaki, A. MacLeod, et al., Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism, Nat.Genet. 8 (1994) 162-170.

- [13] L. Kalaydjieva, B. Morar, Roma (Gypsies): Genetic studies, *Nature Encyclopedia of the Human Genome*. 5 (2003) 160-165.
- [14] N. Novokmet, Z. Pavčec, Genetic polymorphisms of 15 AmpFISTR identifier loci in Romani population from Northwestern Croatia, *Forensic Sci.Int.* 168 (2007) e43-e46.
- [15] M. Eaaswarkhanth, S. Roy, I. Haque, Allele frequency distribution for 15 autosomal STR loci in two Muslim populations of Tamilnadu, India, *Leg.Med.* 9 (2007) 332-335.
- [16] Tahera, Dawoodi Bohra, (2010).
- [17] 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 Science International: Genetics*. In Press, Corrected Proof (2010).
- [18] S. Krithika, R. Trivedi, V.K. Kashyap, T.S. Vasulu, Allele frequency distribution at 15 autosomal STR loci in Panggi, Komkar and Padam sub tribes of Adi, a Tibeto-Burman speaking population of Arunachal Pradesh, India, *Leg.Med.* 9 (2007) 210-217.
- [19] C. Barrot, C. Sánchez, M. Ortega, A. González-Martín, C. Brand-Casadevall, A. Gorostiza, et al., Characterisation of three Amerindian populations from Hidalgo State (Mexico) by 15 STR-PCR polymorphisms, *Int.J.Legal Med.* 119 (2005) 111-115.
- [20] R. Wang, L. Liu, J. Zhao, C. Peng, Genetic distribution on 15 STR loci from a population of Southern Liaoning in northeast of China, *Forensic Science International: Genetics*. 2 (2008) e25-e26.
- [21] S. Maruyama, K. Minaguchi, N. Takezaki, P. Nambiar, Population data on 15 STR loci using AmpF/STR Identifier kit in a Malay population living in and around Kuala Lumpur, Malaysia, *Leg.Med.* 10 (2008) 160-162.
- [22] F. Stanciu, O.R. Popescu, I.M. Stoian, STR data for the AmpFISTR Identifier from Dobruja region (SE Romania), *Forensic Science International: Genetics*. 3 (2009) 146-147.
- [23] Z. Szczerkowska, E. Kapińska, J. Wysocka, L. Cybulska, Northern Polish population data and forensic usefulness of 15 autosomal STR loci, *Forensic Sci.Int.* 144 (2004) 69-71.
- [24] V. Gomes, P. Sánchez-Diz, C. Alves, I. Gomes, A. Amorim, A. Carracedo, et al., Population data defined by 15 autosomal STR loci in Karamoja population (Uganda) using AmpF/STR Identifier kit, *Forensic Science International: Genetics*. 3 (2009) e55-e58.
- [25] A.C.d.S. Góes, D.A. da Silva, E.H. Fonseca Gil, M. da Silva, R. Pereira, E. de Carvalho, Allele frequencies data and statistic parameters for 16 STR loci—D19S433,

D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358—in the Rio de Janeiro population, Brazil, *Forensic Sci.Int.* 140 (2004) 131-132.

[26] L. Souto, C. Alves, L. Gusmão, E. Ferreira, A. Amorim, F. Côrte-Real, et al., Population data on 15 autosomal STRs in a sample from East Timor, *Forensic Sci.Int.* 155 (2005) 77-80.

[27] T. Muro, J. Fujihara, S. Imamura, H. Nakamura, T. Yasuda, H. Takeshita, Allele frequencies for 15 STR loci in Ovambo population using AmpFI STR® Identifiler Kit, *Leg.Med.* 10 (2008) 157-159.

[28] D. Havaš, N. Jeran, L. Efremovska, D. Đorđević, P. Rudan, Population genetics of 15 AmpFI STR Identifiler loci in Macedonians and Macedonian Romani (Gypsy), *Forensic Sci.Int.* 173 (2007) 220-224.

[29] 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, *Forensic Science International: Genetics Supplement Series.* 1 (2008) 306-308.

[30] A.O. Tillmar, G. Bäckström, K. Montelius, Genetic variation of 15 autosomal STR loci in a Somali population, *Forensic Science International: Genetics.* 4 (2009) e19-e20.

[31] E. Martinez-Espin, L.J. Martínez-Gonzalez, F. Fernandez-Rosado, C. Entrala, J.C. Alvarez, J.A. Lorente, et al., Guatemala mestizo population data on 15 STR loci (Identifiler® kit), *J.Forensic Sci.* 51 (2006) 1216-1218.

[32] L.J. Martínez-Gonzalez, E. Martinez-Espin, F. Fernandez-Rosado, M.A. Moguel, C. Entrala, J.C. Alvarez, et al., Mexican population data on fifteen STR loci (Identifiler® kit) in a Chihuahua (North Central Mexico) sample, *J.Forensic Sci.* 50 (2005) 236-238.

[33] E. Martinez-Espin, F. Fernández-Rosado, J.C. Alvarez, C. Entrala, J.A. Lorente, M.O. De Duarte, et al., Paraguayan population data on the fifteen STR loci included in the POWERPLEX 16™ kit, *J.Forensic Sci.* 48 (2003) 253-255.

[34] E. Melendez, E. Martinez-Espin, I.S. Karlson, J.A. Lorente, B. Budowle, Population Data on 15 STR Loci (PowerPlex 16™ Kit) in a Costa Rica (Central America) Sample Population, *J.Forensic Sci.* 49 (2004) 170-172.

[35] M.S. Figueiredo, F. Fernandez-Rosado, I. Kunii, A.C. Pacheco, J.A. Lorente, B. Budowle, Brazilian Caucasian Population Data for 15 STR Loci (PowerPlex 16™ Kit), *J.Forensic Sci.* 49 (2004) 167-169.

[36] R. Rubi-Castellanos, M. Anaya-Palafox, E. Mena-Rojas, D. Bautista-España, J.F. Muñoz-Valle, H. Rangel-Villalobos, Genetic data of 15 autosomal STRs (Identifiler kit) of three Mexican Mestizo population samples from the States of Jalisco (West), Puebla

(Center), and Yucatan (Southeast), *Forensic Science International: Genetics*. 3 (2009) e71-e76.

[37] C.J. Finis, Megaplex STR Analysis from a Single Amplification: Validation of the PowerPlex® 16 System, *Profiles in DNA*. (2001) 3.

[38] A. Tereba, Tools for Analysis of Population Statistics, *Profiles in DNA*. (2001).

[39] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: An integrated software package for population genetics data analysis, *Evolutionary Bioinformatics Online*. 1 (2005) 47.

Capítulo VI: Variabilidad Genética en la Población de Guatemala

*Chapter VI: Genetic Variability in
Guatemalan Population.*



ABSTRACT

This chapter summarizes the most relevant results of the different markers obtained in the previous chapters.

The appearance of the European and Native American genetic flows in the Mestizo population will be considered as far as the emergence of new haplogroups in the population structure. While DNAmT haplogroups are very similar between the Mayan and Mestizos; the Mestizo population has enriched the number of haplogroups for Y chromosome thanks to the presence of the Spanish population.

An asymmetric flow of the genetic contribution of Mayan and Spanish populations to the Mestizo population is shown. The influence has been higher in Y chromosome and lesser in DNAmT than in the rest of the genome because the source of variability was mainly male of the high proportion of male Spanish colonizers.

In the Mestizo population, there are some individuals that have a high percentage of native population instead of other individuals that present a low

proportion of those and high of the immigrant population. This fact is very important to further disease association studies.

To conclude, a Native American population with a high variability is a source of microevolution phenomena: gain of mutations, selection, genetic flows and genetic drift. Native American populations can be defined as a model to study all the microevolution phenomena.

1. INTRODUCCIÓN

Tras realizar el análisis de 400 muestras de Guatemala para 15 STRs autosómicos, de 225 muestras en 17 STRs de cromosoma Y y de 296 análisis para las regiones hipervariables HV1 y HV2 del ADN mitocondrial; se puede observar una gran variabilidad en los resultados obtenidos. La variabilidad se debe a la distinta herencia del conjunto de marcadores genéticos estudiados.

Como ya se describió en el primer capítulo, la herencia de los cromosomas autosómicos difiere de la herencia del ADNmt y del cromosoma Y, estudiar estos tres tipos de marcadores permite ahondar en el abordaje de la estructura global de la población de Guatemala.

En este capítulo, se muestran en síntesis los resultados más relevantes mostrados en los capítulos anteriores de cada uno de los distintos tipos de marcadores.

En primer lugar, se valorará la aparición del flujo genético europeo y nativo americano en la población mestiza y la aparición de nuevos haplogrupos en su estructura poblacional, enriqueciendo y aumentando la variabilidad con respecto a la población maya original. Analizando los haplogrupos, se comprobará la proporción de los nuevos haplogrupos en la población mestiza.

Se describirán las diferencias en la estructura de la población debidas a las diferencias en la herencia de cada uno de los marcadores estudiados. Y como será la composición genética de la población nativa resultante.

Este capítulo se centrará en la influencia europea dejando de lado otros grupos poblacionales, dado a que la influencia de grupos africanos o de otros continentes es mínima en esta población (confirmando los antecedentes descritos por historiadores y antropólogos).

Por último, se describirán los parámetros más importantes que producen la variabilidad encontrada en la población nativa americana.

2. DIFERENCIAS ENTRE LOS MARCADORES ESTUDIADOS

Para los tres tipos de marcadores (nuclear, ADNmt y cromosoma Y), dos de ellos uniparentales; la población mestiza parece siempre en posición intermedia (figuras II.6, III.10 y IV.24) pero con una mayor o menor proximidad a la población nativa americana según el tipo de marcador analizado.

Para poder realizar un estudio a nivel cuantitativo, esta proximidad se observa con la estimación de mestizaje y con la frecuencia de los haplogrupos. Los haplogrupos, como ya se explicó en el capítulo I, poseen una distribución geográfica y herencia específica, la del cromosoma Y por vía paterna y del ADNmt por vía materna.

Si se observan los haplogrupos encontrados en las poblaciones, mayas, mestiza y española, figura VI.1; se ven diferencias claras en la aportación de haplogrupos desde las poblaciones parentales a la población mestiza.

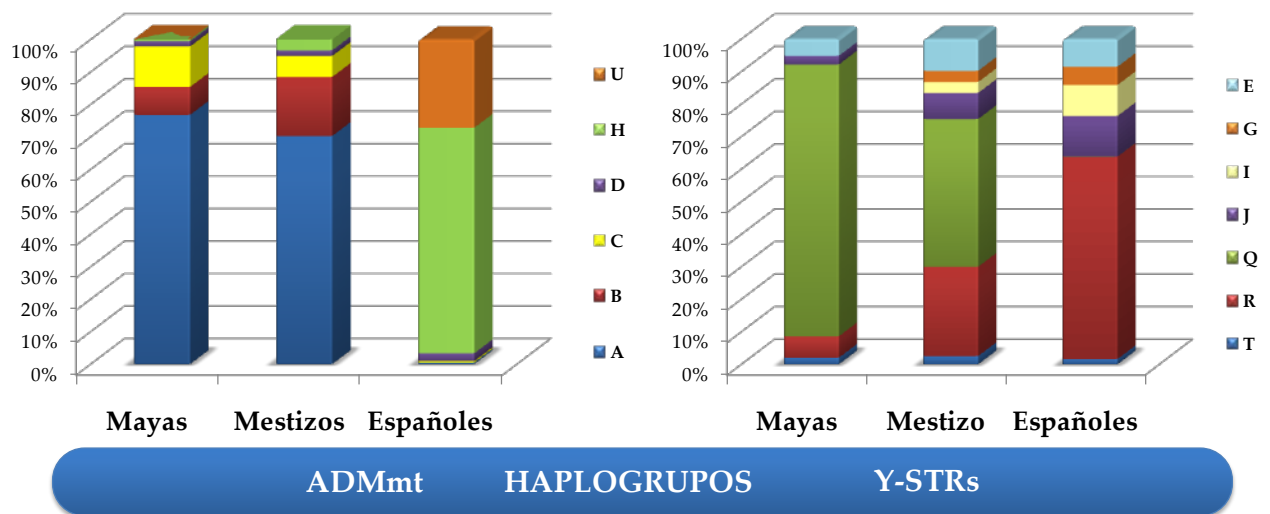


Figura VI.1: Histograma con las proporciones de haplogrupos encontrados en las poblaciones Maya y Mestiza de Guatemala y Española. A la izquierda haplogrupos determinados a partir del ADNmt, a la derecha, los haplogrupos determinados a partir de los Y-STRs.

En esta figura (fig. VI.1), se puede ver que los haplogrupos en ADNmt de la población maya y mestiza son prácticamente iguales, con una pequeña variación en la proporción de B y C (ambos haplogrupos típicos americanos),

mientras que no se encuentra un número significativo de haplogrupos H o R en la población mestiza, típicos de la población española y europeas. En cambio al determinar los haplogrupos de la población mestiza, a partir de los marcadores en cromosoma Y, se puede ver como en ésta se observa un porcentaje bastante alto de individuos con haplogrupo R y también, aunque en menor proporción; aparecen individuos con haplogrupo G e I y aumentan las proporciones del J y E. En general, en cromosoma Y se ha enriquecido el número de haplogrupos gracias a la presencia de la población española.

Si se estudian, de forma conjunta los resultados obtenidos al calcular la proporción del mestizaje calculada en los capítulos anteriores, tablas II.12 (pg.124), III.10 (pg.172) y IV.11 (pg.228), se puede observar claramente como existe un flujo asimétrico genético desde la población europea hasta la población mestiza de Guatemala.

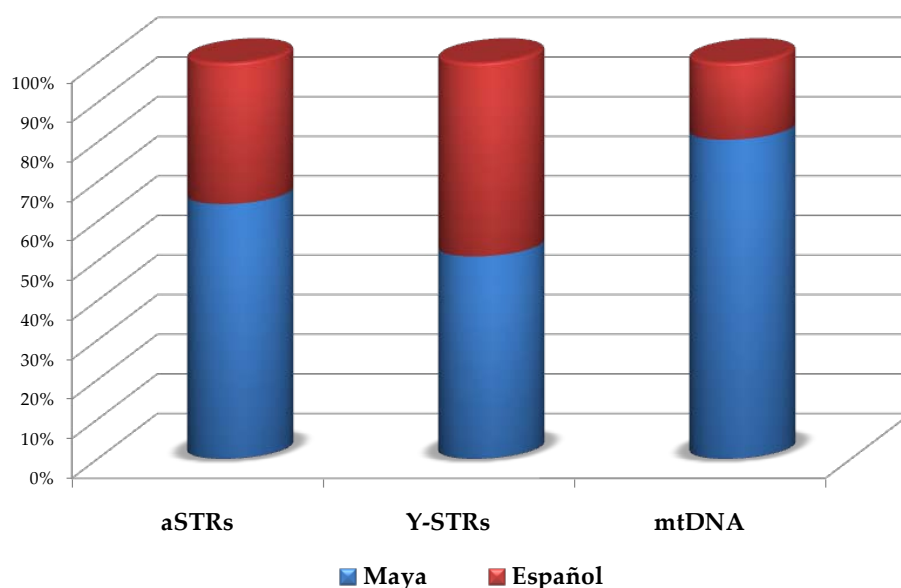


Figura VI.2: Histograma con la media de las proporciones de poblaciones parentales obtenidas por los programas Admix Pas y Admix 95 para los marcadores autosómicos, de cromosoma Y y las secuencias de ADNmt.

Como se muestra en la figura VI.2 las proporciones de marcadores genéticos parentales (población maya y española) aportadas a la población mestiza (a partir de las frecuencias de los marcadores de a-STRs, Y-STRs y ADNmt) son

muy diferentes. La población maya aporta a la mestiza un 64% (a-STRs), un 51% (Y-STRs) y un 80% (ADNmt). La población española, en cambio, aporta un 35% (a-STRs), 48% (Y-STRs) y 19% (ADNmt) a la población mestiza. Es mucho mayor la aportación de marcadores de cromosoma Y, que la que se calcula para ADNmt. Como se podía esperar, el porcentaje en a-STRs tiene una posición intermedia.

En la población de Guatemala, se puede ver que la población mestiza conserva aún gran cercanía a la dotación originaria de las comunidades maya. Aunque es difícil hablar de cifras exactas en estos casos, se podría decir que la población mestiza conserva un 60-65% del pool inicial maya y éste se ha enriquecido con la dotación genética sur europea. Si se comparan todos los resultados obtenidos para cada uno de los conjuntos de marcadores se ve que donde se encuentra mayor aporte español es en el conjunto de marcadores de cromosoma Y, y una menor distancia genética. En cambio un menor aporte y por tanto mayor distancia, se puede ver en las regiones estudiadas de ADN mitocondrial. En los haplogrupos mestizos de cromosoma Y aparecen muchos de origen europeo.

Todas las diferencias encontradas se pueden explicar debido a que los colonizadores españoles fueron varones en su mayoría. Al ser la fuente de variabilidad masculina, la influencia ha sido mayor en cromosoma Y y menor en ADN mitocondrial que en el resto del genoma.

Éste no es un caso aislado, algo similar ocurrió en Colombia con los flujos genéticos africanos y europeos [1,2], donde se pudo demostrar un flujo asimétrico de la contribución genética desde estas poblaciones africanas y europeas hasta la población mestiza de Colombia. En este caso, al igual que en este estudio; la presencia de colonizadores europeos y de esclavos de origen africano, en su mayoría varones, influyen en la dotación genética actual de la población mestiza. Como se puede observar en la población actual de Guatemala.

2.1 DISTRIBUCIÓN DE LA VARIABILIDAD EN LA POBLACIÓN MESTIZA DE GUATEMALA Y SU IMPORTANCIA EN LOS ESTUDIOS DE ASOCIACIÓN

Cuando se aborda el estudio de una población es muy importante conocer la homogeneidad de ésta. Si se desea realizar un estudio de asociación de un carácter genético (SNPs, STRs o indels) a una población, en primer lugar se tiene que conocer como es la estructura de la misma, si está en equilibrio, si es homogénea, etc. Estos parámetros son muy importantes en los estudios poblacionales porque, gracias a ellos, posteriormente se podrán realizar estudios de asociación a marcadores predictivos de enfermedades o eficacia de fármacos asociados a genotipos (farmacogenómica).

