

Procesos microevolutivos en la población romaní residente en el País Vasco

Autor:

Manuel Quintero Reguero

Memoria de Tesis doctoral dirigida por:

Dr. José Ángel Peña García

Dra. Esther M. Rebato Ochoa

Departamento de Genética, Antropología Física y Fisiología Animal

Leioa, 2022



Índice


Índice

N.Resumen	9
N.Abstract.....	11
I. Introducción	15
I.1. La población romaní.....	20
<i>Origen de la etnia</i>	21
<i>Díáspora de la población romaní</i>	22
I.2. Marcadores SNP Informativos de Ascendencia (AIM)	25
I.3 Elementos <i>Alu</i>	29
<i>Estructura de los elementos Alu</i>	30
<i>Origen y evolución de los elementos Alu</i>	31
<i>Retrotrasposición</i>	32
<i>Subfamilias de los elementos Alu</i>	33
<i>Función de las inserciones Alu en estudios genético-evolutivos</i>	36
<i>Elementos Alu analizados</i>	37
ACE	37
APO.....	38
B65.....	38
D1	39
FXIIIB.....	39
PV92	40
A25	40
TPA25	41
I.4 Marcadores STR pertenecientes al Combined DNA Index System (CODIS)	41
I.5 Marcadores de linajes masculinos y femeninos	47
<i>Marcadores del Cromosoma Y</i>	47
<i>Marcadores del ADN mitocondrial (ADNmt)</i>	53
II. Hipótesis de trabajo	61
III. Objetivos	61
III.1. Objetivo general.....	61
III.2. Objetivos específicos.....	62
IV. Materiales y Métodos	67

IV.1. Materiales	67
IV.1.1 Poblacionales analizadas.....	67
<i>Romaníes del País Vasco</i>	67
<i>Inserciones Alu y AIMs en otras poblaciones</i>	67
<i>Otros marcadores</i>	73
CODIS.....	73
Cromosoma Y y ADN mitocondrial.....	75
IV.2. Métodos.....	82
IV.2.1. Extracción del ADN.....	82
IV.2.2 Marcadores genéticos.....	82
<i>Inserciones Alu</i>	82
<i>Marcadores SNPs informativos de ascendencia (AIMs)</i>	87
<i>Otros marcadores genéticos</i>	88
CODIS.....	88
Cromosoma Y	88
ADNmt.....	89
IV.3. Análisis estadísticos	89
<i>Distancia genética de Reynolds</i>	90
<i>Distancia genética de Harpending y Jenkins</i>	90
<i>Análisis de escalamiento multidimensional (MDS)</i>	90
<i>Análisis factorial de correspondencias (AFC)</i>	91
<i>Método del centroide</i>	91
AMOVA.....	91
<i>Dendrograma</i>	92
<i>Análisis de clinas</i>	93
<i>Posible acción de la deriva: $-\log(P)$</i>	93
<i>Coefficientes de correlación de Pearson y Spearman</i>	94
V. Resultados.....	99
V.1. Análisis de la población romaní del País Vasco para AIMs	99
V.1.1. Frecuencias de 48 SNPs AIMs autosómicos.....	99
<i>Análisis molecular de la varianza (AMOVA)</i>	105
<i>Análisis de escalamiento multidimensional (MDS)</i>	107
<i>Análisis factorial de correspondencias (AFC)</i>	112
<i>Método del centroide</i>	115

<i>Análisis de clinas</i>	117
<i>Posible acción de la deriva genética</i>	121
<i>Estimación del grado de mestizaje</i>	124
V.2. Análisis de la población romaní del País Vasco para elementos <i>Alu</i>	125
V.2.1 Frecuencias de inserción de los 8 elementos <i>Alu</i>	125
<i>Análisis molecular de la varianza (AMOVA)</i>	127
<i>Análisis de escalamiento multidimensional (MDS)</i>	128
<i>Análisis factorial de correspondencias (AFC)</i>	133
<i>Método del centroide</i>	137
V.2.2. Análisis de 7 inserciones <i>Alu</i>	139
V.3. Análisis de la población romaní para CODIS	147
<i>Análisis molecular de la varianza (AMOVA) para los marcadores del sistema CODIS</i>	147
<i>Análisis de escalamiento multidimensional (MDS)</i>	148
<i>Análisis factorial de correspondencias (AFC)</i>	151
<i>Método del centroide</i>	152
<i>Análisis de clinas para el sistema CODIS</i>	153
<i>Posible acción de la deriva genética</i>	155
<i>Estimación del grado de mestizaje</i>	160
V.4. Análisis de las poblaciones romaníes europeas para marcadores de linaje.....	163
V.4.1. Análisis de la población romaní para haplogrupos del Cromosoma Y.....	163
<i>Análisis molecular de la varianza (AMOVA)</i>	163
<i>Análisis de escalamiento multidimensional (MDS)</i>	163
<i>Análisis factorial de correspondencias (AFC)</i>	166
<i>Método del centroide</i>	168
<i>Análisis de clinas para haplogrupos del cromosoma Y</i>	169
<i>Posible acción de la deriva genética</i>	172
<i>Estimación del grado de mestizaje</i>	174
V.4.2. Análisis de la población romaní para haplogrupos de ADN mitocondrial	176
<i>Análisis molecular de la varianza (AMOVA)</i>	176
<i>Análisis de escalamiento multidimensional (MDS)</i>	176
<i>Análisis factorial de correspondencias (AFC)</i>	179
<i>Método del centroide</i>	181
<i>Análisis de clinas de haplogrupos del ADNmt</i>	182
<i>Posible acción de la deriva genética</i>	184

<i>Estimación del grado de mestizaje</i>	186
VI. Discusión	191
VI.1 SNPs de ascendencia (AIMs)	192
VI.2. Inserciones <i>Alu</i>	197
VI.3 CODIS	200
VI.4 Marcadores de linaje	206
<i>Análisis con haplogrupos de cromosoma Y</i>	207
<i>Análisis con haplogrupos de ADNmt</i>	210
VI.5 Recapitulación.....	213
VII. Conclusiones	219
VIII. Bibliografía	223
IX. Anexo.....	261



Resumen y Abstract

N.Resumen

La presente Tesis Doctoral ha tratado de evaluar el efecto de los procesos de flujo génico y el impacto de la deriva sobre la diversidad genética de una población romaní residente en el País Vasco, por medio de una serie de análisis de un grupo de marcadores SNPs informativos de ascendencia (AIMs), y un segundo grupo de inserciones *Alu* autosómicas. Además, se han analizado otras poblaciones romaníes europeas a partir de una serie de bases de datos de STRs autosómicos pertenecientes al sistema CODIS, y marcadores de linaje, con el objetivo de extrapolar los resultados a la población romaní objeto de estudio. Igualmente, y en función de los marcadores genéticos utilizados y aquellos disponibles en otros estudios y bases de datos, se han considerado diversas poblaciones europeas que albergan en la actualidad a diversos grupos de origen romaní, poblaciones asiáticas de distintos países, incluida la India, así como de China (los Han), y un grupo africano, los ¡Kung del Sur de Angola. Estos datos poblacionales nos han permitido situar en un determinado contexto geográfico y genético a la población objeto de estudio y analizar, la posible relación entre el porcentaje de población romaní asentada en cada territorio y el grado de mestizaje estimado.

La etnia romaní muestra diversas particularidades tanto históricas (su vasto periplo migratorio desde el subcontinente indio y su asentamiento mayoritariamente a lo largo del continente europeo), como socioculturales (por ejemplo, su estructuración en clanes, su preferencia por los enlaces intra-grupales o su discordancia jerárquica entre los sexos), que posibilitan los estudios sobre el efecto de los agentes microevolutivos y permiten analizar de manera independiente cada linaje. La biografía de este grupo étnico permite observar los efectos de sucesivos eventos fundadores y cuellos de botella en su diversidad genética, y las consecuencias derivadas de posibles flujos génicos con las diferentes poblaciones huésped.

Los resultados obtenidos indican que la población romaní residente en el País Vasco ha estado sometida a períodos de aislamiento y tamaños efectivos reducidos durante su diáspora, como consecuencia de efectos fundadores y cuellos de botella sucesivos, que han propiciado la acción de la deriva genética, encontrándose evidencias de la misma en un gran número de los SNPs y de las inserciones *Alu* analizados. Asimismo, se ha observado un notable impacto del flujo génico en esta población procedente tanto de su población huésped como de otras europeas y, eventualmente, del Próximo Oriente. Este elevado componente europeo se ha detectado en otras poblaciones romaníes europeas, con un amplio rango de variación, situándose la población romaní del País Vasco en un valor intermedio. Respecto a otras poblaciones europeas de etnia romaní, analizadas mediante marcadores STRs y haplogrupos del cromosoma Y así como del

ADNmt, se ha observado una notable variabilidad en las acciones de los agentes microevolutivos. Dicha heterogeneidad pone de manifiesto las diferentes condiciones a las que han estado expuestas y el escaso contacto entre ellas, lo que ha generado posiblemente una subestructuración posterior a su entrada en Europa. También se ha observado un mestizaje diferencial por sexos en las poblaciones romaníes analizadas, con mayores indicios de mestizaje en los marcadores del linaje femenino respecto al masculino. En consecuencia, el flujo génico observado mediante el uso de marcadores autosómicos correspondería en gran parte a la entrada de mujeres de otras poblaciones externas a esta etnia dentro de los grupos romaníes. La dominancia jerárquica de los varones frente a las mujeres, sobre todo en el caso de los enlaces matrimoniales, sería más permisible con los varones en cuanto a las uniones interpoblacionales (con mujeres de otro origen), pero no en sentido contrario, lo que se ha podido traducir en una mayor entrada de haplogrupos del ADNmt con origen externo que de haplogrupos del cromosoma Y.

Los análisis con ambos linajes han contribuido asimismo a explicar el origen indio de esta etnia, pero a su vez han mostrado discordancias en cuanto a la región o a un área geográfica concreta. Por un lado, los análisis del linaje masculino han situado su origen en el Sur del subcontinente indio debido a la presencia del haplogrupo H, mayoritario en las poblaciones romaníes y en las tribus del Sur de la India. Este haplogrupo está sin embargo ausente tanto en las poblaciones europeas como en las del Punjab (en el Noroeste de la India); precisamente esta región se considera como el origen más probable de la etnia romaní, hecho observado en numerosos estudios y sustentado por los resultados obtenidos en esta investigación mediante el análisis de haplogrupos del ADNmt. Dichos resultados podrían ser un indicio de eventos de cuello de botella o de efecto fundador posteriores al inicio de la diáspora romaní, que habrían provocado la desaparición del haplogrupo H del acervo genético de las poblaciones del Noroeste de la India, aún manteniéndose en las tribus y castas del Sur. La permanencia de este haplogrupo en el Sur podría deberse también al aislamiento entre ambos extremos del subcontinente Indio, provocado por la estructuración en castas y la consecuente barrera al flujo génico, lo que generó un gradiente Sur-Norte de la presencia de dicho haplogrupo.