Desde 1986 se comienza a describir la dificultad de estudiar la asociación de enfermedades con base genética en las poblaciones mestizas. La subestructura de las poblaciones produce una estratificación de los resultados al estudiar marcadores genéticos o intentar predecir la asociación de caracteres a estos individuos, lo que hace más difícil poder encontrar una clara asociación de marcadores en este tipo de poblaciones [3]. En las poblaciones mestizas existen individuos que poseen un tanto por ciento muy alto de población nativa y otros, en cambio; que poseen un pequeño porcentaje de éstas y tienen una alta proporción de la población inmigrante (se encuentra en los mestizo un rango que puede ir del 0 al 100%).

Si se observan los resultados de nuestro estudio (con tres tipos de herencia), y como esta herencia ha condicionado a la población Mestiza de Guatemala se observa la aparición de genotipos con distinta gradación dentro del pool genético nativo y español. Por lo tanto, se asume que debe ocurrir lo mismo con el resto de marcadores que no estén asociados a otras fuentes de variación.

En la figura VI.3 se ve la disposición de nuestros individuos, al intentar clasificarlos en tres grupos (*cluster* en la figura) por afinidad genética a los componentes genéticos ancestrales calculados (el programa, Structure v2.3.1

genera el agrupamiento a partir de las distancias a cada una de las 3 componente K, estudiadas en las poblaciones). Los resultados obtenidos en a-STRs, con herencia mendeliana; es la gradación de la que se hablaba anteriormente, deben usarse como patrón, en todos los genes que posean la herencia mendeliana.

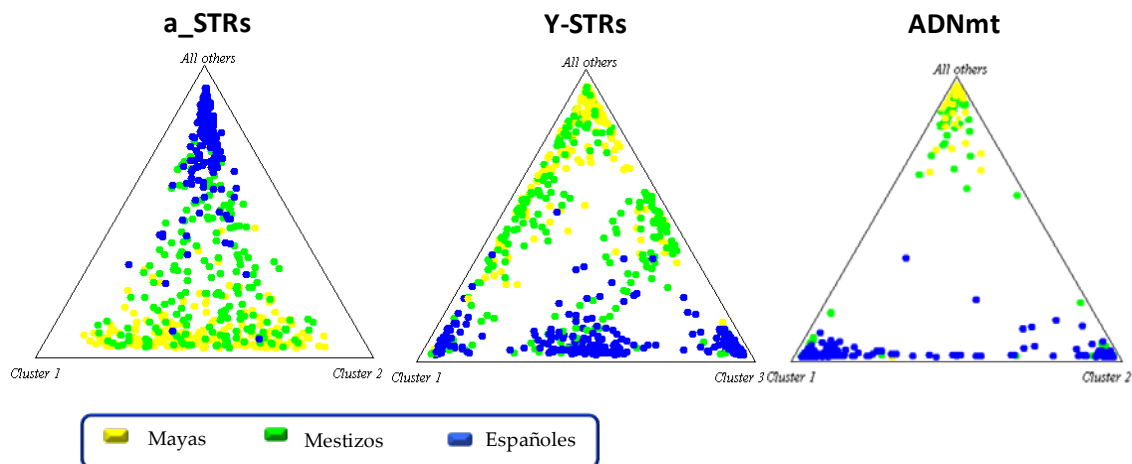


Figura VI.3 Gráfico triangular producido por el programa Structure v2.3.1 para la agrupación de los individuos de una población en tres grupos (1, 2 y 3) por homología genética a cada uno de los grupos.

En la herencia de cromosoma Y en el caso de la población mestiza, segundo triángulo en la fig. VI.6, se ve que hay individuos muy similares a los españoles, otros similares a individuos mayas, mientras que otros se agrupan entre ambos, pero no tienen la dispersión gradual que se encuentra en los marcadores autosómicos. En el ADNmt el mestizaje es mucho menor, la mayoría de los individuos mestizos se solapan con los mayas.

En principio, una población como la de Guatemala, como muchas otras poblaciones mestizas, puede suponer un problema a la hora de realizar estudios de asociación, pero si se conocen estos datos a priori existen herramientas para poder estudiar la variabilidad y clasificar los individuos en tantas subpoblaciones como se puedan definir [4].

Cuando se aborda un estudio de asociación, se puede realizar de dos formas: buscando posiciones variables en marcadores en genes implicados en el desarrollo de una enfermedad, en rutas metabólicas asociadas a estos o en

zonas promotoras, etc., o realizar estudios globales WGs (Estudios de asociación de genoma completo, del inglés, Whole Genome Association Studies). Ambas metodologías van a depender en gran medida de la estructura de la población [5], porque para su análisis es necesario conocer como se distribuirían los marcadores en ausencia de cualquier fuente de variación debida a la selección.

Cuando se estudian marcadores neutros, como los de este trabajo; aunque existen dificultades con cuestiones tales como la imputación y la mezcla individual, la investigación de las diversas poblaciones tiene un potencial significativo para poder avanzar en el proyecto de mapeo de los determinantes genéticos de las enfermedades complejas para la población humana [6]. Estudios como el presente, en marcadores no codificantes, ayudan a poder abordar posteriormente estudios de asociación sabiendo cómo se debe rastrear los caracteres y como debe ser su herencia si ésta no ha estado ligada a ningún carácter selectivo.

3. VARIABILIDAD EN POBLACIONES NATIVAS AMERICANAS

El continente americano, fue el último en ser habitado por el ser humano. Las migraciones, seguramente desde el estrecho de Bering, iniciaron esta colonización. Ya fuese en una o en varias migraciones, en origen la población americana poseía una pequeña variabilidad [7,8] (se calcula que la población ancestral de Beringia sólo contribuyó con seis haplotipos de ADNmt fundadores [9]). Aunque en un origen la población americana sólo poseía un número limitado de secuencias del genoma mitocondrial, hoy en día se han encontrado gran número de sub-haplogrupos.

Si el tamaño estimado de la población fundadora fue menor a 80 personas y, en cambio, ahora existe una gran variabilidad en pueblos nativos, se puede afirmar que han ocurrido una gran cantidad de acontecimientos para poder explicar la divergencia evidente en los pueblos del nuevo mundo [10].

Si se estudian las poblaciones individualmente, se puede ver que algunas son muy poco variables, Huetar [11], Ngöbe [12], Cayapas [13]; en cambio cuando se estudian las diferencias entre ellas, en conjunto existe una gran variabilidad como se ve en los resultados obtenidos en los capítulos anteriores, figura II.9, página 121; en la figura III.10, página 171 y figura IV.23, página 224.

La población nativa americana presenta una gran variabilidad. Ésta es producto de los fenómenos de microevolución: acúmulo de mutaciones fijadas (tasa de evolución), selección natural, flujos genéticos y deriva genética. Se pueden, por tanto, definir las poblaciones nativas americanas como un modelo para estudiar todos los fenómenos de microevolución.

4. BIBLIOGRAFÍA

- [1] L.G. Carvajal-Carmona, I.D. Soto, N. Pineda, D. Ortiz-Barrientos, C. Duque, J. Ospina-Duque, et al., Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in Northwest Colombia, *Am.J.Hum.Genet.* 67 (2000) 1287-1295.
- [2] C. Rodas, N. Gelvez, G. Keyeux, Mitochondrial DNA studies show asymmetrical Amerindian admixture in Afro-Colombian and Mestizo populations, *Human Biology.* 75 (2003) 13-30.
- [3] R. Chakraborty, Gene admixture in human populations: models and predictions, *Am.J.Phys.Anthropol.* 29 (1986) 1-43.
- [4] B. Devlin, K. Roeder, S.A. Bacanu, Unbiased methods for population association studies, *Genet.Epidemiol.* 21 (2001) 273-284.
- [5] M.D. Shriver, G.C. Kennedy, E.J. Parra, H.A. Lawson, V. Sonpar, J. Huang, et al., The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs, *Human genomics.* 1 (2004) 274-286.
- [6] N.A. Rosenberg, L. Huang, E.M. Jewett, Z.A. Szpiech, I. Jankovic, M. Boehnke, Genome-wide association studies in diverse populations, *Nature reviews.Genetics.* 11 (2010) 356.
- [7] T.G. Schurr, S.T. Sherry, Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: Evolutionary and demographic evidence, *Am.J.Hum.Biol.* 16 (2004) 420-439.
- [8] S. Wang, C.M. Lewis Jr, M. Jakobsson, S. Ramachandran, N. Ray, G. Bedoya, et al., Genetic variation and population structure in Native Americans, *PLoS Genetics.* 3 (2007) e185.
- [9] O.A. Derbeneva, R.I. Sukernik, N.V. Volodko, S.H. Hosseini, M.T. Lott, D.C. Wallace, Analysis of mitochondrial DNA diversity in the Aleuts of the Commander Islands and its implications for the genetic history of Beringia, *The American Journal of Human Genetics.* 71 (2002) 415-421.
- [10] J. Hey, On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS biology.* 3 (2005).
- [11] M. Santos, R.H. Ward, R. Barrantes, mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica, *Hum.Biol.* 66 (1994) 963-977.

[12] C.J. Kolman, E. Bermingham, R. Cooke, R. Ward, T.D. Arias, F. Guionneau-Sinclair, Reduced mtDNA diversity in the Ngobe Amerinds of Panama, *Genetics*. 140 (1995) 275.

[13] O. Rickards, C. Martínez-Labarga, J.K. Lum, G.F. De Stefano, R.L. Cann, mtDNA history of the Cayapa Amerinds of Ecuador: Detection of additional founding lineages for the native American populations, *Am.J.Hum.Genet.* 65 (1999) 519-530.

ABREVIATURAS

A	Adenina
a.c.	Antes de Cristo
AC	Análisis de correspondencias
ADN	Acido desoxirribonucleico
aDNA	Ancient DNA (ADN antiguo)
ADNmt	ADN mitocondrial
AMOVA	Análisis molecular de la Varianza
ARN	Ácido ribonucleico
aSTR	Repetición de nucleotidos situada en cromosomas autosomicos
ATP	Adenosin trifosfato
BDCIG	Bases de datos de criminales para identificación genética
C	Citosina
CODIS	Combined DNA Index System (Sistema Combinado de Índice de ADN)
d.c.	Después de Cristo
FBI	Federal Bureau of Investigation (Buró Federal de Investigación de los Estados Unidos)
FTA	Soporte patentado por Whatman para la conservación de muestras de ADN
G	Guanina
Het	Heterocigosidad observada
Hj:	Heterocigosidad esperada en un locus j
HLA	Human leukocyte antigen (Antígenos leucocitarios humanos)
HV1	Región hipervariable 1 del ADNmt

HV2	Región hipervariable 2 del ADNmt
HV3	Región hipervariable 3 del ADNmt
HW	Hardy-Weinberg, Ley del equilibrio Hardy-Weinberg
IP	Índice de paternidad
ISFG	International Society for Forensic Genetics (Sociedad Internacional de Genética Forense)
Kb	Kilobases
LINES	Long Interspersed Nuclear Elements (Elementos nucleares disperses largos)
LR	Likelihood Ratio (Razón de verosimilitud)
Mb	Megabases
MDS	Multidimensional Scaling (Escalamiento multidimensional)
NJ	Algoritmo Neighbor-joining
OXPHOS	Oxidative Phosphorylation (Fosforilación oxidativa)
P	Probabilidad de paternidad
Pb	Pares de bases
PCR	Polymerase Chain Reaction (Reacción en cadena de la polimerasa)
PD	Poder de discriminación
PE	Poder de exclusión
PG	Polimorfismos genéticos
PGM	Proyecto Genoma Humano
PIC	Polymorphism information content (Contenido en información polimórfica)
P_{ij}	Frecuencia del alelo i del locus j
RFLP	Restriction Fragment Length Polymorphism (Polimorfismos de longitud para fragmentos de restricción)
SINEs	Short Interspersed Nuclear Elements (Elementos nucleares dispersos cortos)

SNPs	Single Nucleotide Polymorphisms (Polimorfismos de un único nucleótido)
SPSS	Statistical Package for the Social Sciences (Paquete estadístico , software)
STR	Short Tandem Repeat (Repetición de nucleótidos en tándem)
STRs	Plural de STR
T	Timina
UNICEF	United Nations Children's Fund (Fondo de las Naciones Unidas para la Infancia)
UPGMA	Unweighted pair group method using arithmetic averages (Método de grupo par no ponderado usando promedios aritméticos)
VNTR	Variable Number of Tandem Repeat (Repeticiones en tándem en número variable)
VR1	Región variable 1
VR2	Región variable 2
X-STRs	Short tandem repeat (STR) en el cromosoma X
Y-STRs	Short tandem repeat (STR) en el cromosoma Y
sd	Desviación estándar
μg	Microgramo
μl	Microlitro
μM	Micromolar
WLS	Weighted least squares (minimos cuadrados con errores estandar)
rCRS	Revised Cambridge Reference Sequence (Secuencia de referencia revisada de Cambridge)
DTR	Dye Terminator Removal (Eliminación de terminadores de fluorocromos de secuenciación)

Anexos / Appendices

Publicaciones realizadas hasta la fecha derivadas de esta tesis doctoral:

Población de Guatemala:

1. Distribution of Y chromosomal STRs loci in Mayan and Mestizo populations from Guatemala. **Martínez-González LJ**, Saiz M, Alvarez-Cubero MJ, Gómez-Martín A, Alvarez JC, Martínez-Labarga C, Lorente JA. *Forensic Sci Int Genet*. 2011 May.
2. Guatemala mestizo population data on 15 STR loci (Identifiler Kit). Martinez-Espin E, **Martínez-Gonzalez LJ**, Fernandez-Rosado F, Entrala C, Alvarez JC, Lorente JA, Budowle B, Ovalle de Monroy M. *J Forensic Sci*. 2006 Sep.

Otras poblaciones, sudamericanas y española, empleadas en los análisis de distancias:

1. Genetic variation of 15 autosomal microsatellite loci in a Nayarit population (Mexico). Alvarez-Cubero MJ, **Martinez-Gonzalez LJ**, Virgen-Ponce D, Delgado-Najar EA, Moscoso-Caloca GH, Alvarez JC, Lorente JA. *Leg Med (Tokyo)*. 2011 Sep
2. Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. Alvarez JC, Johnson DL, Lorente JA, Martinez-Espin E, **Martínez-Gonzalez LJ**, Allard M, Wilson MR, Budowle B. *Leg Med (Tokyo)*. 2007 Nov
3. Mexican population data on fifteen STR loci (Identifiler kit) in a Chihuahua (North Central Mexico) sample. **Martínez-Gonzalez LJ**, Martinez-Espin E, Fernandez-Rosado F, Moguel MA, Entrala C, Alvarez JC, Lorente JA, Budowle B. *J Forensic Sci*. 2005 Jan.



Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



Short communication

Distribution of Y chromosomal STRs loci in Mayan and Mestizo populations from Guatemala

Luis J. Martínez-González^{a,*}, María Saiz^a, María J. Álvarez-Cubero^a, Antonio Gómez-Martín^b,
Juan C. Álvarez^{a,d}, Cristina Martínez-Labarga^c, José A. Lorente^{a,d}

^aLaboratory of Genetic Identification, Department of Legal Medicine, Universidad de Granada, Granada, Spain

^bSavia Biotech S. A., Avda. Sudamérica, 94 04740 Roquetas de Mar, Almería, Spain

^cCentre of Molecular Anthropology for Ancient DNA Studies, Dipartimento di Biologia, Università di Roma "Tor Vergata", Rome, Italy

^dPfizer – Universidad de Granada – Junta de Andalucía Centre for Genomics and Oncological GENYO, Centro de Investigación Biomédica, Avda. del Conocimiento s/n, 18100, Armilla, Granada, Spain

ARTICLE INFO

Article history:

Received 15 September 2010

Received in revised form 18 February 2011

Accepted 2 April 2011

Keywords:

Genetic relationship

Admixture

Population genetics

Forensic science

DNA typing

ABSTRACT

In this study, a sample of 225 Guatemalan males, comprising 115 Mestizo-Guatemalan and 110 Mayan-Guatemalan, was typed for 17 Y-short tandem repeats (STRs) loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, YGATA_H4.1 and DYS385a/b). The haplotype diversity ($H = 1$) and discrimination capacity (96.86%) were calculated. Analysis of molecular variance (AMOVA) demonstrated a low but significant interpopulation differentiation when compared with the results obtained when we confront the Mestizo and Mayan populations with the European populations.

Furthermore, the genetic variability and differences among the American, African, Asian, and European populations were analyzed with the software Statistica 9.1. In addition, the genetic distances were also calculated using other published data. Reynolds and Slatkiñs genetic distance was visualized using the multidimensional scaling (MDS) analysis. All the analysis performed locates the Mayan population next to the Native American population, while Guatemalan-Mestizo population was found to be between these populations and the European population, similar to other Mestizo one.

The implementation of the estimation of individual ancestry proportions of the whole population sample showed the presence of two well-differentiated population groups.

© 2011 Elsevier Ireland Ltd. All rights reserved.

1. Population

A total of 225 DNA samples was obtained from unrelated male donors from both the populations studied, comprising 115 Guatemalan-Mestizo inhabitants and 110 Guatemalan-Mayan individuals [Kaqchiquel (22), K'iché (37), Mam (27), and Q'eqchi (24)] (see Fig. 1). Samples from Mayan individuals were obtained directly at their communities. About 25 μ l of blood samples were spotted on FTA[®] Whatman paper under informed consent.

The population of Guatemala includes 12,727,566 individuals [1] and according to INE Guatemala, 48.92% of them are men [2]. Nowadays, the population of Guatemala can be divided into two principal groups: Native-American groups (40.1%) and the Spanish-speaking population (Mestizo and European descendents) (59.4%) [1]. The main Mayan ethnic groups living in the country are K'iche' (9.1%), Kaqchikel (8.4%), Q'eqchi' (6.3%), and Mam (7.9%) [1].

The great majority of Guatemalan-Mestizos are the result of admixture between these Mayan ethnic groups and Spaniards.

In order to perform the analysis of molecular variance (AMOVA), it was included a sample of 175 Spanish individuals previously analyzed.

2. Extraction

DNA extract was prepared for PCR analysis using standard FTA[®] protocols (Whatman, Clifton, NJ, USA) and 1.2 mm punch of FTA cards was used in each amplification.

3. PCR amplification

The samples were amplified using AmpFISTR[®] Yfiler[®] kit including 17 loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, YGATA_H4.1 and DYS385a/b) (Applied Biosystems, Foster City, CA) [3].

* Corresponding author. Tel.: +34 958249950; fax: +34 958246107.

E-mail address: luisjaviermg@ugr.es (L.J. Martínez-González).



Fig. 1. Mayan and Mestizo distribution in Guatemala.

4. Typing

Alleles were separated and detected using an Applied Biosystems ABI 310 genetic analyzer. Fragment sizes were analyzed using the GeneMapper ID-X v.1.1 software (Applied Biosystems, Foster City, CA). The alleles were named according to the number of repeated units based on the sequenced allelic ladder (ISFG recommendations) [4].

5. Quality control

The regular participation of the lab in the proficiency testing of the GEP-ISFG Working Group (<http://www.gep-isfg.org>) and Proficiency testing of the GITAD (<http://gitad.ugr.es/principal.htm>) is certified.

The Y-STR Haplotype Reference Database [5] accession number for the population sample “Guatemala [Mestizo]” $n = 115$ is YA003664; and the YHRD accession number for Guatemala, Mayan, $n = 110$ are Chimaltenango, Guatemala [Maya, Kaqchiquel] YA003663, Santa Cruz del Quiché, Guatemala [Maya, K'iché] YA003666, Alta Verapaz, Baja Verapaz, Petén e Izabal, Guatemala [Maya, Qeqch'i] YA003662 and Northeast, Guatemala [Maya, Mam] YA003665 (<http://www.yhrd.org/>).

6. Analysis of data

Haplotype frequencies were calculated by the gene counting method. Haplotype diversities were calculated according to Nei [6,7] using Arlequin v3.0 software [6]. Previously published haplotypes and frequencies data from several populations were compiled, which comprised South American (Toba, Argentina [8]; El Beni, Bolivia [9]; Rio de Janeiro, Brazil [10]; Rio Grande do Sul, Brazil [11]; Mestizo, Ecuador [12]; and Kichwa, Ecuador [12]), North American (African descendent; European descendent; Hispanic descendent) [13], African (Algeria [14]; Equatorial Guinea [15]; Ovambo, Namibia [16]), Asian (Japan [17]; Korea [18]; Chinese Tibetan [19]; Chaoshan, China [20]), and European (Spain [21]; Portugal [22]; Italy [23]; Croatia [24]; Serbia [25]; and Austria [26]). To study the ancestry relationships, six markers (DYS385a/b, DYS439, DYS456, DYS458 and DYS635) with the

highest mutation rates were eliminated [27]. A correspondence analysis was calculated with Statistica 9.1 (Statsoft Inc. Tulsa USA) to see the association between allele frequencies and populations. Arlequin v3.0 software was used to calculate the distributions of the observed allele frequencies in each group, F-statistics and Reynolds and Slatkińs genetic distances. Arlequin v3.0 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 [28]. To represent the distances in a more appropriate way, genetic distances (Reynolds and Slatkińs genetic-distance matrix) between the American and European populations were graphically summarized by nonmetric multidimensional scaling (NM-MDS) [29], using the software SPSS v.15.0. NM-MDS uses an iterative process to transform a similarity/dissimilarity matrix into distances represented in a Euclidean n -dimensional space. Four markers were eliminated owing to their high mutation rates (DYS439, DYS458, DYS456 and DYS635) [27].

Furthermore, the admixture proportions of the Mayan and South West European populations (90% Spanish and 10% Portugal data) in Guatemalan-Mestizo population were estimated by means of the weighted least squares (WLS), ADMIX.PAS program [30] and ADMIX95 program [31]. We implemented the estimation of individual ancestry proportions with the program STRUCTURE 2.3.1 [32–34] <http://pritch.bsd.uchicago.edu/structure.html>. Replicate runs of STRUCTURE using different burn-in period and interactions were obtained, and here, we have reported the estimations with a burn-in period of 50,000 interactions followed by an additional 100,000 interactions ($K = 1, 2$ and 3) with all the populations. We used two models of the available options, namely admixture and no admixture models.

7. Results

A total of 225 samples from two Guatemalan populations, Mayan and Mestizo, were investigated in this study. All the 225 haplotypes obtained were different, and therefore, in our case, the haplotype diversity was 1.0. See electronic supplementary material, Tables S1 and S2 for haplotype distribution of the Y-chromosome obtained among the 225 individuals analyzed. The distributions of the observed allele frequencies in Mestizo and Mayan populations from Guatemala for 17 short tandem repeats (STRs) loci studied are shown in Tables 1 and 2. These tables show us the power of discrimination (PD) of the markers studied. The most informative marker was found to be DYS458 in both the populations. The less descriptive locus was observed to be DYS393 in the Guatemalan-Mestizo population and DYS437 in the Guatemalan-Mayan population. The combined PD in both the populations was higher than 0.999999.

8. Other remarks

Population genetic variation studies have demonstrated that there is an overall low level of differentiation in human populations [35]; however, local factors, such as geography and differential settlement can greatly enhance genetic discontinuity.

In this study, we have carried out an extensive analysis of Y-chromosomal diversity in Guatemalan populations, examining 17 STRs. These loci have allowed us to study the level of variability and substructure of the actual populations of Guatemala. AMOVA revealed that the highest variation is mainly within the populations. Both the Guatemalan populations displayed a low but significant interpopulation differentiation when compared with the results obtained when we confronted the Mestizo and Mayan populations with the European populations. The results are presented in Table 3.

Table 1
Y-STR haplotypes and locus diversity from 115 Mestizo individuals from Guatemala.

Allele	DYS456	DYS 389 I	DYS 390	DYS 389 II	DYS 458	DYS 19	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	Y GATA H4	DYS 437	DYS 438	DYS	Allele	DYS 385 A/B
															448		
8														0.00870		10.14	0.00870
9								0.04348						0.10435		11.11	0.00870
10								0.66087	0.02609			0.00870	0.03478	0.21739		11.12	0.00870
11							0.00870	0.29565	0.45217			0.23478	0.30435	0.39130		11.13.2	0.00870
11.2																11.14	0.14783
12		0.19130				0.02609	0.15652		0.38261			0.60000		0.27826		11.15	0.04348
13	0.02609	0.48696				0.40000	0.73913		0.12174			0.33913	0.06087	0.00870		11.16	0.01739
13.2																11.18	0.00870
<14					0.00870											11.2.12	0.00870
14	0.06957	0.31304			0.02609	0.40870	0.08696		0.00870		0.31304		0.65217			12.14	0.03478
15	0.50435	0.00870			0.21739	0.11304	0.00870		0.00870		0.08696		0.23478			12.16	0.01739
16	0.29565				0.22609	0.03478					0.01739		0.10435			13.14	0.02609
17	0.06957				0.31304	0.01739										13.15	0.03478
18	0.02609				0.15652										0.095652	13.16	0.02609
18.2					0.00870											13.17	0.02609
>18	0.00870															13.18	0.02609
19					0.04348					0.00870						0.417391	13.20
20										0.00870						0.313043	13.2.18
21			0.01739													0.130435	14.14
22			0.05217													0.034783	14.15
23			0.26957													0.008696	14.16
24			0.47826														14.17
25			0.16522														14.18
26			0.00870														14.19
27			0.00870														14.20
28				0.10435													15.15
29				0.35652													15.16
30				0.37391													15.17
31				0.12174													15.18
32				0.02609													16.16
33				0.01739													16.17
PD	0.64711	0.62820	0.66813	0.70639	0.77641	0.65800	0.42147	0.47395	0.63350	0.70261	0.72393	0.54246	0.50858	0.71123	0.70034		0.93899

Abbreviations: PD – power of discrimination.

Table 2

Y-STR haplotypes and locus diversity from 110 Mayan individuals from Guatemala.

Allele	DYS456	DYS 389 I	DYS 390	DYS 389 II	DYS 458	DYS 19	DYS 393	DYS 391	DYS 439	DYS 635	DYS 392	Y GATA H4	DYS 437	DYS 438	DYS 448	Allele	DYS 385 A/B
9								0.06364						0.02727		11.13	0.01818
10		0.01818						0.77273	0.03636		0.00909			0.09091		11.14	0.02727
11		0.00909					0.01818	0.16364	0.41818		0.05455	0.15455		0.76364		12.12	0.00909
12		0.02727				0.00909	0.15455		0.39091			0.70909		0.10000		12.16	0.00909
13		0.70909				0.59091	0.718182		0.154545		0.136364	0.1181818	0.00909	0.018182		13.14	0.00909
13.2						0.02727										13.15	0.05455
>13												0.01818				13.16	0.00909
<14					0.00909											13.17	0.05455
14	0.08182	0.20909			0.00909	0.3	0.10000				0.66364		0.854545			13.18	0.01818
15	0.53636	0.01818			0.17273	0.05455	0.00909				0.11818		0.11818			13.19	0.01818
16	0.22727				0.34545	0.01818					0.01818		0.01818			14.14	0.03636
17	0.11818	0.00909			0.28182											14.15	0.00909
17.2					0.00909											14.16	0.14545
18	0.03636				0.11818										0.072727	14.17	0.14545
19					0.03636										0.581818	14.18	0.13636
2					0.018182										0.318182	14.19	0.00909
21										0.04545					0.027273	14.20	0.00909
21.3										0.00909						15.15	0.04545
22			0.02727							0.65455						15.16	0.06364
23			0.21818							0.23636						15.17	0.04545
24			0.60909							0.04545						15.18	0.04545
25			0.10909							0.00909						15.19	0.00909
26			0.02727	0.00909												16.16	0.01818
27			0.00909	0.00909												16.17	0.02727
28				0.02727												16.18	0.01818
29				0.27273												17.17	0.00909
30				0.34545													
31				0.22727													
32				0.09091													
33				0.01818													
PD	0.63868	0.45190	0.56793	0.74512	0.75554	0.55669	0.44992	0.37207	0.64711	0.51140	0.52364	0.45901	0.25537	0.39752	0.554215		0.91769

Abbreviations: PD – power of discrimination.

Table 3
AMOVA design and results from 110 Mayan, 115 Mestizo and 175 Spanish males.

AMOVA design and results:					
Structure desing	Source of variation:	d.f.	Sum of squares	Variance components	Percentage of variation
2 Groups:	Among groups	1	115,552	0.45497 Va	9.43%
Group 1 = Mestizo and Mayan	Among populations within groups	1	27,634	0.20873 Vb	4.32%
Group 2 = Spanish	Within populations	375	1,561,163	4.16310 Vc	86.25%
	Total	377	1,704,349	4.82681	Fixation Indices
					FSC: 0.04775
					FST: 0.13750
					FCT: 0.09426
2 Groups:	Among groups	1	96,284	0.34484 Va	7.13%
Group 1 = Mestizo and Spanish	Among populations within groups	1	46,902	0.32549 Vb	6.73%
Group 2 = Mayan	Within populations	375	1,561,163	4.16310 Vc	86.13%
	Total	377	1,704,349	4.83344	Fixation indices
					FSC: 0.07252
					FST: 0.13869
					FCT: 0.07135

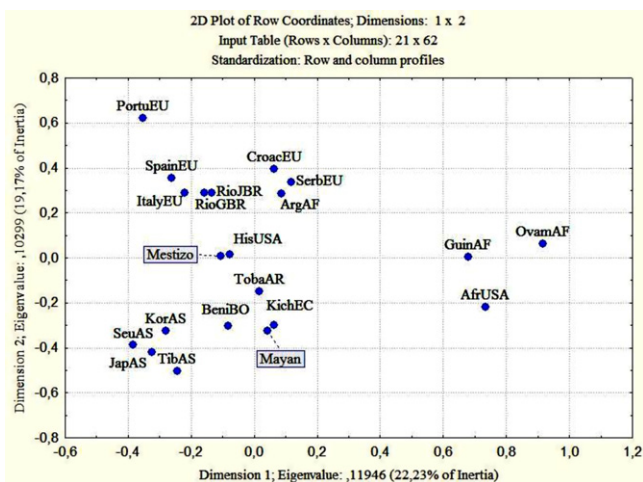


Fig. 2. Two-dimensional plot of correspondence analysis (for the interpretation of the abbreviation used in this figure, refer to Table S1).

A correspondence analysis with the Asian, European and African populations was performed with Statistica 9.1 (Fig. 2). To make the interpretation of the data easier, this plot only contains results from the population and data from each marker are omitted. Our Mayan sample has been located with all the rest of the Native American populations that are found between Asian and Mestizo American population. All the Mestizo populations are found to be located at the middle of the image, between the Native American populations and the European colonizers. Furthermore, Europeans are observed to be situated at the top left part and at the center of this representation, owing to the differences found in *DYS438* (alleles 9 and 12), and *DYS437* (allele 16). This observation correlates with the genetic homogeneity between these populations. Although it does not correlate with the geographical distribution, it correlates with the history of these population groups. It can be noted that all Native South American populations are situated at the bottom right part of the figure. Next to them, the two Asian populations studied (Japan and Korea) can be observed at the left bottom part of the image. This position is because of the genetic markers located further from the center of the representation: *DYS389I* (allele 12), and *DYS 389II* (alleles 27 and 28). This

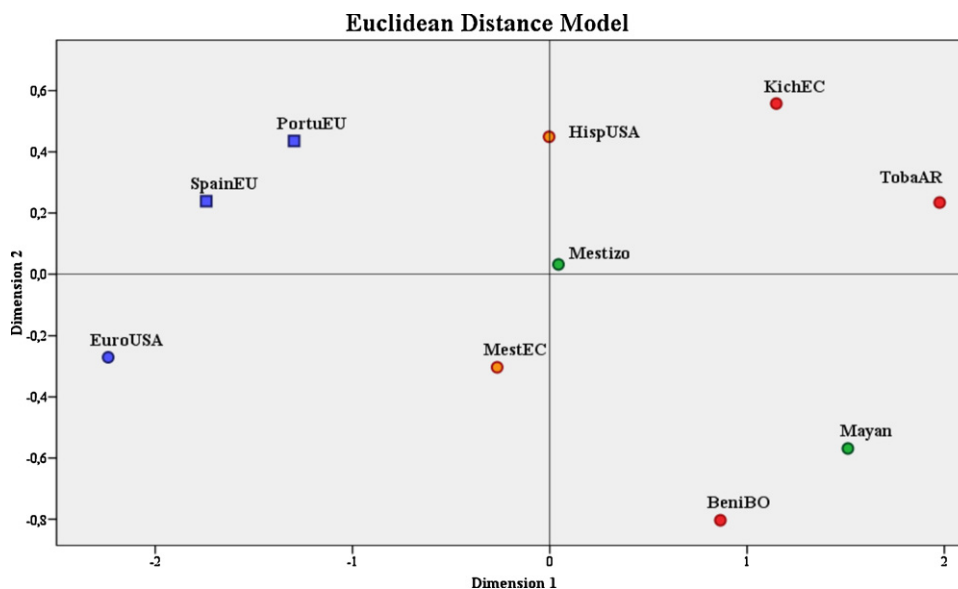


Fig. 3. Multidimensional scaling plot of haplogroup Fst Reynold's distances showing the admixture effect; stress value 0.14073 (for interpretation of the abbreviation in this Fig. the reader is referred to Table S1).

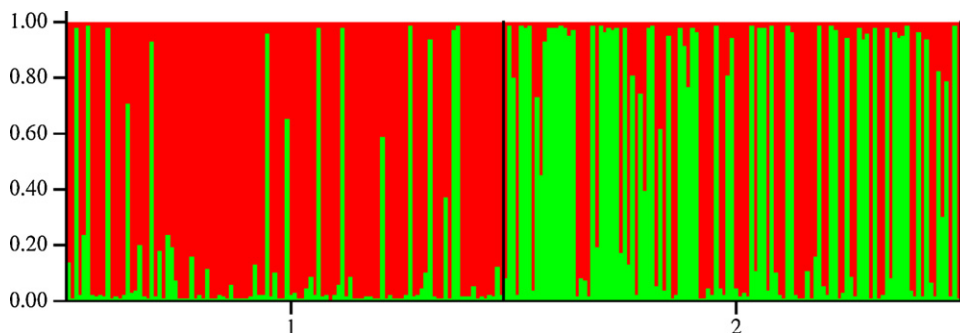


Fig. 4. Individual ancestry estimates of 225 samples from Guatemala; (1) Guatemalan-Mestizo and (2) ethnic groups. Bars represent individual ancestry estimates considering two parental populations.

proximity between Native American and Asian populations can be explained due to the Asian origin of Native American populations.

African populations are located in the top right part of the representation owing to the differences found in the markers: DYS390 (allele 21), DYS393 (allele 15), DYS437 (allele 13), DYS392 (allele 12), DYS389II (allele 32), and DYS19 (allele 17). Algeria is found to be next to the European Mediterranean populations.

To study the relationships between our populations, other American populations and European populations in a more accurate way; genetics distances, F -statistics, and Reynolds and Slatkin's genetic distances were calculated from previously published haplotypes and frequencies data. Multidimensional scaling (MDS) was carried out using SPSS v.15.0 to visualize the genetic distance matrix. As we can see in Fig. 3, there has been an admixture effect in the South American populations. Furthermore, we can observe all Guatemalan-Mestizo, Ecuador-Mestizo and USA-Hispanos populations clustered at the center of the image and the Mayan and other tribal populations situated at the right of the image. Iberian-origin samples from Portugal and Spain are located at the left part of the image.

The admixed proportions of our groups were calculated with ADMIX.PAS [30] and ADMIX95 program [31]. Using a two-parental model, the estimated admixture confirms that the Mestizo population is approximately a 50% mixture of the South-West Europe populations and Mayan population (The results are summarized in Table S4).

The composition of our population was also studied with STRUCTURE 2.3.1 [34,36] to investigate the presence of the population structure in the whole Guatemalan sample ($n = 225$). If we assumed more than one ancestral population ($K > 1$), then the estimated natural logarithm probability $\ln P(D)$ was found to be higher. We obtained the same results with the Mestizo population ($\ln P(D)$ became higher as K increased). For the Mayan population, $K = 1$ (admixture method) and $K = 3$ (no admixture method) were the highest values. Fig. 4 shows the two components that the program differentiates in the whole population structure of Guatemala. The differences between the Mestizo population (1) located at the left side and the Mayan population (2) located at the right side have been marked. However, the result, $K = 2$, is not the highest probability. In all the studied models, an increase in $\ln P(D)$ with the increasing K has been observed. Guatemala population has a great number of well-differentiated ancestry populations owing to the huge number of Mayan population groups (Kaqchiquel, K'iché, Mam, etc.) present before the origin of Mestizo populations.