N.Abstract

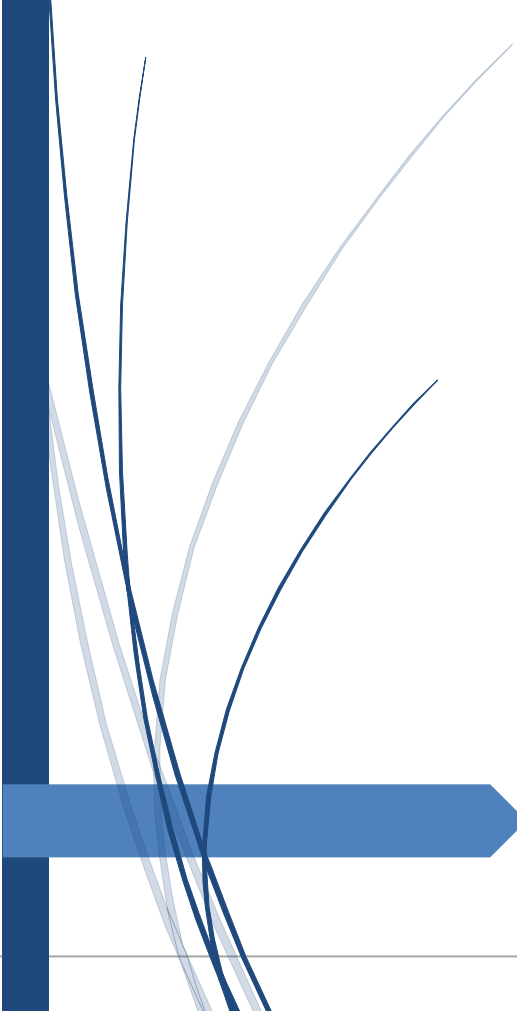
This PhD Thesis has attempted to evaluate the effect of gene flow processes and the impact of drift on the genetic diversity of a Roma population residing in the Basque Country, through a series of analyses of a group of ancestry informative SNP markers (AIMs), and a second group of autosomal Alu insertions. In addition, other European Roma populations have been analysed from a series of databases of autosomal STRs belonging to the CODIS system, as well as lineage markers, with the aim of extrapolating these results to the Roma population under study. Likewise, and depending on the genetic markers used, along with those available in other studies and databases, we have taken into consideration various European populations that are currently home to various groups of Roma origin, Asian populations from different countries, including India, as well as from China (the Han), and an African group, the iKung from Southern Angola. These population data have allowed us to place the population under study in a specific geographical and genetic context and to analyse the possible relationship between the percentage of Roma settled in each territory and the estimated degree of interbreeding.

The Roma ethnic group shows various particularities, both historical (its vast migratory journey from the Indian subcontinent and its settlement mostly along the European continent) and socio-cultural (for example, its clan structure, its preference for intra-group links or its hierarchical discordance between the sexes), which make it possible to study the effect of microevolutionary agents and allow each lineage to be analysed independently. The biography of this ethnic group makes it possible to observe the effects of successive founding events and bottlenecks on its genetic diversity, as well as the consequences of possible gene flows with the different host populations.

The results obtained indicate that the Roma population residing in the Basque Country has been subjected to periods of isolation and reduced effective sizes during its diaspora, as a consequence of successive founder effects and bottlenecks, which have favoured the action of genetic drift, evidence of which has been found in a large number of the SNPs and Alu insertions analysed. Likewise, a notable impact of gene flow has been observed in this population, both from its host population and from other European and possibly Middle Eastern populations. This high European component has been detected in other European Roma populations, with a wide range of variation, placing the Roma population of the Basque Country at an intermediate value. With respect to other European Roma populations, analysed using STRs and Y-chromosome haplogroups as well as mtDNA markers, a remarkable variability in the actions of

microevolutionary agents has been observed. This heterogeneity highlights the different conditions to which they have been exposed and the scarce contact between them, which has possibly generated a substructured population after their entry into Europe. Differential interbreeding by sex has also been observed in the Roma populations analysed, with greater evidence of interbreeding in the markers of the female lineage compared to the male. Consequently, the gene flow observed using autosomal markers would largely correspond to the entry of women from other non-Roma populations into Roma groups. The hierarchical dominance of males over females, especially in the case of marital links, would be more permissible with males in terms of inter-population unions (with women of other origin), but not in the opposite direction, which could have been translated into a greater entry of mtDNA haplogroups of external origin than of Y-chromosome haplogroups.

Analyses with both lineages have also contributed to explaining the Indian origin of this ethnic group, but have also shown discordance in terms of the region or a specific geographical area. On the one hand, analyses of the male lineage have placed its origin in the south of the Indian subcontinent due to the presence of haplogroup H, which is predominant in the Roma populations and tribes of southern India. This haplogroup is, however, absent in both European and Punjabi populations (in North-West India); this region is considered to be the most probable origin of the Roma, a fact observed in numerous studies and supported by the results obtained in this research through mtDNA haplogroup analysis. These results could be an indication of bottleneck or founder effect events after the onset of the Roma diaspora, which would have led to the disappearance of haplogroup H from the gene pool of the populations of north-western India, while still remaining in the tribes and castes of the South. The permanence of this haplogroup in the South could also be due to the isolation between the two extremes of the Indian subcontinent, caused by caste structuring and the consequent barrier to gene flow, which generated a South-North gradient in the presence of this haplogroup.



Introducción

I. Introducción

EL origen de los humanos anatómicamente modernos se sitúa entre los 300.000-200.000 años, aunque aún existe discusión sobre la exactitud de esta fecha y su rango de variación (Stringer, 2016; Weaver, 2012). Algunos autores se basan en que el último ancestro común compartido con nuestra especie hermana, *Homo neanderthalensis*, data de hace unos 400.000 años, iniciándose un periodo en el que no había una única línea de progresión en la morfología de los *Homo sapiens* arcaicos, pudiéndose superponer, a nivel cronológico, morfologías “arcaicas” y “modernas”, siendo difícil determinar un “punto” en el tiempo como origen de nuestra especie (Stringer, 2016; Weaver, 2012). En lo que si hay un gran consenso es en que el origen de *Homo sapiens* se emplaza en el continente africano como señalan los fósiles, entre otros, los de Omo Kibish en el valle del río Omo al Sur de Etiopía (Foley et al., 2016; Leakey, 1969). Posteriormente, hace unos 200.000 años, los *Homo sapiens* migraron desde África extendiéndose por Eurasia (Harvati et al., 2019); este evento no ocurrió en una única oleada migratoria, sino en varias, siendo la más reciente hace unos 70.000 años, como ha señalado Wallace (2013) al analizar el registro de haplotipos del ADNmt. Desde ese momento el tamaño poblacional de la especie humana experimentó un aumento progresivo, salvo durante el último periodo glacial, donde el número de individuos se redujo considerablemente (principalmente por motivos climáticos) en una serie de zonas geográficas como el refugio franco-cantábrico (Achilli et al., 2004)

Durante buena parte de nuestra historia evolutiva (Paleolítico) hemos sido una especie nómada, cazadora-recolectora, pero hace unos 13.000 años (fecha variable según el área geográfica y las distintas evidencias registradas), durante la Revolución del Neolítico (Thaler, 2016; Zeder, 2008), surgieron la agricultura, la domesticación, los primeros asentamientos y, por lo tanto, las primeras poblaciones humanas instaladas de manera permanente en un territorio (paso de nomadismo a sedentarismo). La revolución neolítica supuso una gran transformación en la forma de vida de la humanidad, lo cual, junto con los procesos migratorios previos y sus consecuentes efectos fundadores, además de los cuellos de botella que sin duda tuvieron lugar durante las glaciaciones de Riss (hace unos 200.000 años) o la de Würm (hace unos 110.000), jugaron un papel fundamental en la variabilidad genética humana actual. Dicha variabilidad es muy baja si se compara con otras especies de primates (Barbujani y Colonna, 2010), siendo la similitud genética entre dos individuos escogidos al azar superior al 99% (Glazko et al., 2005); esta escasa variabilidad puede ser explicada como resultado de un corto y reciente periodo de

evolución de nuestra especie y un bajo tamaño efectivo de las poblaciones fundadoras, cifrado inicialmente en unos 10.000 individuos, aunque también se barajan valores más reducidos (Fagundes et al., 2007; Hawks et al., 2000). El tamaño poblacional inicial podría justificarse tras eventos de cuello de botella acaecidos desde la divergencia del ancestro común compartido con el *Homo neanderthalensis* (Hawks et al., 2000), como por ejemplo los acontecidos durante las glaciaciones anteriormente citadas.

El estudio de la variabilidad genética humana es clave para entender el devenir de nuestra especie, para lo cual se necesitan conocimientos que parten de disciplinas tales como la Antropología Física (o Biológica), la Genética, la Estadística o las Matemáticas, entre otras, y el uso de diferentes herramientas como, por ejemplo, el denominado índice de fijación (F_{ST}), una medida de la diferenciación de la población debido a la estructura genética. Dos de las definiciones más comúnmente usadas para el índice F_{ST} en un locus dado se basan en la varianza de las frecuencias alélicas entre las poblaciones y en la probabilidad de “identidad por descendencia”. El desarrollo de este índice, que será usado a lo largo de esta Tesis Doctoral, se debe al genetista estadounidense Sewall Wright (1889-1988) que lo publicó en 1951 (también se denomina “índice de fijación de Wright”) y lo consideró como una medida para valorar las diferencias genéticas dentro de las poblaciones y entre ellas, siendo por tanto una medida de la variabilidad genética tanto intra como inter-poblacional (Wright, 1951; Sherier et al., 2021). Posteriormente, en 1983, el grupo de investigación de John Reynolds concluyó que este valor aumentaba de manera constante desde el tiempo de divergencia de dos especies y por lo tanto permitía la estimación temporal de ese periodo (Reynolds et al., 1983).