The Y-linked markers analyzed allowed us to estimate the paternal ancestral components in Guatemalan-Mestizos. We observed that the Guatemalan-Mestizos' Y chromosomes studied are of South-West European origin, which must have obvious consequences regarding the absence of significant differences for most of the populations analyzed. All these results have

considerable implications for forensic examinations, because they point out that it is unnecessary to incorporate geographic structuring into Y-STRs databases, at least for most of the admixed populations.

This study represents one of the first steps in understanding Guatemala populations' admixture to produce successful experimental designs for admixture mapping. In particular, there has been an increasing interest about admixed populations in the biomedical field, because it is possible to make use of recent admixture for mapping genes underlying ethnic variation in disease risk.

This study followed the guidelines for publication of population data requested by the journal [4,37].

Acknowledgments

The authors thank the DNA donors and INACIF (Instituto Nacional de Ciencias Forenses de Guatemala) staff for kindly providing the DNA samples and their great effort in collecting information about the origin and ethnicity of each individual.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2011.04.003.

References

- [1] Central Intelligence Agency, The World Factbook, 2009.
- [2] Instituto Nacional de Estadística, XI Censo Nacional de Población y VI de Habitación, 2002.
- [3] J.J. Mulero, C.W. Chang, L.M. Calandro, R.L. Green, Y. Li, C.L. Johnson, et al., Development and validation of the AmpF/STR[®] Yfiler[™] PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system, *J. Forensic Sci.* 5 (2006) 64–75.
- [4] 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, DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, *Forensic Sci. Int.* 157 (2006) 187–197.
- [5] S. Willuweit, L. Roewer, Y chromosome haplotype reference database (YHRD): update, *Forensic Sci. Int. Genet.* 1 (2007) 83–87.
- [6] 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.
- [7] M. Nei, Genetic distance, in: S. Brenner, J.H. Miller (Eds.), *Encyclopedia of Genetics*, Academic Press, New York, 2001, pp. 828–832.
- [8] U. Toscanini, L. Gusmao, G. Berardi, A. Amorim, Á. Carracedo, A. Salas, et al., Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data, *Forensic Sci. Int. Genet.* 2 (2008) 274–280.
- [9] M. Tirado, A.M. López-Parra, C. Baeza, F. Bert, A. Corella, A. Pérez-Pérez, et al., Y-chromosome haplotypes defined by 17 STRs included in AmpF/STR Yfiler PCR Amplification Kit in a multi ethnic population from El Beni Department (North Bolivia), *Leg. Med.* 11 (2009) 101–103.
- [10] D.A. Silva, E. Carvalho, G. Costa, L. Tavares, A. Amorim, L. Gusmao, Y-chromosome genetic variation in Rio de Janeiro population, *Am. J. Hum. Biol.* 18 (2006) 829–837.

- [11] S.P. Schwengber, T. Kommers, C.H.F. Matte, P.E. Raimann, B.A. Carvalho, F.P.N. Leite, et al., Population data of 17 Y-STR loci from Rio Grande do Sul state (South Brazil), *Forensic Sci. Int. Genet.* 4 (2009) e31–e33.
- [12] F. González-Andrade, L. Roewer, S. Willuweit, D. Sánchez, B. Martínez-Jarreta, Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis, *Forensic Sci. Int. Genet.* 3 (2009) e83–e91.
- [13] J.M. Butler, A.E. Decker, P.M. Vallone, M.C. Kline, Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples, *Forensic Sci. Int.* 156 (2006) 250–260.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] B. Budowle, J. Ge, J. Low, C. Lai, W.H. Yee, G. Law, et al., The effects of Asian population substructure on Y STR forensic analyses, *Leg. Med.* 11 (2009) 64–69.
- [18] H.Y. Lee, M.J. Park, U. Chung, H. Young Lee, W.I. Yang, S.H. Cho, et al., Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father–son pairs, *Int. J. Legal Med.* 121 (2007) 128–135.
- [19] B. Zhu, Y. Wu, C. Shen, T. Yang, Y. Deng, X. Xun, et al., Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic group residing in Qinghai province of China, *Forensic Sci. Int.* 175 (2008) 238–243.
- [20] M. Shi, R. Bai, X. Yu, J. Lv, B. Hu, Haplotype diversity of 22 Y-chromosomal STRs in a southeast China population sample (Chaoshan area), *Forensic Sci. Int. Genet.* 3 (2009) e45–e47.
- [21] A. Gómez-Martín, L.J. Martínez-González, M.J. Alvarez-Cubero, C. Entrala, F. Fernández-Rosado, A. González-Liñan, et al., Population data for Y-chromosome haplotypes and haplogroups defined by 16 STRs (AmpF/STR[®] Yfiler[™] Kit) in Spain, *IALM*. XXI (2009).
- [22] A.M. Bento, M. Carvalho, V. Lopes, A. Serra, H.A. Costa, L. Andrade, et al., Distribution of Y-chromosomal haplotypes in the Central Portuguese population using 17-STRs, *Forensic Sci. Int. Genet.* 4 (2009) e35–e36.
- [23] 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.
- [24] N. Haliti, M. Čarapina, M. Mašić, D. Strinović, I.M. Klarić, M. Kubat, Evaluation of population variation at 17 autosomal STR and 16 Y-STR haplotype loci in Croats, *Forensic Sci. Int. Genet.* 3 (2009) e137–e138.
- [25] 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.
- [26] B. Berger, A. Lindinger, H. Niederstätter, P. Grubwieser, W. Parson, Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay, *Int. J. Legal Med.* 119 (2005) 241–246.
- [27] M. Goedbloed, M. Vermeulen, R.N. Fang, M. Lembring, A. Wollstein, K. Ballantyne, et al., Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpF ISTR[®] Yfiler[®] PCR amplification kit, *Int. J. Legal Med.* (2009) 1–12.
- [28] M. Raymond, F. Rousset, An exact test for population differentiation, *Evolution* 41 (1995) 823–833.
- [29] J.B. Kruskal, Nonmetric multidimensional scaling: a numerical method, *Psychometrika* 29 (1964) 115–129.
- [30] J.C. Long, The genetic structure of admixed populations, *Genetics* 127 (1991) 417–428.
- [31] R. Chakraborty, Estimation of race admixture. A new method, *Am. J. Phys. Anthropol.* 42 (1975) 507–511.
- [32] M. Stephens, P. Donnelly, Inference in molecular population genetics, *R. Stat. Soc. Ser. B* (2000).
- [33] 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.
- [34] 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.
- [35] L. Excoffier, Human diversity: our genes tell where we live, *Curr. Biol.* 13 (2003) R134–R136.
- [36] J.K. Pritchard, M. Stephens, N.A. Rosenberg, P. Donnelly, Association mapping in structured populations, *Am. J. Hum. Genet.* 67 (2000) 170–181.
- [37] 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.

FOR THE RECORD

Esther Martinez-Espin,¹ M.S.; Luis J. Martínez-Gonzalez,¹ M.S.; Francisco Fernandez-Rosado,² M.S.; Carmen Entrala,² Ph.D.; J. Carlos Alvarez,¹ Ph.D.; Jose A. Lorente,¹ M.D., Ph.D.; Bruce Budowle,³ Ph.D.; and Myriam Ovalle de Monroy,⁴ M.A.

Guatemala Mestizo Population Data on 15 STR Loci (Identifiler[®] Kit)

POPULATION: Guatemala mestizo.

KEYWORDS: forensic science, DNA typing, population genetics, Guatemala, Identifiler[®] kit, D3S1358, TH01, D21S11, D18S51, D2S1338, D5S818, D13S317, D7S820, D16S539, CSFIPO, D19S433, vWA, D8S1179, TPOX, FGA

Sample Preparation

Blood samples were obtained by venipuncture from unrelated individuals ($n = 200$) living in the State of Guatemala, and spotted on an FTA paper (Whatman, Florham Park, NJ).

PCR

Approximately 1 ng of DNA was used in each amplification. The samples were amplified using the AmpF/STR[®] Identifiler[®] kit (Applied Biosystems, Foster City, CA). The alleles were separated and detected using an Applied Biosystems ABI310 genetic analyzer.

Analysis of Data

The frequency of each allele for each locus was calculated from the numbers of each genotype in the sample set (i.e., the gene

count method). Unbiased estimates of expected heterozygosity were computed as described by Edwards et al. (1). Possible divergence from Hardy–Weinberg expectations (HWE) was tested by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies (1–4) and the exact test (5), based on 2000 shuffling experiments. The program for this analysis was kindly provided by R. Chakraborty (University of Cincinnati, Center for Genome Research, Department of Environmental Health, Cincinnati, OH).

The distributions of the observed allele frequencies for the 15 STR loci are shown in Table 1. The most informative loci are FGA and D18S51, and the least discriminating are TPOX and D3S1358. There were no detectable departures from HWE for the 15 loci. The combined power of discrimination is >0.99999999 , and the combined power of exclusion is 0.99999875.

The complete data are available either by accessing <http://www.lorgen.com/forense> or emailing the contact author (jllorente@ugr.es).

TABLE 1—Observed allele frequencies for the 15 Identifiler[®] loci.

Allele	D3S1358	TH01	D21S11	D18S51	D2S1338	D5S818	D13S317
4		0.0050					
5		0.0025					
6		0.4200					
7		0.2500				0.0675	
8		0.0500				0.0050	0.0550
9		0.0900		0.0025		0.0825	0.2475
9.3		0.1675					
10		0.0150				0.0425	0.1175
11				0.0025		0.4475	0.1850
12				0.1000		0.2400	0.1900
13	0.0025			0.1150		0.1050	0.1375

¹Laboratory of Genetic Identification, Department de Medicina Legal, Facultad de Medicina, Universidad de Granada, 18012 Granada, Spain.

²LORGEN GP, R & D Division, PT, Ciencias de la Salud—BIC, 18100 Armilla, Granada, Spain.

³FBI Laboratory Division, Quantico, VA 22135.

⁴Departamento Tecnico Cientifico Ministerio Publico de Guatemala, Ciudad de Guatemala, Guatemala.

TABLE 1—Continued.

Allele	D3S1358	TH01	D21S11	D18S51	D2S1338	D5S818	D13S317
14	0.0600			0.2300		0.0100	0.0675
15	0.4825			0.1475	0.0025		
16	0.2375			0.0875	0.0075		
16.2	0.0025						
17	0.1250			0.1450	0.1150		
18	0.0850			0.0700	0.0325		
19	0.0050			0.0475	0.1800		
20				0.0225	0.1600		
21				0.0225	0.0225		
22				0.0050	0.1225		
23				0.0025	0.2350		
24					0.0625		
25					0.0450		
26			0.0025		0.0150		
27			0.0050				
28			0.0750				
29			0.2550				
30			0.2350				
30.2			0.0025				
31			0.0525				
31.2			0.1425				
32			0.0100				
32.2			0.1475				
33			0.0025				
33.2			0.0575				
34			0.0025				
34.2			0.0050				
35			0.0050				
Homozygosity test*	0.855	0.728	0.392	0.463	0.189	0.884	0.447
Exact test†	0.750	0.381	0.203	0.056	0.331	0.522	0.204
PD	0.85800000	0.87855000	0.94850000	0.96040000	0.95690000	0.87745000	0.94460000
PE	0.45403932	0.49390020	0.65700908	0.73085064	0.70273600	0.50223640	0.65646563

Allele	D7S820	D16S539	CSF1PO	D19S433	vWA	D8S1179	TPOX	FGA
7	0.0075		0.0050					
8	0.0675	0.0050	0.0100			0.0050	0.4975	
9	0.0475	0.1125	0.0225			0.0075	0.0450	
10	0.2075	0.2400	0.2350	0.0050		0.0600	0.0425	
10.2				0.0025				
10.3	0.0050							
11	0.3075	0.2550	0.2625	0.0075		0.0400	0.2575	
11.2				0.0025				
12	0.2825	0.2550	0.4000	0.0625		0.1425	0.1525	
12.2				0.0150				
13	0.0675	0.1200	0.0600	0.2100	0.0050	0.3300	0.0050	
13.2				0.1600				
14	0.0075	0.0100	0.0025	0.2600	0.0800	0.2650		
14.2				0.0400				
15		0.0025	0.0025	0.1175	0.0775	0.1175		
15.2				0.0850				
16				0.0175	0.3675	0.0300		
16.2				0.0100				
17				0.0050	0.3100	0.0025		
18					0.1050			0.0025
19					0.0425			0.0975
20					0.0100			0.0500
21					0.0025			0.0850
22								0.1200
23								0.1175
24								0.1675
24.2								0.0050
25								0.1700
26								0.1450
27								0.0350
28								0.0025
> 30								0.0025
Homozygosity test*	0.437	0.921	0.565	0.774	0.754	0.145	0.241	0.653
Exact test†	0.131	0.348	0.463	0.473	0.953	0.694	0.242	0.660
PD	0.90845000	0.91390000	0.86970000	0.95035000	0.89700000	0.91970000	0.82615000	0.96735000
PE	0.56031198	0.57576700	0.46487616	0.67634348	0.52640202	0.58221995	0.41414762	0.74300963

* χ^2_{df} based on unbiased estimate with 2000 shufflings.

†Exact test based on 2000 shufflings.

PD, power of discrimination; PE, power of exclusion.

Acknowledgment

We thank Applied Biosystems for kindly providing the Identifier[®] kit.

References

1. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 1992;12:241–53.
2. Chakraborty R, Smouse PE, Neel JV. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. *Am J Hum Genet* 1988;43:709–25.
3. Chakraborty R, Fornage M, Guegue R, Boerwinkle E. Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In: Burke T, Dolf G, Jeffreys AJ, Wolff R, editors. *DNA fin-*

- gerprinting: approaches and applications. Berlin: Birkhauser Verlag, 1991: 127–43.
4. Nei M, Roychoudhury AK. Sampling variances of heterozygosity and genetic distance. *Genetics* 1974;76:379–90.
 5. Guo SW, Thompson EA. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 1992;48:361–72.

Additional information and reprint requests:

Jose A. Lorente, M.D., Ph.D.
Department Medicina Legal
University of Granada
Av. Madrid 11
E-18012 Granada
Spain
E-mail: jlorente@ugr.es

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

Legal Medicine

journal homepage: www.elsevier.com/locate/legalmed

Announcement of Population Data

Genetic variation of 15 autosomal microsatellite loci in a Nayarit population (Mexico)

M.J. Alvarez-Cubero^a, L.J. Martinez-Gonzalez^{a,*}, D. Virgen-Ponce^b, E.A. Delgado-Najar^b, G.H. Moscoso-Caloca^b, J.C. Alvarez^{a,c}, J.A. Lorente^{a,c}

^aLaboratory of Genetic Identification, Department of Legal Medicine, University of Granada, Granada, Spain

^bGeneral Office of Expert Criminalistics Services, Laboratory of Forensic Genetic, Av. Tecnológico No. 3200 Ote, C.P. 63175 Tepic, Nayarit, Mexico

^cPfizer Center, University of Granada, Junta of Andalucía of Genomics and Oncology Research, Biomedical Research Center, Av. del Conocimiento s/n, Granada, Spain

ARTICLE INFO

Article history:

Received 3 May 2011

Received in revised form 22 July 2011

Accepted 25 July 2011

Available online xxx

Keywords:

STRs

Nayarit

Genetic similarities

Homogeneity

Mestizo

Genetic database

ABSTRACT

Fifteen STRs are studied to determine the allele frequencies' distribution and to evaluate the homogeneity of Nayarit populations. This study allows the identification of forensic efficiency parameters to be used in forensic genetics and to explore the genetic similarities between Nayarit and the neighboring countries such as Mexico, Brazil, Puerto Rico, Guatemala, Honduras, Bolivia and Costa Rica. The Hardy–Weinberg equilibrium, expected heterozygosity, matching probability, and power of discrimination, were calculated in the Nayarit population. We found that with respect to the studied markers, Nayarit genetic structure is homogeneous. In this study, it is established that Nayarit is genetically similar to the South American Mestizo population. The distribution of a set of these 15 STRs was analyzed with other South American populations as well as in the extensive set of neighboring populations from the literature (USA, Europe and Africa). We found significant differences exist between the isolated populations (Huastecos, Otomi from Sierra Madre and from Ixmiquilpan Valley) and Mestizo populations. Statistical analysis supports that Americans actual inhabitants and Europeans are genetically similar, while Africans and isolated populations from South America have more genetic differences.

© 2011 Elsevier Ireland Ltd. All rights reserved.

Population

Samples (231 individuals) from Nayarit, Mexico were analyzed for 15 STR loci, however for doing anthropological analysis and population's studies we should reduce the markers to 13. The selection of 13 STR loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, TH01, TPOX, CSF1PO and AMEL) is due to they constitute the core Genetic Loci of the United States national database, CODIS (Combined DNA Index System) [1]. The population samples consist of five regional populations in Nayarit: Acajoneta: 29, Santiago: 27, Buceras: 31, Ixtlan del Rio: 32 and Tepic: 112. It was also completed with population data from Mexico and other countries, such as South and Central America (Guatemala, Costa Rica, Puerto Rico, Brazil, Honduras and Bolivia), Europe (Spain and Italy), Africa (Angola and Tanzania) and some American descendent (Native American, Afri-

can, Hispanic and European). The map of Nayarit and the five populations of this region are represented in Figs. 1 and 2.

DNA extraction

Blood samples and buccal swabs were collected by Forensic Genetic Laboratory of the Attorney General's Office of Nayarit state, using the Helsinki protocol and collected with informed consent which was approved by Ethics Committee. Some personal questions were asked in the collection of the samples in order to have sufficient anthropological information of their ancestors and an accurate selection of the individuals.

PCR

Approximately 0.5–1 ng/μl of template DNA was amplified using Power Plex16 kit (Promega®) [2] (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D19S433, vWA, TPOX, D18S51, D5S818, Penta D, Penta E, and FGA), by a GeneAmp System 2400 thermal cycler. The kit Power Plex16 was selected because it is the common kit for forensic uses, however if we want to do anthropological analysis and population's studies we should reduce the markers to 13. The selection of 13 STR loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818,

Abbreviations: FST, fixation index; MDS, multidimensional scale; PCR, polymerase chain reaction; PD, power of discrimination; PE, power of exclusion; PI, power of inclusion; PIC, polymorphism information content; STRs, short tandem repeats; TPI, typical paternity index.

* Corresponding author. Tel.: +34 958 24 99 50; fax: +34 958 24 61 07.

E-mail address: luisjaviermg@ugr.es (L.J. Martinez-Gonzalez).