El índice de fijación oscila desde un valor 0, lo que indica que la población no está estructurada (la estructura genética de una población viene determinada por su historia evolutiva y expresa la cantidad de diversidad genética que alberga y cómo se reparte dentro de la población), a un valor 1 que significa que para cada locus se ha fijado un alelo en cada población. En definitiva, este índice representa la probabilidad de que al tomar un alelo aleatorio de una población este sea idéntico por descendencia respecto a una población ancestral (Ochoa et al., 2021). Nuestra especie presenta un rango de valores medios de F_{ST} que van desde 0,052 según una base de datos de más de 400.000 SNPs en casi 4000 individuos de 4 continentes (Auton et al., 2009), a 0,13, según un estudio de 600.000 SNPs de 4 poblaciones caucásicas, chinas, japonesas y yorubas (Weir et al., 2005). Estos valores quedarían algo alejados de los observados en otras especies de primates cercanos filogenéticamente a nuestra especie, como el gorila, con un valor de 0,38 (Thalmann et al., 2007). Este índice es también muy relevante

debido a que marcadores genéticos con altos valores se utilizan como informativos de ascendencia (AIMs) (Kidd et al., 2014). Esto se debe a la acumulación de mutaciones neutras que generan alelos específicos para las poblaciones o diferencias significativas entre ellas para sus frecuencias (Mao et al., 2007; Tian et al., 2006, 2007), lo cual permite caracterizar poblaciones, por ejemplo, mediante la estimación de las tasas de mestizaje, o identificar su origen (Tvedebrink y Eriksen, 2019). En definitiva, el análisis de las variaciones genéticas de las poblaciones permite una mejor comprensión de nuestra historia evolutiva y de los agentes que han intervenido en ella.

La evolución de nuestra especie se puede analizar desde un nivel poblacional dentro de la especie, es decir, en un marco microevolutivo, o a nivel de formación de una segunda especie que constituiría el llamado marco macroevolutivo (Reznick y Ricklefs, 2009). Los agentes implicados en el proceso evolutivo serían la selección natural, la mutación, el flujo génico y la deriva genética. Los dos primeros tendrían una mayor relevancia a nivel macroevolutivo y los otros dos en el microevolutivo (Arnold et al., 2001; Relethford, 2012). Son precisamente estas dos últimas fuerzas evolutivas las que presentan una mayor importancia a la hora de estudiar la diversidad genética de las poblaciones. El flujo génico consiste en el intercambio de genes entre poblaciones siendo por lo tanto una fuerza homogeneizadora que tiende a reducir las diferencias genéticas entre poblaciones (Slatkin, 1987); es decir, puede hacer que poblaciones distantes sean similares genéticamente entre sí. Este flujo dependerá tanto del tamaño efectivo de las poblaciones (N_e) (número de individuos dentro de una población con capacidad reproductora en cada generación), como de la tasa de migración (m) entre ellas (Hansen et al., 2007; Tigano y Friesen, 2016). Como consecuencia, la posibilidad de que dos poblaciones se unifiquen o mantengan sus estructuras genéticas de manera autónoma será dependiente de la intensidad del flujo génico entre ambas (Bolnick y Nosil, 2007).

Como fuerza evolutiva de efecto contrario se encuentra la deriva genética, que corresponde a los cambios aleatorios en la frecuencia de un alelo en una población debido al azar (Gould, 2002). La deriva puede provocar tanto que un alelo se afiance en una población, como que las frecuencias de un alelo se vuelvan mínimas, pero estables, promoviendo muchas veces la fijación de variantes genéticas poco comunes (Masel, 2011). Una acumulación de dichas variantes a lo largo del tiempo promovería la diferenciación de la población que estuviera sujeta al efecto de la deriva de manera intensa. Como en el caso del flujo génico, el efecto de la deriva genética es dependiente del tamaño efectivo de la población (N_e), debido a que las poblaciones pequeñas están más expuestas a fluctuaciones aleatorias en sus frecuencias alélicas que las

poblaciones de mayor tamaño, que pueden (y suelen) mantener niveles más altos de variabilidad (Lynch et al., 2016; Wang et al., 2020).

La presente Tesis Doctoral ha tratado de evaluar el efecto de los procesos de flujo génico y el impacto de la deriva sobre la diversidad genética de una población romaní del País Vasco, por medio de una serie de análisis de un grupo de marcadores SNPs (polimorfismos de un solo nucleótido) informativos de ascendencia o de ancestría (AIMs), y un segundo grupo de inserciones *Alu*. Precisamente, los polimorfismos de inserción *Alu* son unos marcadores genéticos muy interesantes para los estudios de evolución humana, ya que se producen por un único evento mutacional y se conoce su estado ancestral. Además, se han analizado otro grupo de poblaciones romaníes europeas a partir de una serie de bases de datos de marcadores de linaje y STRs pertenecientes al sistema CODIS, con el objetivo de extrapolar los resultados a la población romaní objeto de estudio.

Las particularidades culturales e históricas de la etnia romaní la dotan de un gran interés para los estudios de Genética de poblaciones. Culturalmente, la etnia romaní promueve los matrimonios dentro de su mismo grupo y a su vez posiciona en diferente jerarquía social a los varones y a las mujeres (Kalaydjieva et al., 2001a). Dichos comportamientos, junto a un tamaño poblacional reducido, posibilitan los estudios sobre aislamiento y/o deriva genética, y permiten analizar de manera independiente cada linaje. Por otro lado, desde el punto de vista histórico, los romaníes son el paradigma de una etnia nómada con origen en el noroeste del subcontinente indio (Melegh et al., 2017), del que salieron mediante una serie de oleadas migratorias a lo largo de varios siglos, principalmente hacia el continente europeo (Gusmão et al., 2010). Este contexto permite observar los efectos de sucesivos eventos fundadores en la diversidad genética de este grupo étnico, junto a las consecuencias derivadas de posibles flujos génicos con las diferentes poblaciones receptoras.