1344-6223/\$ - see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.legalmed.2011.07.003

Please cite this article in press as: Alvarez-Cubero MJ et al. Genetic variation of 15 autosomal microsatellite loci in a Nayarit population (Mexico). *Leg Med* (2011), doi:10.1016/j.legalmed.2011.07.003



Fig. 1. Map of Mexico.



Fig. 2. Nayarit map.

D13S317, D7S820, D16S539, TH01, TPOX and CSF1PO) is due to they constitute the core Genetic Loci of the United States national database, CODIS (Combined DNA Index System) [1].

Typing

Electrophoresis, detection of PCR products, and genotyping were carried out on the ABI PRISM 310 Genetic Analyzer (Applied Biosystems®) using GeneMapper® ID Software v3.2 (Applied Biosystems®, Foster city, CA). The raw data obtained was also analyzed by GeneMapper® ID-X Software v1.1 and all the haplotypes were confirmed in the Identification Laboratory, University of Granada, Spain.

Quality control

Positive and negative controls as specified in the Power Plex16 kit (Promega®) user's manual. The data were analyzed and verified by two independent analysts.

Analyses of data

The population's genetic structure was analyzed using methods implemented in the Arlequin v.3.5 software [3], obtaining data, such as Hardy–Weinberg equilibrium, expected heterozygosity and observed heterozygosity. Matching probability, power of discrimination, polymorphism information content, probability of exclusion, and typical paternity index, which were calculated with a modified version of Powerstats v.1.2 [4]. To test for population stratification, the STRUCTURE v.2.1 [5] program was used. A correspondence analysis was calculated with STATISTICA v.7.1 (StatSoft, Inc., Tulsa, USA). The statistical significance of the pairwise genetic distances FST values was estimated by permutation analysis using 2000 permutations. Reynolds' distance [6] was calculated with the Gendist genetic tool of the Phylip v.3.1 software pack [7]. Multidimensional scaling (MDS) [8] analysis of Reynolds' distance values was performed using the software package SPSS v.15.0.

Results

The allele frequency distributions for the 15 STR loci studied in the Nayarit population and statistical parameters (Hardy–Weinberg equilibrium, expected heterozygosity, observed heterozygosity, matching probability, power of discrimination, polymorphism information content, probability of exclusion, and typical paternity index) are summarized in Table 1. To evaluate whether the main regions of the present day Nayarit population are homogeneous with respect to the studied STRs, the analysis of molecular variance (AMOVA) [9] on grouped Mexican populations was performed.

The expected heterozygosity, the power of inclusion (PI), and the power of discrimination (PD) calculated from the allele frequencies obtained for the Nayarit population revealed that in combination, the 15 STR loci have a high forensic efficiency.

Table 1
Allele frequencies and statistical parameters of the 15 STRs loci in the Nayarit population ($n = 231$). The minimum frequency is 0.0021645.

Allele	D3S1358	TH01	D21S11	D18S51	Penta E	D5S818	D13S317	D7S820	D16S539	CSF1PO	Penta D	vWA	D8S1179	TPOX	FGA
1	-	-	-	-	-	-	-	-	-	-	0.1100	-	-	-	-
5	-	0.0020	-	-	0.0190	-	-	-	-	-	0.0020	-	-	-	-
6	-	0.2710	-	-	-	-	-	-	-	-	-	-	-	0.0040	-
7	-	0.3920	-	0.0020	0.0910	0.0400	-	0.0130	-	0.0020	0.0020	0.0020	-	0.0040	-
8	-	0.1020	-	-	0.0260	0.0110	0.0690	0.1210	0.0300	0.0040	0.0120	-	0.0130	0.5280	-
9	-	0.0950	-	0.0020	0.0060	0.0550	0.2190	0.0690	0.1000	0.0320	0.1250	0.0020	0.0090	0.0610	-
9.2	-	0.0020	-	-	-	-	-	-	-	-	-	-	-	-	-
9.3	-	0.1800	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	0.0190	-	0.0060	0.0350	0.0440	0.0950	0.2770	0.1710	0.2340	0.1790	0.0020	0.0760	0.0630	-
11	-	-	-	0.0170	0.0710	0.4430	0.2400	0.3140	0.2970	0.2810	0.1270	0.0020	0.0670	0.2160	-
12	-	-	-	0.1620	0.1820	0.3060	0.2010	0.1750	0.2730	0.3610	0.1570	-	0.0740	0.1130	-
13	0.0040	-	-	0.1060	0.0840	0.0950	0.1190	0.0240	0.1150	0.0710	0.1930	0.0020	0.3460	0.0040	-
13.2	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0040	-
14	0.0610	-	-	0.1560	0.0930	0.0040	0.0540	0.0040	0.0150	0.0040	0.0740	0.0670	0.2730	0.0040	-
15	0.4330	-	-	0.1580	0.1150	-	0.0020	-	-	0.0040	0.0100	0.1150	0.0910	0.0020	-
15.2	-	-	-	-	-	-	-	-	-	-	-	0.0090	-	-	-
16	0.2730	-	-	0.0910	0.0970	0.0020	-	-	-	-	0.0020	0.3070	0.0450	-	-
17	0.1430	-	-	0.1450	0.0580	-	-	-	-	-	-	0.2920	0.0020	-	-
18	0.0800	-	-	0.0670	0.0220	-	-	-	-	-	-	0.1280	-	-	0.0150
19	0.0060	-	-	0.0350	0.0280	-	-	-	-	-	-	0.0610	0.0020	-	0.0640
20	-	-	0.0040	0.0220	0.0370	-	-	-	-	-	-	0.0110	0.0020	-	0.0880
21	-	-	-	0.0060	0.0220	-	-	-	-	-	-	-	-	-	0.1190
22	-	-	-	0.0110	0.0060	-	-	0.0020	-	-	0.0080	-	-	-	0.1310
23	-	-	-	0.0090	-	-	-	-	-	-	-	-	-	-	0.1440
24	-	-	-	-	0.0060	-	-	-	-	-	-	-	-	-	0.1420
25	-	-	-	0.0020	-	-	-	-	-	-	-	-	-	-	0.1370
25.2	-	-	0.0020	-	-	-	-	-	-	-	-	-	-	-	-
26	-	-	0.0020	-	-	-	-	-	-	-	-	-	-	-	0.1260
27	-	-	0.0110	-	-	-	-	-	-	-	-	-	-	-	0.0180
28	-	-	0.0700	-	-	-	-	-	-	-	-	-	-	-	0.0100
29	-	-	0.2190	-	-	-	-	-	-	-	-	-	-	-	0.0040
29.2	-	-	0.0020	-	-	-	-	-	-	-	-	-	-	-	-
30	-	-	0.3090	-	-	-	-	-	-	-	-	-	-	-	-
30.2	-	-	0.0130	-	-	-	-	-	-	-	-	-	-	-	-
31	-	-	0.0700	-	-	-	-	-	-	-	-	-	-	-	-
31.2	-	-	0.1100	-	-	-	-	-	-	-	-	-	-	-	-
32	-	-	0.0180	-	-	-	-	-	-	-	-	-	-	-	-
32.2	-	-	0.1210	-	-	-	-	-	-	-	-	-	-	-	-
33.2	-	-	0.0350	-	-	-	-	-	-	-	-	-	-	-	-
34.1	-	-	0.0020	-	-	-	-	-	-	-	-	-	-	-	-
34.2	-	-	0.0110	-	-	-	-	-	-	-	-	-	-	-	-
Homozygosity*	0.2908	0.2319	0.1798	0.1222	0.0925	0.3047	0.1753	0.2249	0.2140	0.2698	0.1615	0.2186	0.2212	0.3441	0.1151
Exacttest**	0.8026	0.3856	0.0242	0.2554	0.4501	0.1956	0.3967	0.2455	0.9433	0.4230	0.2257	0.3583	0.3169	0.3566	0.0148
PD	0.8610	0.9080	0.9450	0.9680	0.9810	0.8560	0.9410	0.9070	0.9170	0.8740	0.9590	0.9170	0.9100	0.8320	0.9680
PE	0.5150	0.5010	0.5800	0.7000	0.7520	0.4860	0.6170	0.5530	0.6170	0.4510	0.7110	0.5770	0.5850	0.3520	0.8550
PIC	0.6600	0.7300	0.8000	0.8600	0.9000	0.6500	0.8000	0.7400	0.7500	0.6800	0.8400	0.7500	0.7500	0.6100	0.8700
TPI	2.0300	1.9600	2.3800	3.4000	4.1300	1.8900	2.6300	2.2200	2.6300	1.7500	3.5200	2.3600	2.4100	1.4300	7.0600

Abbreviations: PD – power of discrimination; PE – power of exclusion; PIC – polymorphism information content; TPI – typical paternity.

* χ^2_{df} based on unbiased estimate with 2000 shufflings.

** Exact test based on 2000 shufflings.

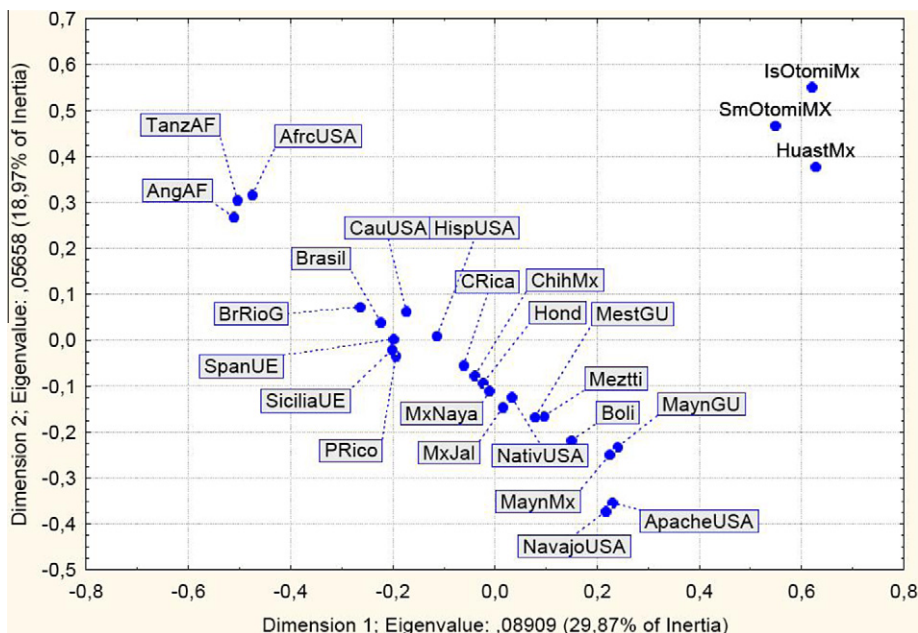


Fig. 3. Two-dimensional plot of correspondence analysis. (For the interpretation of the abbreviations used in this figure, the reader is referred to Appendix A). To make an easier interpretation of the data, this plot only contains data from the population, and those from the markers are omitted.

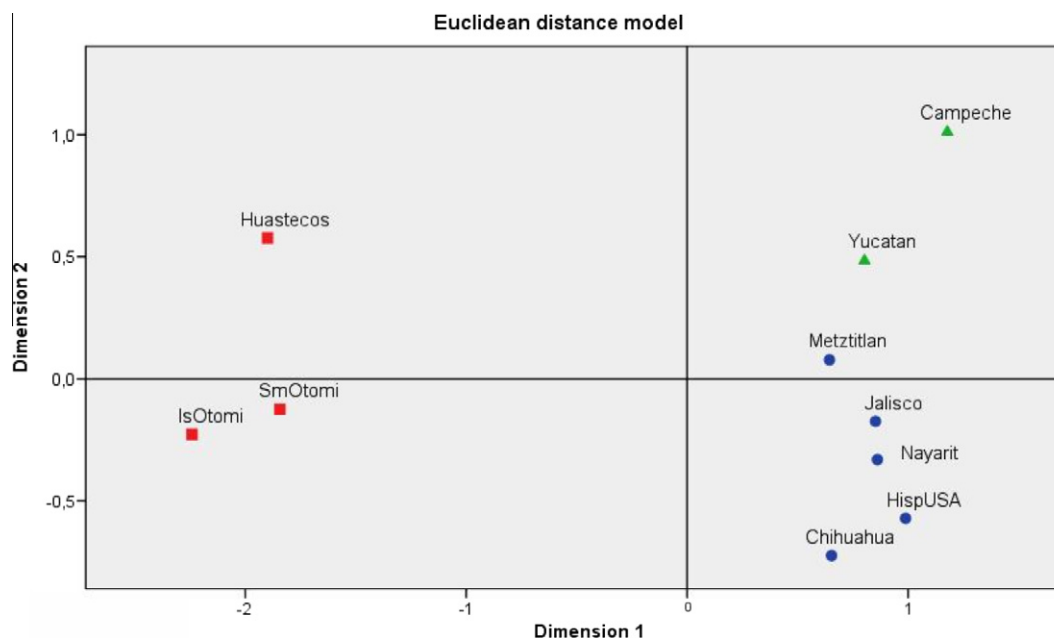


Fig. 4. Euclidean distance model of an MDS representation plot from the analysis of Reynolds distance for the Mexican populations and Hispans from the USA.

The combined Power of discrimination (PD) and the combined Power of exclusion (PE) for the 15 tested STR loci were 0.999999999 and 0.999999333, respectively.

To evaluate whether the main regions of the present day Nayarit population are homogeneous with respect to the studied STRs, AMOVA performed. We were unable to detect statistically significant differences between the subregions of Nayarit (Acaponeta, Bucerías, Itxlan del Rio, and Tepic), which indicates genetic homogeneity among the analyzed populations. We also observed that all loci are in Hardy–Weinberg equilibrium. STRUCTURE v.2.1 was also used to analyze homogeneity, and we always observed the highest probability in $K = 1$ with all the studied models. Thus, it can be concluded that the Nayarit population is homogeneous [5]

and we cannot find a statistically significant component of indigenous population.

Other remarks

To further study the genetic distances and the differences between the American populations, we also included the African and European populations. STATISTICA v.7.1 program was performed with all the populations cited in Appendix A to calculate the correspondence analysis. We analyzed the differences between the populations' STR frequencies. We reduced the analysis to 13 STRs to achieve unification of the data [10].

All the Mestizo populations from America and Europe are clustered, while African population and populations from Central and

North America, Mayan, and Nadane (Navajo and Apache) (Fig. 3) [10] are quite dispersed. The most dispersed are the Mexican Huastecos groups (Otomi from Sierra Madre and Ixmiquilpan and Huastecos) [11], while premayan population groups are in fact isolated. There are some Mestizo populations like Brazil [12] genetically quite similar to the African populations and others like Mestizos from Bolivia [13] or Meztitlan [14] to the Native American populations. We can see that this situation correlates with the genetic homogeneity between these populations, and that the distribution of populations on the correspondence analysis plot did not fully correlate with geography, but with the ancestry of these population groups [15]. The distribution of all the populations depends on a range of markers that make a population suitable in one specific part of the plot.

We studied the genetic distances and the differences between the Mexican population and USA hispanic descendent, and calculated the matrix of coancestry coefficients [6] using the tool GenDist (Phylip v.3.1 software). The Euclidean distance model of a multidimensional scale (MDS) and a representation plot from the MDS analysis is shown in Fig. 4. MDS was calculated with the SPSS software. As can be seen in Fig. 4, the Mexican population can be divided into natives and Mestizos. We could observe all Mestizo population groups clustered on the right part at the top of the image, and the Mayan population at the bottom. Isolated protomayan populations are dispersed and situated on the left of the image.

There were significant differences between the population of North America, indicating that isolated groups are different from the others owing to the geographical and cultural factors. Furthermore, differences between the Mayan and Mestizo population could also be a consequence of these factors, while we also observed that populations from Jalisco, Meztitlan, and Nayarit have more genetic similarities, owing to their nearer geographical locations.

Acknowledgments

This study was supported by the Forensic Genetic Laboratory of the Attorney General's Office of Nayarit state. The software was provided by the University of Granada. We would like also to thanks to the Ministry of Education (Spain) for the grant conceded to one of the authors (Maria Jesus Alvarez Cubero). This grant is called FPU (Formation for Professors).

Appendix A. Supplementary data

Supplementary data (completed list of populations used in the present study for comparative analyses) associated with this article can be found, in the online version, at [doi:10.1016/j.legalmed.2011.07.003](https://doi.org/10.1016/j.legalmed.2011.07.003).

References

- [1] Baechtel FS, Monson KL, Forsen GE, Budowle B, Kearney JJ. Tracking the violent criminal offender through DNA typing profiles – a national database system concept. *EXS* 1991;58:356–60.
- [2] Finis CJ. Megaplex STR Analysis from a Single Amplification: validation of the PowerPlex® 16 System. *Profiles in DNA* 2001;3.
- [3] Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resources* 2010;10:564–7. Arlequin suite ver 3.5.
- [4] Tereba A. Tools for Analysis of Population Statistics. *Profiles in DNA*. 2001.
- [5] Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. *Am J Hum Genet* 2000;67:170–81.
- [6] Reynolds J, Weir BS, Cockerham CC. Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 1983;105:769–79.
- [7] Felsenstein J. PHYLIP (Phylogeny Inference Package), version 3.62 [computer program]. Department of Genome Sciences, University of Washington; 2004.
- [8] Witten DM, Tibshirani R. Supervised multidimensional scaling for visualization, Classification, and bipartite ranking. *Comput Stat Data Anal* 2011;55:789–801.
- [9] Raymond M, Rousset F. An exact test for population differentiation. *Evolution* 1995;41:823–33.
- [10] Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations. *J Forensic Sci* 2001;46:453–89.
- [11] Barrot C, Sánchez C, Ortega M, González-Martín A, Brand-Casadevall C, Gorostiza A, et al. Characterisation of three Amerindian populations from Hidalgo State (Mexico) by 15 STR–PCR polymorphisms. *Int J Legal Med* 2005;119:111–5.
- [12] De Assis Poiars L, De Sá Osorio P, Spanhol FA, Coltre SC, Rodenbusch R, Gusmão L, et al. Allele frequencies of 15 STRs in a representative sample of the Brazilian population. *Forensic Science International*. *Genetics* 2010;4:e61–3.
- [13] Riccardi LN, Melean G, Rada A, Tirado N, Genuardi M, Ricci U. Genetic profiling of Bolivian population using 15 STR markers of forensic importance. *Leg Med* 2009;11:149–51.
- [14] Gorostiza A, González-Martín A, Ramírez CL, Sánchez C, Barrot C, Ortega M, et al. Allele frequencies of the 15 AmpF/Str Identifier loci in the population of Meztitlan (Estado de Hidalgo), México. *Forensic Sci Int* 2007;166:230–2.
- [15] Licea-Cadena R, Rizzo-Juárez R, Muñoz-Lozano E, Páez-Riberos L, Rangel-Villalobos H. Population data of nine STRs of Mexican-mestizos from Veracruz (Central South-Eastern, Mexico). *Leg Med* 2006;8:251–2.

Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set

J. Carlos Alvarez^{a,1}, Diana L.E. Johnson^{b,*}, Jose A. Lorente^a, Esther Martinez-espín^a, Luis J. Martinez-Gonzalez^a, Marc Allard^b, Mark R. Wilson^c, Bruce Budowle^d

^a Department of Legal Medicine, University of Granada, Madrid, Granada, Spain

^b Department of Biological Sciences, George Washington University, Lisner Hall 340, Washington, DC 20052, USA

^c Federal Bureau of Investigation Academy, Chem. Biol. Science Unit, FBI Laboratory, Quantico, VA 22135, USA

^d Federal Bureau of Investigation, Laboratory Division, 2501 Investigation Pkwy, Quantico, VA 22135, USA

Received 19 May 2006; received in revised form 2 May 2007; accepted 9 May 2007

Available online 5 July 2007

Abstract

Population data on the hypervariable regions of the mitochondrial DNA (mtDNA) genome are used to convey the relative rarity of mtDNA profiles obtained from evidence samples and of profiles used to identify missing persons. In this study, mtDNA profiles of Spanish individuals ($n = 312$) were analyzed to describe haplogroup distributions and to determine relevant single nucleotide polymorphisms (SNPs) of those haplogroups. All nine common European haplogroups were observed in the sample, and these were divided into subgroups when possible. Haplogroup H was the most common haplogroup. The haplogroups U, J, T, and V were the next most frequent groups, each occurring at a frequency of 6.4% or greater. In addition, African and Asian sequences were present though rare in the samples. The data were compared with and found to be similar to other published data sets. There were 109 SNPs observed in the data set, including 10 positions not previously reported. The most variable sites are consistent with other studies.

© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Mitochondrial DNA; Haplogroup designation; Haplotype; Spanish; Control region

1. Introduction

Mitochondrial DNA (mtDNA) has been studied extensively in order to describe human variation and population substructure, infer information about human migration patterns, and for human identity testing. There are several characteristics of mtDNA which make it useful for these roles. Mitochondria are maternally inherited and their genome is believed to be free of recombination. Therefore, the mitochondrial genome can be treated as a single locus with an effective population size that is 1/4 of the size for nuclear

autosomal genes [1]. MtDNA also has been observed to have a high mutation rate which contributes to the high levels of variability [2]. Within the mtDNA the hypervariable regions (HVI and HVII) evolve particularly rapidly [3]. The HVI includes nucleotide positions 16,024–16,365. The HVII includes nucleotide positions 73–340.

The high variability in mtDNA sequence has been used by forensic scientists for identity testing in both criminal and missing person cases. When one cannot exclude that an evidence sample and a reference sample may originate from the same source based on mtDNA analysis, an inference of the rarity of the sequence is required. To estimate the frequency of a mtDNA profile, population databases are used [4]. This study describes the mtDNA genetic variation in a set of individuals from Spain.

* Corresponding author. Tel.: +1 202 994 7121; fax: +1 202 994 6100.

E-mail address: dejohnsn@gwu.edu (D.L.E. Johnson).

¹ These authors contributed equally to this work.

2. Materials and methods

2.1. Subjects

DNA samples were obtained from unrelated individuals ($n = 312$) residing and born in Spain and whose mother also was born in Spain (Fig. 1 and Table 1). Seventeen of the individuals' mothers were born in Spain, but immigrated back to Spain after giving births outside the country. The study was approved by the UGR Ethical Committee and by the Spanish Central Agency for Data Protection. All individuals gave informed consent.

2.2. Data collection

DNA was obtained from buccal cells and extracted by a standard Proteinase K and phenol chloroform method. The HVI (nucleotide positions 16,024–16,365) and HVII (nucleotide positions 73–340) regions were amplified as previously described employing a GeneAmp System 2400 thermal cycler [5]. The amplified fragments were purified by filtration through Microcon 100[®] tubes (Millipore, Billerica, MA). Sequencing was generated using the kit ABI PRISM[®] BigDye Terminator Cycle Sequencing Ready Reaction Kits and primers identical to those employed for amplification, according to the protocol provided with the kit. DNA products were purified by ethanol precipitation and the fragments were analyzed using an automated DNA sequencer ABI PRISM 310. Each template was sequenced in both directions and the consensus sequence was aligned and compared with the Anderson reference sequence [6] using the software Navigator (ABD) [7].

2.3. Data availability

The sequence designations are based on comparisons to the revised Cambridge Reference Sequence (rCRS, [6,8]). All sequences are available at GenBank, Accession Nos. DQ811134–DQ811757.

2.4. Analysis and haplogroup classification

The analysis consisted of constructing an alignment, performing parsimony analysis based on the alignment, and generating a phylogenetic tree derived from the parsimony analysis. The software packages Winclada and Nona were used to carry out parsimony analysis of the data. At least 2000 replicates of the parsimony ratchet were used to determine the most parsimonious solution. The characters were the variable nucleotide sequence positions and states listed in reference to the rCRS [6,8]. After independently determining the positions and states for all major groups on the tree, the data were compared to published mtDNA population data [7,9–21]. The haplogroups present in this study were determined by comparison of the defining characters to sites listed as important for distinguishing groups [12,22 and others listed below]. This approach allowed validation of the data set through independent analysis. Special emphasis was placed on the comparison of the patterns of mtDNA variation in this data set to Helgason, et al. [12] and Maca-Meyer, et al. [14]. Other studies that were used included Larruga, et al. [13] and Plaza, et al. [17]. The choice of studies was based on the rigorous, recent treatment of their data.

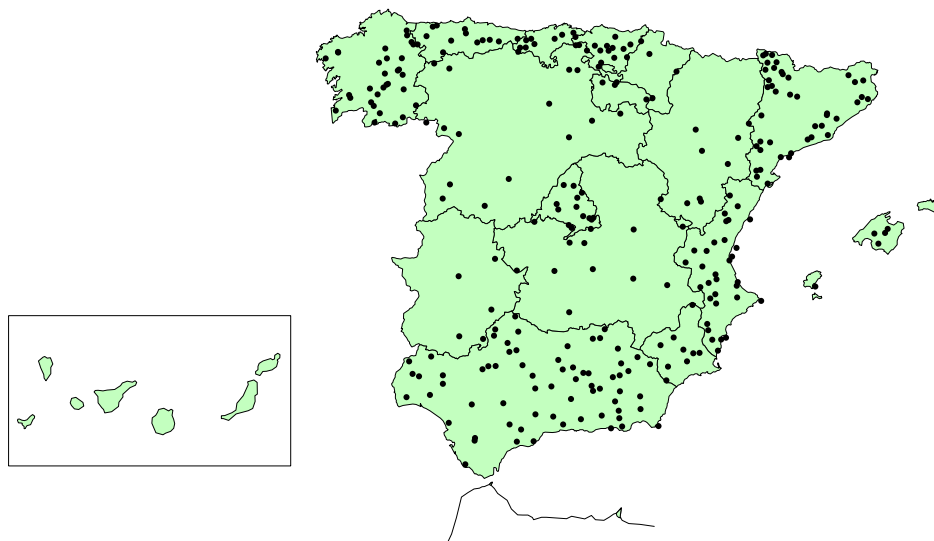


Fig. 1. Geographical distribution of the individuals sampled in this study. Each dot represents one individual.

Table 1
Geographical locations of samples

Samples	Province	Area within Spain
GRX151, GRX231, GRX271, GRX274, GRX275, GRX277, GRX282, GRX302	Canarias	Canarias
GRX021, GRX076, GRX077, GRX134	Albacete	Central
GRX043, GRX045, GRX129	Huesca	Central
GRX119, GRX146, GRX162, GRX163	La Rioja	Central
GRX041, GRX306, GRX307, GRX308, GRX309, GRX310, GRX311	Leon	Central
GRX012, GRX013, GRX014, GRX016, GRX023, GRX034, GRX035, GRX052, GRX090, GRX099, GRX100, GRX113, GRX117, GRX182, GRX190	Madrid	Central
GRX082	Palencia	Central
GRX203	Salamanca	Central
GRX048, GRX181	Segovia	Central
GRX046, GRX132	Soria	Central
GRX121	Teruel	Central
GRX096, GRX097, GRX122, GRX123, GRX217, GRX237, GRX239	Toledo	Central
GRX169, GRX276	Valladolid	Central
GRX047, GRX072, GRX087, GRX176	Zaragoza	Central
GRX028, GRX044, GRX078, GRX145, GRX243, GRX246, GRX251, GRX261, GRX283, GRX284, GRX285, GRX288, GRX297, GRX301, GRX303, GRX304	Alicante	East
GRX093, GRX118, GRX120, GRX130, GRX137, GRX258	Baleares	East
GRX017, GRX022, GRX026, GRX037, GRX050, GRX053, GRX058, GRX064, GRX073, GRX075, GRX079, GRX085, GRX086, GRX089, GRX094, GRX095, GRX104, GRX106, GRX109, GRX175, GRX221, GRX235, GRX299	Barcelona	East
GRX249, GRX250, GRX252, GRX253, GRX254, GRX255	Castellon	East
GRX127, GRX128	Girona	East
GRX056, GRX074, GRX098, GRX105, GRX180	Lleida	East
GRX024, GRX208, GRX209, GRX210, GRX211, GRX212, GRX225, GRX226	Murcia	East
GRX054, GRX092, GRX103, GRX156, GRX238, GRX260, GRX269, GRX270, GRX272, GRX279, GRX296, GRX300	Tarragona	East
GRX018, GRX027, GRX066, GRX069, GRX126, GRX172, GRX183, GRX184, GRX192, GRX216, GRX244	Valencia	East
GRX038, GRX061, GRX067, GRX110, GRX138, GRX139, GRX144, GRX167, GRX205, GRX223, GRX240, GRX259, GRX267, GRX268, GRX293, GRX294, GRX295	A coruña	North
GRX015, GRX222	Alava	North
GRX025, GRX029, GRX114, GRX161, GRX262, GRX263, GRX264, GRX265, GRX266, GRX278, GRX280, GRX281, GRX286	Asturias	North
GRX031, GRX040, GRX068, GRX084, GRX091, GRX136, GRX148, GRX171, GRX177, GRX178, GRX312	Cantabria	North
GRX063, GRX065, GRX080, GRX108, GRX133, GRX298	Guipuzcoa	North
GRX036, GRX147	Navarra	North
GRX215	Ourense	North
GRX062, GRX150, GRX173, GRX174, GRX193, GRX194, GRX195, GRX196, GRX197, GRX201, GRX202, GRX204, GRX206, GRX207	Pontevedra	North
GRX011, GRX081, GRX142, GRX152, GRX170, GRX256, GRX257	Vizcaya	North
GRX003, GRX107	Almeria	South
GRX032, GRX111	Badajoz	South
GRX290, GRX291, GRX305	Caceres	South
GRX070, GRX124, GRX125, GRX149, GRX157, GRX160, GRX198, GRX214, GRX227, GRX234	Cadiz	South
GRX030, GRX185, GRX186, GRX187, GRX188, GRX189, GRX248	Cordoba	South
GRX001, GRX002, GRX004, GRX006, GRX007, GRX008, GRX009, GRX102, GRX131, GRX213	Granada	South
GRX010, GRX088, GRX116, GRX166, GRX179, GRX191, GRX289	Huelva	South
GRX020, GRX042, GRX049, GRX051, GRX055, GRX059, GRX060, GRX158, GRX220	Jaen	South
GRX033, GRX039, GRX083, GRX135, GRX200, GRX218, GRX219, GRX232, GRX241, GRX242, GRX273	Malaga	South
GRX019, GRX071, GRX101, GRX153, GRX154, GRX164, GRX229, GRX230, GRX233, GRX245, GRX247	Sevilla	South
GRX287	Alemania	#N/A
GRX292	Belgica	#N/A
GRX115	Francia	#N/A
GRX140, GRX155, GRX168, GRX224, GRX236	Inglaterra	#N/A
GRX199, GRX228	Irlanda	#N/A
GRX112, GRX141, GRX143, GRX159, GRX165	Marruecos	#N/A
GRX005	Mexico	#N/A
GRX057	Principado De Andorra	#N/A

2.5. Sequence alignment

The sequences used in the analyses were the Spanish profiles ($n = 312$), as well as the rCRS, and one chimpanzee sequence that served as an outgroup [23]. The Spanish data set is composed of samples for which regions HVI (np 16,024–16,365) and HVII (np 73–340) were sequenced. Sequences were aligned based on a number of rules that provided for the consistent placement of gaps [24,25]. In the phylogenetic analysis, N's were inserted where data were absent. Using the gap-placement rules, the final alignment was 1175 bp. Two additional insertions, 16538.1 and 16538.2, that occurred only in the chimpanzee, were added giving a final length of 1177 bp. Fifty-five insertions were included in the mtDNA multiple alignment and compared to the rCRS [6,8].

2.6. Statistical methods

Statistical analysis was performed with the Arlequin 2000 software package [26].

3. Results and discussion

In studying the evolutionary tree of human mtDNA, a variety of phylogenetic methods have been employed and these have produced both common branching patterns and lists of characters that define the lineages.

3.1. Variation at SNP sites

In the group of 312 Spaniards, there were 221 haplotypes (70.8%) and 188 had unique substitutions found only in one individual. The haplotype diversity is 0.997. Twenty-four individuals (7.7%) carried the most frequent haplotype which was 263G, 315.1C.

SNPs were identified from a data set of human mtDNA hypervariable region sequences. The samples were collected from 312 individuals from Spain (Fig. 1). There were 10 new substitutions found that have not been reported in MITOMAP as human mitochondrial control region sequence polymorphisms (MITOMAP, May 4, 2005 update). These 10 sites all occur in one or two individuals except for 256T which is present in three individuals. Observation of these probably reflects sampling or some localized genetic variation.

3.2. Spanish haplogroup frequencies

There have been a number of studies on mtDNA variation in regions of Spain, Spain as a whole, and the Iberian Peninsula [7,10,12–14,16,17,19–21]. In addition there have been characterizations of mtDNA variation in Europe, the Middle East and Africa. (See for example: [9, 12, 15, 17, 21, 27, and 28] and references therein). The HVI and HVII sequences of the individuals in this study and the basic informative SNPs used to assign individuals to

haplogroups are listed in Table 2. There are many similarities between the published data sets for Spain, regions of Spain and the Iberian Peninsula [7,10,12–14,16,17,19–21] and the Spanish sample described here. The observed haplogroup frequencies for the Spanish individuals in this study are shown in Fig. 2. The Spanish haplogroup frequencies are similar to those compiled for the Iberian Peninsula in Helgason, et al. [12], Larruga, et al. [13], and Maca-Meyer, et al. [14] as seen in Fig. 3.

3.3. The H haplogroup

The H haplogroup cluster was observed in 42.3% of the individuals ($n = 132$) and was defined by 73A and lacked substitutions characteristic of other haplogroups. Twenty-four individuals (7.6%) carried 263G and 315.1C. There were five individuals who lacked 73A but who were classified tentatively as H based on their position on the phylogenetic tree. It was assumed that these had a reversion at the 73 site. The frequency of the haplogroup in this study is lower than Helgason et al. [12] reported, but probably is the result of different geographical sources of individuals typed in the data sets used in the Helgason report versus this study. Both Larruga, et al. [13] and Maca-Meyer et al. [14] have reported varying frequencies of the H haplogroup in different regions of Spain with the lowest frequency being 26.8% in Pasiegos samples and the highest being 60.1% in Leonese samples. The weighted average frequency of H for those two studies in Fig. 2 depends on relatively high frequencies of some rather small and isolated groups and may not be representative of a random sample. Based on the control region SNPs, the H haplogroup is not well defined and may include haplotypes with different common ancestors [29].

The H haplogroup was broken into clusters based on a variety of SNPs. The ones which produced the largest clusters ($n \geq 7$) were 16129A, 16189C, 16362C, and 152C. Clusters defined as H1, H4, and H5 [12] were observed and counted as part of the H haplogroup.

3.4. I haplogroup

The I haplogroup was characterized by 16129A, 16223T, 199C, 204C, and 250C. Five individuals (1.6%) were classified in this group. One individual classified as I lacked the 204C character, and this may be due to a reversion. Other characters that subdivided the group included 152C and 207A.