Aunque se describirá de forma técnica en el Apartado de Material y Métodos, queremos avanzar tanto aquí como en los párrafos 1.2 a 1.5, la descripción de las propiedades de los marcadores empleados en esta Tesis Doctoral y, por ende, la justificación de su uso para la consecución de los objetivos de este estudio. Inicialmente se seleccionaron 48 SNPs con alto poder discriminante de ascendencia (AIMs). Los SNPs son, en su mayoría, marcadores bialélicos con tasas bajas de mutación, que constituyen la forma de variación genética más abundante en nuestro genoma (Chen et al., 2016; Hu et al., 2017; 1000 Genomes Project Consortium et al., 2015); además, debido a que los SNPs no cambian mucho de una generación a otra (se heredan de forma muy estable), es sencillo seguir su evolución en estudios poblacionales. Su distribución

a lo largo del genoma no es homogénea, y se barajan diferentes hipótesis para explicar este hecho, entre ellas la presencia o ausencia de microsatélites (Varela et al, 2010). Los SNPs seleccionados en esta investigación fueron aquellos que presentaban frecuencias significativamente diferentes entre poblaciones de varias localizaciones geográficas (Santos et al., 2016), lo que nos permitirá analizar el grado de mestizaje de la población romaní objeto de estudio con la población huésped española general y con otras del País Vasco, así como el grado de similitud con su población ancestral en la India y la distancia genética con otras poblaciones.

Además, se seleccionaron 8 inserciones *Alu*. Dichas inserciones son marcadores bialélicos que conforman más del 10% del genoma humano (Aktas et al., 2017), siendo su origen más probable el gen 7SL RNA (Gal-Mark et al., 2008). Esta familia de marcadores presenta una estructura heterodimérica, compuesta por dos monómeros relacionados pero no equivalentes, de una longitud total aproximada de 300 pb (Häsler y Strub, 2006). Son de gran utilidad para los estudios de genética de poblaciones debido a que dichas inserciones se consideran idénticas por descendencia. Esto significa que la probabilidad de que dos individuos no emparentados presenten la misma inserción en un lugar determinado es prácticamente nula (Batzer et al., 2002; Novick et al., 1998; Salem et al., 2005). Por otro lado, son marcadores selectivamente neutros por lo que sus frecuencias están afectadas únicamente por el efecto de la deriva y el flujo génico (Comas et al., 2000). También presentan una alta estabilidad, siendo muy poco probable que sufran una delección y, aún más improbable, que ésta sea completa (Batzer et al., 1994, 2002; Novick et al., 1998; Salem et al., 2005). Como consecuencia de todo lo anterior, las inserciones *Alu* tienen un estado ancestral conocido que viene dado por la ausencia de inserción; esta característica permite generar modelos evolutivos añadiendo una hipotética población ancestral, es decir, con ausencia total de las inserciones analizadas (Batzer et al., 1994, 2002; Novick et al., 1998; Salem et al., 2005).

Por último, se tomaron frecuencias de marcadores de linaje del cromosoma Y y del ADNmt, a partir de una serie de bases de datos y artículos científicos (ver Tabla 1 del Anexo). Esto es de especial interés debido a que la diferente posición social en que se encuentran varones y mujeres dentro de la etnia romaní podría verse reflejada en la diversidad genética de cada linaje. Además, las personas estudiadas, debido a que sólo presentan un alelo por individuo, están sujetas a un mayor efecto de la deriva. Por otro lado, también se tomaron frecuencias de STRs (*Short Tandem Repeats*) que son repeticiones consecutivas cortas (entre 2-7 pb) de número variable, que conforman cerca del 3% del genoma humano (Butler y John, 2011; Fan et al., 2007). Estos marcadores, junto a los marcadores de linaje, presentan altas tasas de

mutación (Cabrera, 2021; Fan et al., 2007; Helgason et al., 2015). Las elevadas tasas de mutación de los STRs junto a su localización en el genoma, mayoritariamente en regiones no codificantes (Ellegren, 2000; Fan et al., 2007), puede promover la acumulación de mutaciones aumentando la variabilidad de repeticiones consecutivas entre individuos. Como consecuencia, tendrían un alto valor como marcadores identificativos de individuos.

En resumen, y avanzando ya parte de los objetivos, mediante esta investigación se han tratado de valorar los efectos de los agentes de microevolución sobre la diversidad genética de una población romaní del País Vasco, considerando la posibilidad de que sus características socio-culturales que propician aún en la actualidad un cierto grado de endogamia (Bianco et al., 2020), hayan sido una barrera al flujo génico (Melegh et al., 2017), considerando además la discrepancia jerárquica entre los linajes masculino y femenino (Kalaydjieva et al., 2001a).

I.1. La población romaní

El término población gitana o romaní hace referencia a un grupo étnico que ha experimentado uno de los fenómenos de diáspora humana recientes más importante y a la vez menos documentados (Martínez-Cruz et al., 2016). Este proceso es asimismo de gran interés para evaluar las repercusiones de una serie de migraciones y eventos fundadores recientes, múltiples y generalizados, sobre el patrimonio genético de una población humana tan relevante en el ámbito de la Antropología Física y de la Antropología social y cultural. Sus características sociales y diversos factores políticos suelen dificultar la obtención de datos precisos sobre esta etnia, pero se les reconoce como la minoría étnica más numerosa de Europa, con una población de hasta 10 millones de personas repartidas por todo el continente, que se concentra principalmente en el centro y sudeste de Europa (Martínez-Cruz et al., 2016).