3.5. J haplogroup

Twenty-one individuals (6.7%) were found to carry the J haplogroup or one of its subgroups. The J haplogroup designation was based on 16069T, 16126C, and 295T. One individual classified as J lacked 16126T, but had a haplotype that was very similar to other individuals in the group. Overall 17 individuals (5.4%) were placed in

Table 2
 HVSI and HVSII haplotypes (with nucleotides characterizing haplogroup clusters) for the Spanish sample^a

Haplogroup/cluster (defining motif)	HVSI	HVSII
<i>H (73 A)</i>		
24 Samples ^b		
GRX130		73A
GRX301		73G
GRX098		73A, 93, 95
GRX061		73A, 125, 127
GRX144		73A, 150
GRX201		73A, 151
		73G, 152
5 Samples ^c		
GRX026, GRX 249, GRX267		73A, 152
GRX059		73A, 152, 182
GRX004		73A, 152, 191.1, 194
GRX282		73A, 152, 272
GRX138		73A, 185
GRX111		73A, 195
GRX027		73A, 195, 257
GRX295	16079, 16217, 16218, 16234	73A, 228
GRX006	16086, 16129, 16223	73A
GRX030, GRX 160	16093	73G, 152, 199, 207, 239, 250
GRX038	16093, 16126	73A, 93
GRX139	16093, 16270	73A
GRX194	16093, 16270, 16343	73A
GRX166	16093, 16355	73A
GRX246	16111, 16288, 16362	73A
GRX082, GRX 242	16126, 16209, 16362	73A, 146, 195
GRX141,	16126, 16355, 16362	73A
GRX308	16126, 16362	73G, 146, 152, 195
6 Samples ^d		
	16129	73A, 152
GRX119	16129, 16172	73A
GRX029	16140	73A
GRX147	16166, 16304	73A
GRX181	16169, 16184	73A
GRX010	16176	73A, 195
GRX037, GRX039	16176, 16219	73A, 146, 257
GRX291	16182, 16184, 16185, 16189, 16356	73A
GRX264	16183, 16184, 16185, 16189	73A
GRX044, GRX153	16183, 16187.1, 16189, 16239	73A, 152
GRX114	16184	73A, 146
GRX087, GRX088	16187.1, 16189	73A
GRX152	16187.1, 16189, 16192.1, 16290	73A, 152
GRX011	16187.1, 16189, 16214, 16234	73A, 152
GRX309	16189	73A
GRX085, GRX 200, GRX 212	16192	73A
GRX253	16209, 16311	73A
GRX135	16213	73A
GRX092	16216, 16311	73A
GRX066	16218	73A
GRX009	16218, 16328	73A
GRX220	16218, 16328, 16362	73A, 214, 292
GRX294	16221	73A, 143, 152
GRX280	16221, 16291	73G
GRX122	16224, 16249	73A, 235
GRX289	16235, 16291	73A
GRX178	16235, 16291	73A, 93
GRX304	16239, 16242	73A
GRX045	16248	73A, 152, 200
GRX129	16256	73A
GRX203	16256	73A, 152
GRX001	16260, 16311	73A, 93, 152
GRX162	16261	73A
GRX115	16262	73A
GRX189	16263, 16288, 16362	73A, 146, 195, 198
GRX105	16278	73A

(continued on next page)

Table 2 (continued)

Haplogroup/cluster (defining motif)	HVSI	HVSII
GRX217	16288, 16311, 16362	73A, 114, 146, 152, 195
GRX104, GRX 137	16291	73A
GRX148, GRX196	16293	73A
GRX211	16293, 16311	73A, 195
4 Samples ^e	16304	73A
GRX169	16304, 16311	73A, 93
GRX265	16304, 16319	73A, 152
GRX110, GRX 168	16311	73A
GRX240	16311, 16355	73A, 150
GRX245	16316	73A
GRX250	16325	73A, 152
GRX049	16335	73A, 237
GRX205	16344	73A, 93
GRX024, GRX 229	16354	73A
GRX079	16355	73A, 152
GRX157, GRX167	16362	73A
GRX046, GRX176, GRX 278	16362	73A, 150, 239
GRX076	16362	73A, 152
GRX302	16362	73A, 239
<i>I (16129A, 16223T, 199C, 204C, and 250C)</i>		
GRX042	16129, 16223	73G, 199, 204, 250
GRX140	16129, 16223	73G, 152, 199, 204, 207, 250
GRX171	16129, 16223, 16255	73G, 199, 204, 250
GRX172	16129, 16223, 16311	73G, 152, 199, 204, 250
<i>J (16069T, 16126C, and 295T)</i>		
GRX204	16069, 16126	73G, 146, 185, 188, 222, 228, 295
GRX293	16069, 16126	73G, 228, 295
GRX276	16069, 16126, 16153	73G, 185, 295
GRX033	16069, 16126, 16179, 16192, 16278, 16311	73G, 150, 152, 295
GRX055	16069, 16126, 16189	73G, 146, 195, 295
GRX287	16069, 16126, 16189, 16265, 16319	73G, 150, 185, 295
GRX003	16069, 16126, 16192, 16278, 16311	73G, 150, 152, 295
GRX034	16069, 16126, 16234, 16278	73G, 185, 188, 228, 295
GRX093	16069, 16126, 16243	73G, 185, 228, 295
GRX266	16069, 16126, 16266,	73G, 185, 228, 295
GRX023	16069, 16126, 16278	73G, 146, 185, 188, 228, 295
GRX126, GRX193	16069, 16126, 16278	73G, 185, 188, 228, 295
GRX288	16069, 16126, 16278	73G, 185, 188, 295
GRX133, GRX247	16069, 16126, 16300	73G, 185, 228, 295
GRX195	16069, 16278	73G, 185, 188, 228, 295
<i>J1a (16069T, 16126C, 16145A, 16231C, 16261T, and 295T)</i>		
GRX065, GRX077	16069, 16126, 16145, 16231, 16261	73G, 150, 152, 195, 215, 295, 319
<i>J1b (16069T, 16126C, 16145A, 16222T, 16261T, and 295T)</i>		
GRX270	16069, 16126, 16145, 16222, 16261	73G, 295
<i>J1b1 (16069T, 16126C, 16145A, 16172C, 16222T, 16261T, and 295T)</i>		
GRX218	16069, 16126, 16145, 16172, 16222, 16261	73G, 242, 295
<i>K (16224C and 16311C)</i>		
GRX235	16086, 16189, 16224, 16311	73G, 195, 204
GRX086	16086, 16224, 16311	73G, 93, 95, 152, 195
GRX068	16086, 16224, 16311	73G, 146, 152
GRX208	16224, 16234, 16261, 16311	73G, 146, 150
GRX258	16224, 16311	73G
GRX107	16224, 16311	73G, 146, 152, 324
GRX156	16224, 16311	73G, 152
GRX192	16224, 16311	73G, 195
<i>K2 (16093C, 16224C, 16311C)</i>		
GRX008	16093, 16224, 16311	73G
GRX151	16093, 16224, 16311	73G, 114
GRX025	16093, 16224, 16311	73G, 146

Table 2 (continued)

Haplogroup/cluster (defining motif)	HVSI	HVSII
GRX015	16093, 16224, 16311	73G, 146, 195
GRX202	16093, 16224, 16311	73G, 280
<i>K2b (16224C, 16311C, 16319A)</i>		
GRX238	16093, 16224, 16311, 16319	73G, 152, 199
GRX286	16224, 16311, 16319	73G, 152
<i>L1b (16126C, 16187T, 16189C, 16223T, 16270T, 16278T, 16311C, 152C, 182T, 185T, 195C, and 247A)</i>		
GRX116, GRX159	16126, 16187, 16189, 16223, 16264, 16270, 16278, 16293, 16311, 16362	73G, 152, 182, 185, 189, 195, 247, 263
GRX083	16126, 16187, 16189, 16223, 16264, 16270, 16278, 16311	73G, 152, 182, 185, 189, 195, 247
GRX226	16126, 16187, 16189, 16223, 16270, 16278, 16293, 16311	73G, 152, 182, 185, 195, 247
<i>L3d (16223T and 152C)</i>		
GRX269	16124, 16223	73G, 150, 152
GRX127, GRX175	16145, 16176, 16223	73G, 152
GRX112	16147, 16172, 16223, 16224, 16248, 16355, 16357	73G, 151, 152, 199, 204
GRX241	16179, 16223, 16243, 16256, 16284, 16311, 16320	73G, 152, 189, 194, 195
<i>M1 (16129A, 16189C, 16223T, 16249C, 16311C)</i>		
GRX109	16129, 16183, 16187.1, 16189, 16223, 16249, 16311, 16359	73G, 195, 215
<i>T (16126C and 16294T)</i>		
GRX230	16126, 16147, 16294, 16297, 16304	73G
GRX272	16126, 16223, 16294, 16296, 16304	73G, 199, 204
GRX022	16126, 16239, 16294, 16296, 16304	73G, 151
GRX173	16126, 16239, 16294, 16296, 16304, 16335	73G, 151
GRX233	16126, 16248, 16294	73G
GRX142	16126, 16292, 16294	73G, 146, 152, 279
GRX268	16126, 16292, 16294	73G, 146, 249, 279
GRX199	16126, 16292, 16294	73G, 146, 279
GRX012	16126, 16294, 16296, 16304	73G, 146, 151
GRX053	16126, 16294, 16296	73G
GRX252	16126, 16294, 16296, 16304	73G
GRX106	16126, 16294, 16296, 16304	73G, 151
GRX207	16126, 16294, 16296, 16304, 16311	73G, 152
GRX164	16126, 16294, 16296, 16324	73G
<i>T1 (16126C, 16189C/A, and 16294T)</i>		
GRX290	16104, 16126, 16163, 16187.1, 16189, 16294	73G, 152, 195
GRX177	16126, 16163, 16172, 16186, 16189, 16294, 16304	73G, 151, 152
GRX182	16126, 16163, 16186, 16189, 16294	73G, 146, 152, 195
GRX032	16126, 16163, 16186, 16189, 16294	73G, 150, 152, 195, 225
GRX054	16126, 16163, 16186, 16189, 16294	73G, 152, 195
GRX254	16126, 16163, 16187.1, 16189, 16294	73G, 152
GRX206	16126, 16186, 16189, 16294	73G
<i>T2 (16126C, 16153A and 16294T)</i>		
GRX062	16126, 16153, 16189, 16209, 16294	73G, 150
GRX048	16126, 16153, 16294	73G, 150
GRX080, GRX184	16126, 16153, 16294, 16296	73G, 150
GRX132	16126, 16153, 16294, 16296, 16304	73G, 146
<i>U (16192)</i>		
GRX050, GRX 161	16192	73G, 150
<i>U2 (16051G and 16129C)</i>		
GRX073	16051, 16126, 16129, 16183, 16186, 16189, 16362	73G, 152, 217, 340
GRX237	16051, 16129, 16182, 16183, 16189, 16362	73G, 152, 217
GRX070	16051, 16129, 16183, 16187.1, 16189, 16213, 16362	73G, 152, 217, 340
GRX016	16051, 16129, 16183, 16187.1, 16189, 16362	73G, 152, 217, 340
<i>U3 (16343G)</i>		
8 Samples [†]	16343	73G, 150
<i>U4 (16356C and 195C)</i>		
GRX043	16129, 16134, 16356	73G, 152, 195, 296
GRX013	16134, 16356	73G, 195
GRX305	16265, 16356, 16362	73G, 195, 247
GRX057	16356	73G, 195

(continued on next page)

Table 2 (continued)

Haplogroup/cluster (defining motif)	HVSI	HVSII
<i>U5 (16270T)</i>		
GRX060	16270	73G, 150
<i>U5a (16192T, 16270T)</i>		
GRX051	16067, 16192, 16270, 16304, 16311	73G, 150, 228
GRX197, GRX 312	16129, 16192, 16270, 16304	73G, 150, 228
GRX251	16167, 16192, 16270, 16311, 16356	73G, 150
GRX084	16172, 16189, 16192, 16270, 16311	73G, 150, 194
GRX121	16172, 16192, 16270	73G, 150, 195, 327
GRX131	16189, 16192, 16270	73G, 150, 152
GRX078	16189, 16192, 16270, 16293, 16311, 16336	73G, 150
GRX020	16189, 16192, 16270, 16320	73G, 150
GRX186	16192, 16249, 16270, 16304	73G, 150, 228, 251
GRX007	16192, 16270	73G, 150, 189, 195
GRX028	16192, 16270, 16304	73G, 150, 228
5 Samples ^g	16192, 16270, 16319	73G, 150
<i>U5a1 (16192T, 16256T, and 16270T)</i>		
GRX056	16192, 16256, 16270, 16291	73G
GRX094	16114, 16192, 16256, 16270, 16294	73G
GRX099	16192, 16256, 16270, 16362	73G
GRX124	16192, 16256, 16270, 16311	73G, 153, 228
GRX191	16192, 16256, 16270, 16298	73G
GRX219	16192, 16256, 16270, 16311	73G
<i>U5a1a (16256T and 16270T)</i>		
GRX158	16129, 16256, 16270, 16293	73G
GRX036, GRX 273	16227, 16256, 16270, 16316, 16320	73G, 152
GRX150	16256, 16270	73G
<i>U5b (16189C and 16270T)</i>		
GRX296	16172, 16185, 16189, 16270, 16274, 16311, 16325	73G, 150, 152, 195, 196
GRX170	16187.1, 16189, 16270	73G, 150, 217
<i>U6 (16172C and 16219G)</i>		
GRX075	16093, 16172, 16183, 16187.1, 16189, 16219, 16278	73G, 185
GRX190	16093, 16172, 16183, 16189, 16219, 16274, 16278	73G, 185
GRX179	16172, 16183, 16187.1, 16189, 16219, 16278, 16311	73G, 146, 152, 185, 188
GRX299	16172, 16219, 16278, 16290, 16311	73G, 235
<i>V (16298C)</i>		
GRX232	16075, 16189, 16298	73A, 200
GRX248	16172, 16298	73A, 150
10 Samples ^h	16298	73A
GRX260	16298	73A, 146
GRX118, GRX125	16298	73A, 195
GRX185	16298	73A, 195, 198
GRX165	16298	73A, 199
GRX035, GRX134	16298, 16311	73A
GRX262	16298, 16311	73A, 195
<i>W (16223T, 16292T, 195C, 204C, and 207A)</i>		
GRX100	16086, 16223, 16292	73G, 119, 189, 195, 204, 207
GRX143	16192, 16209, 16223, 16292, 16325	73G, 152, 189, 194, 195, 204, 207
GRX256	16192, 16223, 16292, 16325	73G, 143, 189, 194, 195, 204, 207
GRX149	16223, 16292	73G, 119, 189, 195, 204, 207, 214, 227
GRX180	16223, 16292, 16295	73G, 189, 195, 204, 207
GRX113	16223, 16292, 16318	73G, 195, 204, 207
GRX095	16223, 16292, 16320	73G, 189, 195, 204, 207
GRX101	16223, 16292, 16362	73G, 189, 194, 195, 204, 207
<i>X (16189C, 16223T, 16278T)</i>		
GRX089	16093, 16129, 16187.1, 16189, 16223, 16278, 16294, 16309,	73G, 143, 146, 152, 195
GRX040	16108, 16183, 16187.1, 16189, 16223, 16255, 16278	73G, 153, 195, 225, 227
GRX224	16108, 16184, 16185, 16189, 16223, 16255, 16038, 16278	73G, 153, 195, 225, 227
GRX071	16124, 16183, 16187.1, 16189, 16223, 16278, 16362	73G, 143, 152, 195, 225, 226
GRX210	16181, 16182, 16183, 16189	73G, 146, 153
GRX063, GRX223, GRX306	16182, 16183, 16189, 16223, 16278	73G, 146, 153, 256

Table 2 (continued)

Haplogroup/cluster (defining motif)	HVSI	HVSII
GRX047, GRX198	16183, 16187.1, 16189, 16223, 16278, 16355	73G, 153, 195, 225, 226
GRX283	16185, 16189, 16223, 16278,	73G, 146, 153, 195, 198, 225, 226
GRX188	16189, 16192, 16223, 16278, 16304	73G, 195, 207
<i>U/R</i>		
GRX174	16092, 16129, 16239	73G
GRX123	16180, 16189	73G
GRX259	16188.1, 16256, 16261	73G
GRX103	16224	73G, 152
GRX091	16278, 16311	73G
<i>A</i>		
GRX005	16223, 16290, 16319, 16362	73G, 146, 153, 235
<i>C</i>		
GRX052	16051, 16172, 16223, 16298, 16325, 16327	73G, 146, 194, 195, 249, 289, 290
<i>N9</i>		
GRX310	16169, 16192, 16195, 16223, 16243, 16261	73G, 150
<i>R</i>		
GRX183	16129, 16248, 16304	73G, 146, 152, 268
GRX227, GRX234	16182, 16183, 16189, 16234, 16284, 16304, 16311, 16324	73G, 195
<i>pre Y</i>		
GRX215	16126, 16231, 16318	73G, 146, 152

^a All individuals' haplotypes contained 263G and 315.1C.

^b GRX002, GRX014, GRX074, GRX081, GRX128, GRX108, GRX136, GRX187, GRX209, GRX213, GRX228, GRX231, FRX239, GRX243, GRX271, GRX274, GRX275, GRX277, GRX281, GRX285, GRX292, GRX297, GRX300, GRX303.

^c GRX019, GRX041, GRX069, GRX072, GRX120.

^d GRX031, GRX067, GRX225, GRX236, GRX244, GRX284.

^e GRX018, GRX117, GRX155, GRX307.

^f GRX017, GRX058, GRX090, GRX097, GRX154, GRX257, GRX263, GRX311.

^g GRX064, GRX096, GRX214, GRX221, GRX021.

^h GRX163, GRX145, GRX146, GRX102, GRX216, GRX222, GRX255, GRX261, GRX279, GRX298.

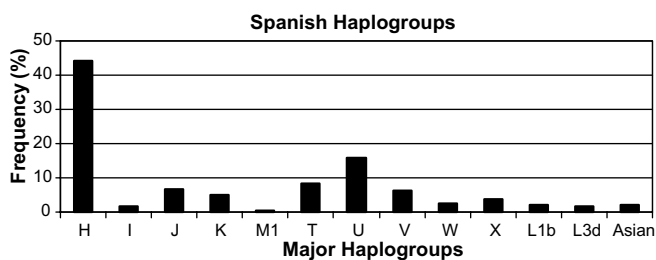


Fig. 2. Haplogroup frequencies observed in the Spanish data set ($n = 312$). Numbers and letters refer to the major haplogroups as defined in the text.

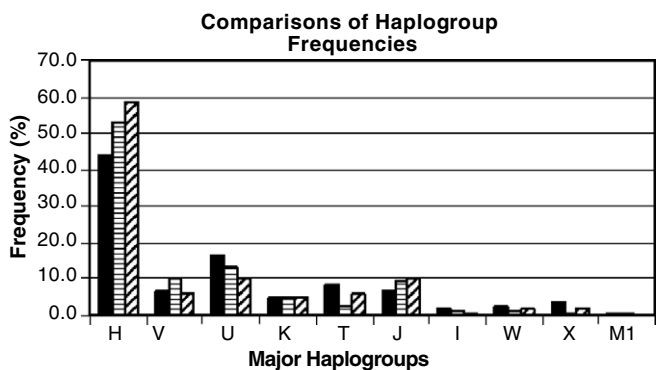


Fig. 3. European haplogroup frequencies for groups observed in this Spanish data set (black $n = 312$), the summation of Spanish haplogroups from Larruga, et al. [13] and Maca-Meyer, et al. [14] (horizontal lines) including all separate, non-repeated Spanish groups sampled, and the compilation of Iberian haplogroups from Helgason, et al. [12] from their Table 3 (diagonal). H is the combination of all H groups, HV, preHV, and U/R. T is the sum of T and T3, T4, and T5 haplogroups.

the J haplogroup. Two individuals (0.6%) were classified as J1a based on the additional SNPs of 16145A, 16231C, and 16261T. One individual (0.3%) was classified as J1b because the haplotype included 16145A and 16222T. Another individual (0.3%) had 16145A and 16222T and 16172C and was classified as a J1b1 haplogroup member. Other characters that subdivided the group further were 16278T, 152C, 185A, 188G, and 228A.

3.6. K haplogroup

The K haplogroup was another group with intermediate representation ($n = 15$, 4.8%). It was defined by 16224C and 16311C. Several sites divided this group further, 146, 152, and 195. The haplogroup was also subdivided by Helgason, et al. [12] into K2 defined by the presence of 16093C and K2b defined by 16319A. This Spanish sample contained five individuals with the K2 haplogroup and two individuals with the K2b haplogroup. Only one of the K2b individuals carried 16093C.

3.7. M1 haplogroup

Only one individual (0.3%) was found to belong to the M1 haplogroup. This group was defined by 16129A, 16189C, 16223T, 16249C, 16311C. Other studies of regions of Spain report this haplogroup as either missing or rare in samples [13,14,17].

3.8. *T* haplogroup

This is one of the more common haplogroups observed. Twenty-six individuals (8.3%) were classified in this group based on the presence of 16126C and 16294T. A separate subgroup ($n = 7$) was recognized based on 16189C and defined as T1. Helgason, et al. [12] include 16186T in the definition of T1, but the change at this site was absent from some of the T1 individuals in this Spanish data set. Five T2 sequences were recognized based on 16153A [12]. In this data set, 16296T, 16304C, and 151T also subdivided the group into separate clusters but each character was subject to reversals within the group.

3.9. *U* haplogroup

U is the second most common haplogroup in the study ($n = 51$, 16%). Helgason, et al. [12] defined a subgroup, U2, based on the presence of 16051G and 16129C and there were four (1.3%) individuals with those characters in the Spanish data set. We also found the additional characters 16189C, 16362C, and 217C in these individuals which are consistent with that described by Finnilä, et al. [11]. Based on the presence of characters 16343G and 150T in the Spanish samples, eight individuals (2.6%) belonged to the subgroup U3. Four individuals had 16356C and 195C and thus were classified as members of the U4 subgroup [11]. The U5 subgroup usually is divided further. One individual had only the distinguishing characteristic of the U5 group 16270T and could be designated as simply U5. In addition to 16270T, seventeen (5.4%) individuals also carried 16192T and were placed in the U5a cluster. This cluster was further subdivided by 16304C, 16319A, and 228A. The U5a1 cluster was defined by the presence of 16270T, 16192T, and 16256T and found in six individuals (1.9%). Four individuals (1.3%) were placed in the U5a1a cluster which contains 16270T and 16256T. Finally two individuals (0.6%) were identified as U5b based on the presence of 16189C in addition to the 16270T character. In this study, 150T also was present in all U5, U5a, and U5b individuals. The 150T substitution had previously been reported by Finnilä, et al. [11] for U5b. The final U cluster observed was U6 defined by 16172C and 16219G and containing four individuals (1.3%). In addition to the defined U clusters, two individuals were assigned to be members of the U haplogroup and five (1.3%) were defined as being U/R because they carried 73G but lacked other defining characters in the hypervariable regions.

Based on hypervariable regions I and II sequences only, the U clusters do not share a common character and branch off at different regions. Indeed even within the U5 group, the U5a1 and U5a1a clusters are placed separately from the other U5 clusters. It is possible that these groups do not share a common ancestor. However, others have cited common coding region changes between members of all U clusters ([12] and references therein) and additional

mitochondrial sequence data would be needed to establish the common traits for this group [29].

3.10. *V* haplogroup

This is another haplogroup of intermediate frequency ($n = 20$, 6.4%). The presence of 16298C characterizes the members of this group. The group is further subdivided by 16311C. All these individuals lack changes at 16,183 and 225 and do not meet Finnilä, et al.'s [11] criteria for classification as V1. As others have noted, V appears as a subgroup of haplogroup H.

3.11. *W* haplogroup

In these data, a cluster appeared that was characterized by 16223T, 16292T, 195C, 204C, and 207A and was classified as W with eight members (2.6%). Seven of the eight also had 189G. There has been some disagreement about which characters adequately define this group. Most authors [8,12,15,29] include 189G which was lacking from one of the individuals in this group. This may represent a rare reversion in the individual lacking the change. With respect to the other hypervariable region II changes, McCaulay, et al. [15] only include 204C and 207A in their definition, while Helgason, et al. [12] include 195C and 204C. However, based on the region I characters, W seems a reasonable assignment for the group. In this data set, further subdivisions occur due to 16192T, 16325C, 119C, and 194T.

3.12. *X* haplogroup

Twelve individuals (3.8%) shared the 16189C, 16223T, 16278T characters and were classified as X. While Macaulay et al. [15] make 16223T and 16278T the only defining characters, Helgason et al. [12] proposed that 16189C, 195C and 225A should be considered part of the definition. However, Allard, et al. [29] found that 195C was a variable character in the X haplogroup. In this Spanish cluster, 195C subdivides the group as do 146C, 153G, 225A, 226C, and 256T.

3.13. *The African haplogroups*

African haplogroups have been reported throughout many of the Western Mediterranean populations that have been sampled [17]. They have also been reported in low frequencies in regions of Spain, but not all regions sampled have contained the haplogroups found in this study [13,14]. It is difficult to know whether there are true regional differences in frequencies of the haplogroups or whether this reflects sampling. Overall nine individuals (2.9%) were placed in African groups.

3.14. *L1b* haplogroup

Four individuals (1.3%) were placed in the L1b haplogroup. In this study the group was defined by 16126C,

16187T, 16189C, 16223T, 16270T, 16278T, 16311C, 152C, 182T, 185T, 195C, and 247A. Three of the four individuals also had 16264T which is usually included as a characteristic of this haplogroup [27].

3.15. L3d haplogroup

This haplogroup occurs at a low frequency ($n = 5$, 1.6%), in these data and is characterized by 16223T and 152C. Although Salas, et al. [27] also indicate that 16124C is associated with the group, Allard, et al. [28] did not observe that association, but they also saw the 152C character.

3.16. Asian haplogroups

Four individuals were found with characters typically associated with Asian haplogroups based on characters described elsewhere [30]. These individuals include one each of haplogroups A, C, N9, and preY.

3.17. R haplogroup

Three individuals were placed in haplogroup R. Without coding region sequence or RFLPs, further classification is not possible.

4. Conclusions

The mtDNA lineages found in Spaniards were characterized predominantly by the commonly found European haplogroups. All of the major European haplogroups, H, I, J, K, T, U, V, W, and X were present. A 10th haplogroup also associated with European populations [9], M, was found in one individual and has been reported in regions of Spain and Portugal, but in low frequencies [13,14,17]. The most common haplogroup in Spain was H (42.3%) as is common in Europe. The other common haplogroups were J, K, T, U, and V all of which had frequencies of 4.8% or greater. Within the U haplogroup, the U5a cluster was particularly common at 5.4%. The SNPs that defined each haplogroup were determined and these agree well with published analyses of European haplogroups. In addition to the European haplogroups, African haplogroups and Asian haplogroups were also detected although at low frequencies (1.2–0.3%).

Haplogroup assignments have been made with extant data. One should recognize that additional sequence data, such as diagnostic coding-region mutations, could better resolve or clarify haplogroup assignments. Diagnostic coding region mutations may contradict HVI and HVII assignments on rare occasions. Many of the variable sites found in this study have been observed in other studies [11,12,26]. The character combinations used to define haplogroups in these data were compared to others and they allow further refinement of current haplogroup definitions. This information will be useful for forensic analysis of Spanish populations.

Acknowledgments

This research was supported in part by an appointment to the Visiting Scientists Program to MWA at the Federal Bureau of Investigation, Counterterrorism Forensic Science Research Unit administered by the Research Participation Program of the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the FBI-CTFSRU. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

References

- [1] Avise JC. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annu Rev Genet* 1991;25:45–69.
- [2] Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, et al. A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* 1997;15:363–8.
- [3] Meyer S, Weiss G, von Haeseler A. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 1999;152:1103–10.
- [4] Monson K, Miller K, Wilson M, DiZinno J, Budowle B. The mtDNA population database: an integrated software and database resource for forensic comparison. *Forensic Sci Commun* 2002;4:2.
- [5] Wilson MR, DiZinno JA, Polansky D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int J Legal Med* 1995;108:68–74.
- [6] Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 1999;23:147.
- [7] Crespillo M, Luque JA, Paredes M, Fernandez R, Ramirez E, Valverde JL. Mitochondrial DNA sequences for 118 individuals from northeastern Spain. *Int J Legal Med* 2000;114(1–2):130–2.
- [8] Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. *Nature* 1981;290:457–65.
- [9] Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, et al. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 1996;144:1835–50.
- [10] Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D. Human mitochondrial DNA variation and the origin of Basques. *Ann Hum Genet* 1995;59(Pt. 1):63–81.
- [11] Finnilä S, Lehtonen MS, Majamaa K. Phylogenetic network for European mtDNA. *Am J Hum Genet* 2001;68:1475–84.
- [12] Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R, et al. mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 2001;68:723–37.
- [13] Larruga JM, Diez F, Pinto FM, Flores C, Gonzalez AM. Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain. *Eur J Hum Genet* 2001;9:708–16.
- [14] Maca-Meyer N, Sanchez-Velasco P, Flores C, Larruga JM, Gonzalez AM, Oterino A, et al. Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Ann Hum Genet* 2003;67:329–39.
- [15] Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, et al. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 1999;64:232–49.
- [16] Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM. Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* 1996;60:321–30.

- [17] Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, et al. Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann Hum Genet* 2003;67:312–28.
- [18] Richards MB, Macaulay VA, Bandelt HJ, Sykes BC. Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 1998;62:241–60.
- [19] Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, et al. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* 1996;60(Pt. 4):331–50.
- [20] Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A. mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* 1998;6(4):365–75.
- [21] Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. Geographic patterns of mtDNA diversity in Europe. *Am J Hum Genet* 2000;66(1):262–78.
- [22] Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, et al. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 2000;67(5):1251–76.
- [23] Foran DR, Hixson JE, Brown WM. Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nucleic Acids Res* 1988;16(13):5841–61.
- [24] Wilson M, Allard MW, Monson KL, Miller K, Budowle B. Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. *Forensic Sci Int* 2002;129:35–42.
- [25] Wilson M, Allard MW, Monson KL, Miller K, Budowle B. Further discussions of the consistent treatment of length variants in the human mitochondrial DNA control region. *Forensic Sci Commun* 2002;4:4.
- [26] Schneider S, Roessli D, Excoffier L. Arlequin ver. 2000: a software for population genetic data analysis. Switzerland: Genetics and Biometry Laboratory, University of Geneva; 2000.
- [27] Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, et al. The making of the African mtDNA landscape. *Am J Hum Genet* 2002;71(5):1082–111.
- [28] Allard MW, Polansky D, Miller K, Wilson MR, Monson KL, Budowle B. Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. *Forensic Sci Int* 2005;148(2–3):169–79.
- [29] Allard MW, Miller K, Wilson M, Monson K, Budowle B. Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. Scientific working group on DNA analysis methods. *J Forensic Sci* 2002;47(6):1215–23.
- [30] Allard MW, Wilson MR, Monson KL, Budowle B. Control region sequences for East Asian individuals in the scientific working group on DNA analysis methods forensic mtDNA data set. *Leg Med (Tokyo)* 2004;6(1):11–24.

FOR THE RECORD

Luis J. Martínez-Gonzalez,¹ M.S.; Esther Martínez-Espin,¹ M.S.; Francisco Fernández-Rosado,¹ M.S.; Mayra A. Moguel,² M.S.; Carmen Entrala,¹ Ph.D.; J. Carlos Alvarez,¹ Ph.D.; Jose A. Lorente,¹ M.D., Ph.D.; and Bruce Budowle,³ Ph.D.

Mexican Population Data on Fifteen STR Loci (Identifiler[®] Kit) in a Chihuahua (North Central Mexico) Sample

POPULATION: Chihuahua, Northern Mexico, Mexico

KEYWORDS: forensic science, DNA typing, population genetics, Mexico, Identifiler

Sample preparation—Blood samples were obtained by venipuncture from unrelated individuals ($n = 161$) living in the State of Chihuahua, North Central Mexico, and spotted on FTA paper (Whatman, Clifton, NJ).

PCR—Approximately 1 ng of DNA were used in each amplification. The samples were amplified using the AmpF/STR[®] Identifiler[®] kit (Applied Biosystems, Foster City, CA) and the alleles were separated and detected using an Applied Biosystems ABI310 genetic analyzer.

Analysis of data—The frequency of each allele for each locus was calculated from the numbers of each genotype in the sample set (i.e., the gene count method). Unbiased estimates of expected heterozygosity were computed as described by Edwards et al. (1). Possible divergence from Hardy-Weinberg expectations (HWE) was tested by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies (1–4) and the exact test (5), based on 2000 shufflings experiments. An interclass correlation criterion (6) for two-locus associations was used for detecting disequilibrium between the STR loci. The program for this analysis was kindly provided by R. Chakraborty (University of Texas, School of Biomedical sciences, Houston Texas).

The distributions of the observed allele frequencies for the 15 STR loci are shown in Table 1. The most informative loci are D18S51 and FGA, and the least discriminating are TPOX and CSF1PO. The fifteen loci meet Hardy-Weinberg expectations. The combined power of discrimination is >0.99999999 , and the combined power of exclusion is 0.99999926. The complete data are

available either by accessing <http://www.gitad.org/poblaciones> or emailing the corresponding author.

Acknowledgment

This study was partially supported by the “Fundacion M. Botin”, Santander, Spain. We thank Applied Biosystems for kindly providing the Identifiler[®] kit.

References

1. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 1992;12:241–53.
2. Chakraborty R, Smouse PE, Neel JV. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. *Amer J Hum Genet* 1988;43:709–25.
3. Chakraborty R, Fornage M, Guegue R, Boerwinkle E. Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In Burke T, Dolf G, Jeffreys AJ, Wolff R, editors. *DNA fingerprinting: approaches and applications*. Berlin: Birkhauser Verlag 1991;127–43.
4. Nei M, Roychoudhury AK. Sampling variances of heterozygosity and genetic distance. *Genetics* 1974;76:379–90.
5. Guo SW, Thompson EA. Performing the exact test of Hardy Weinberg proportion for multiple alleles. *Biometrics* 1992;48:361–72.
6. Karlin S, Cameron EC, Williams PT. Sibling and parent-offspring, correlation estimation with variable family size. *Proc Natl Acad Sci USA* 1981;78:2664–8.

¹ Laboratory of Genetic Identification. Dept. de Medicina Legal. Facultad de Medicina. Universidad de Granada. E-18012 Granada, Spain.

² DNA Laboratory. D.G. Servicios Periciales. P.G.J. del Estado de Chihuahua. Chihuahua, Mexico.

³ FBI Laboratory Division. Quantico, VA 22135.

Additional information and reprint requests:
Prof. Dr. Jose A. Lorente
Department of Medicina Legal—University of Granada
Av. Madrid 11
E-18012 Granada
Spain

TABLE 1—Observed allele frequencies for the fifteen Identifiler® loci.

Allele	D3S1358	TH01	D21S11	D18S51	D2S1338	D5S818	D13S317	
2.2								
<5		0.00311						
5								
6		0.23913						
7		0.30435				0.05901		
8		0.06832				0.01553	0.06832	
9		0.16460				0.04037	0.22981	
9.3		0.21118						
10		0.00621		0.01863		0.05590	0.09627	
>10		0.00311						
11				0.02174		0.40994	0.26398	
<12	0.00621							
12	0.00311			0.13975		0.29814	0.18012	
13	0.00621			0.14286		0.11180	0.09627	
14	0.05280			0.13975		0.00621	0.06522	
15	0.31677			0.16460		0.00311		
16	0.26087			0.13665	0.03106			
17	0.16770			0.10248	0.16460			
18	0.14596			0.06522	0.04348			
19	0.04037			0.03727	0.23913			
20				0.01863	0.17391			
21				0.00621	0.02795			
22				0.00311	0.06211			
23				0.00311	0.14286			
24					0.05901			
25					0.04348			
25.2			0.00311					
26					0.01242			
27			0.01863					
28			0.09317					
29			0.20807					
30			0.29503					
30.2			0.02795					
31			0.05901					
31.1			0.00311					
31.2			0.13043					
32			0.00311					
32.2			0.09006					
32.3			0.00311					
33			0.00311					
33.2			0.04658					
34.2			0.00932					
34.3			0.00311					
35.2			0.00311					
Homozygosity test*	0.762	0.446	0.148	0.680	0.228	0.342	0.301	
Exact test**	0.017	0.459	0.116	0.824	0.972	0.073	0.012	
PD	0.91007291	0.90675514	0.94649126	0.96909841	0.95675321	0.85405656	0.93368311	
PE	0.57058241	0.55668406	0.66976540	0.75147290	0.70663845	0.49866824	0.64004831	
Allele	D7S820	D16S539	CSF1PO	D19S433	vWA	D8S1179	TPOX	FGA
2.2								
3.2								
5								
6							0.00311	
7	0.01863							
8	0.14596	0.03416	0.00311			0.00621	0.58385	
9	0.06832	0.11180	0.02484			0.00932	0.06832	
10	0.30124	0.20186	0.28261			0.07143	0.05280	
11	0.27019	0.21739	0.29193	0.00621	0.00311	0.05901	0.22671	
11.2				0.01242				
12	0.16460	0.27640	0.34783	0.03727		0.11801	0.06211	
12.2				0.01242				
13	0.02484	0.13975	0.04348	0.22360		0.30124	0.00311	
13.2				0.07764				
14	0.00621	0.01863	0.00621	0.31056	0.07143	0.29193		
14.2				0.02484				
15				0.16770	0.07453	0.11801		
15.2				0.06211				
16				0.02484	0.29503	0.02174		
16.2				0.03416				
17					0.30745	0.00311		0.00311

TABLE 1—Continued.

Allele	D7S820	D16S539	CSF1PO	D19S433	vWA	D8S1179	TPOX	FGA
17.2				0.00621				
18					0.16149			0.01242
19					0.07143			0.09006
20					0.01242			0.08385
21					0.00311			0.09938
22								0.10870
22.2								
23								0.17391
23.2								0.00311
24								0.15839
25								0.14596
26								0.09317
27								0.02484
28								0.00311
Homozygosity test*	0.800	0.916	0.157	0.845	0.382	0.428	0.239	0.377
Exact test**	0.523	0.448	0.672	0.072	0.057	0.648	0.705	0.268
PD	0.91709425	0.92789630	0.85112457	0.92889935	0.90289726	0.92805062	0.79086455	0.96801821
PE	0.57759331	0.60855433	0.45289261	0.63921864	0.57252133	0.59304848	0.36674141	0.75055161

PD = power of discrimination.

PE = power of exclusion.

* χ^2_{df} based on unbiased estimate with 2000 shufflings.

** Exact test based on 2000 shufflings.