La sociedad romaní está conformada por una estructura compleja, donde el Clan (grupo de personas unidas por lazos de parentesco y ascendencia) es la unidad principal. Esta característica ha llamado la atención durante mucho tiempo a los antropólogos que describen la organización de los romaníes actuales como un "mosaico fluido de grupos diversificados" (Fraser, 1992; Marushiakova y Popov, 1997; Petulengro, 1915-16; Liégeois, 1994). Esto se refiere a que la etnia presenta una población fundadora fragmentada, formada por un gran número de subgrupos endogámicos socialmente diversos, y localizados a lo largo del continente europeo (Gusmão et al., 2010). La identidad de grupo y las divisiones sociales resultantes se basan en una variedad de criterios, como costumbres, etnónimos (nombre que reciben distintos grupos étnicos y que describen oficios tradicionales) y dialectos, que reflejan la historia de las migraciones; sin embargo, el término romaní se emplea como nombre que engloba a todos,

pese a que, técnicamente, hace referencia únicamente a los gitanos de Europa del Este (Kalaydjieva et al., 2005). De hecho, el término gitano es mayoritario en español y ellos se hacen llamar así, en sentido positivo, aunque existen muchas veces connotaciones peyorativas por parte de personas que no pertenecen a esta etnia. A escala internacional, sobre todo en España, donde también se usa el término “calé” o gitanos ibéricos, existe una propuesta común para utilizar *rom*, tanto como nombre del pueblo como del idioma, aunque dicha propuesta está aún en discusión.

En los romaníes puede hablarse de un efecto fundador combinado: un proceso inicial que afectó a la totalidad de la población romaní, y de diversos microprocesos secundarios que generaron posteriores subgrupos poblacionales que se dispersaron a través del tiempo y del espacio. La mayor diversidad parece encontrarse en los Balcanes, donde existen numerosas poblaciones romaníes con fronteras sociales bien definidas (Marushiakova y Popov, 1997, 2001). Estudios genéticos relativamente recientes han mostrado que las mutaciones fundadoras pueden ser compartidas por poblaciones romaníes socialmente diversas y geográficamente dispersas y, al mismo tiempo, aquellas que viven en una proximidad geográfica pueden mostrar frecuencias genéticas diferentes (Kalaydjieva et al., 2001b). Por tanto, las prácticas sociales, así como los datos genéticos, sugieren una importante subestructura en los romaníes. Sin embargo, la relación entre las divisiones de grupos tradicionales y las afinidades biológicas no está clara y es muy compleja. Los actuales patrones genéticos, al igual que los sociales, podrían ser el producto de diversos escenarios, con diferentes implicaciones para la comprensión del flujo génico.

Origen de la etnia

Los romaníes, como ya se ha señalado, se consideran la minoría étnica más grande de Europa; establecer un valor fijo respecto a su tamaño efectivo global es un proceso de gran complejidad, ya que este grupo étnico carece habitualmente de crónicas o investigaciones históricas exhaustivas que puedan aportar información detallada sobre su pasado (Kalaydjieva et al., 2005). A pesar de esta problemática, los análisis lingüísticos y antropogenéticos han permitido establecer la raíz de este pueblo en el subcontinente indio (Chaix et al., 2004; Fraser, 1992; Gresham et al., 2001; Hancock, 2000; Kalaydjieva et al., 2001b; Marushiakova y Popov, 1997; Morar et al., 2004, 2005; Melegh et al., 2017). Sin embargo, aunque el origen indio de las poblaciones romaníes europeas está lingüística y genéticamente bien establecido, la identificación del lugar geográfico preciso ha sido una cuestión de debate. Algunos estudios lingüísticos sostienen que la población fundadora “proto-romaní” se situaba en el noroeste de

la India (Fraser et al., 1992; Hancock et al., 2002; Iovita et al., 2004), aunque sus propios “mitos de origen” sugieren que su cuna se emplazaba en la llanura indo-gangética (Chaubey et al., 2010; Moreau et al., 2002). Por otro lado, los análisis sobre las castas y el comportamiento endogámico presentes en la cultura romaní muestran una gran similitud con la cultura india. Estas costumbres limitan el matrimonio a únicamente entre miembros de los propios clanes romaníes, y por lo tanto entre emparentados (Melegh et al., 2017). Este origen parece confirmarse en un gran número de estudios realizados sobre marcadores genéticos clásicos y del ADNmt, que sugieren una afinidad cercana de los romaníes con los Rajpust y Punjabís, poblaciones de Rajasthan y Punjab, respectivamente, respecto a otras poblaciones de la India (Figura 1) (Mastana et al., 1992; Mendizábal et al., 2011).



Figura 1: Posible origen geográfico de la población romaní.

Díaspóra de la población romaní

Actualmente se desconocen las razones principales que desencadenaron la migración de la población romaní, pero existe un cierto consenso en la hipótesis que explica que fue promovida por la invasión del norte de la India por los persas (S.VII) debido a la ausencia de vocablos árabes en la lengua romaní (Fraser, 1992; Gusmão et al., 2010; Rochow y Matschke, 1991). Posteriormente migraron a Irán y Armenia, donde fueron desplazados hacia los Balcanes por los invasores selúcidas y mongoles (Font-Porterías et al., 2019). Durante los siglos XI y XII la población romaní se asentó en la Península balcánica adquiriendo costumbres más

sedentarias. Tras ese período comenzó la migración por el resto del continente. Esta dispersión por oleadas coincide con los gradientes genéticos encontrados a lo largo del continente en el eje este-oeste (Fraser, 1992; Gusmão et al., 2010; Rochow y Matschke, 1991)

La formación de las poblaciones romaníes actuales de los países europeos es el producto compuesto de las primeras migraciones de los Balcanes a Europa occidental y otras posteriores que se completaron en el siglo XV (Gresham et al., 2001). Durante la migración la población romaní se fue subestructurando en unidades con tamaños efectivos más reducidos. Los descendientes de una de estas subpoblaciones (Romaní Vlax o valaco romaníes) provienen de aquellos que migraron a la región de Valaquia (actual Rumanía), donde fueron sometidos a la esclavitud (Salihovic et al., 2011). Otro grupo se asentó próximo a la salida de la península balcánica dentro del territorio del imperio austrohúngaro (los denominados romungro romaníes). Posteriormente, diversos grupos se trasladaron en dirección a Europa central y occidental durante la Edad Media (Martínez-Cruz et al., 2016). La llegada de la población romaní a la Península Ibérica está datada durante el siglo XV, apareciendo en Barcelona en el año 1447 (Font-Porterías et al., 2019) (Figura 2), si bien el primer documento escrito que los menciona se remonta a 1425 y se refiere a un salvoconduto emitido por el rey Alfonso de Aragón para que los gitanos fueran admitidos a la entrada de pueblos y ciudades (Lermo et al., 2006).



Figura 2: Mapa de las migraciones de las poblaciones romaníes por el continente europeo. Basado en Haywood (2008).

Ya entrando en los siglos XIX-XX se sucedieron varias oleadas migratorias superpuestas: la primera a finales del siglo XIX con motivo de la abolición de la esclavitud gitana en Rumanía (Fraser, 1992; Gresham et al., 2001; Hancock, 1987) que, posteriormente, la Guerra Fría frenó principalmente en Albania, Bulgaria, Checoslovaquia, Rumania y la ex Yugoslavia (Claude y Elspeth, 2008; Viorel, 2004); la segunda se originó en la antigua Yugoslavia entre 1960 y 1970; y la tercera, durante la década de los 90 como consecuencia de cambios en el contexto político y económico en Europa del Este (Claude y Elspeth, 2008; Gresham et al., 2001; Reyniers, 1995). Dichas oleadas migratorias sucesivas han promovido la subdivisión de los grupos poblacionales en subgrupos más pequeños y aislados. Estos subgrupos presentan una estricta estructura jerárquica que limita las uniones entre miembros del mismo clan y, de manera más marcada, entre clanes. Todo ello en el marco de un conjunto de normas desiguales para ambos sexos (Kalaydjieva et al., 2001a).

Como consecuencia de esta estructura social, del tamaño poblacional reducido de cada subgrupo y del aislamiento geográfico de los clanes, los romaníes presentan un alto grado de endogamia. Un ejemplo lo constituye el clan “Kalajdzii” en Bulgaria, en el que se prohíbe

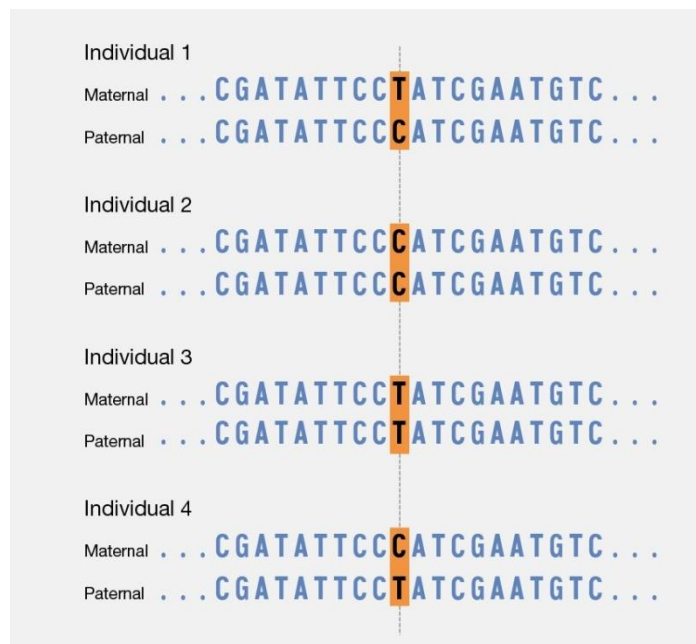
cualquier casamiento fuera del clan, manteniendo la costumbre de que los jóvenes varones se relacionen con las mujeres en eventos sociales organizados por el clan y tengan que ofrecer una dote por su mano. Esta dote es aceptada (o rechazada) por el patriarca de la familia de la posible novia, aunque en la actualidad las jóvenes Kalajdzii deciden ellas mismas si aceptar o no esa petición de mano (Marushiakova y Popov, 2008, 2013). En lo que respecta a las uniones entre grupos, la violación de las normas sociales tiene como consecuencia la expulsión de la pareja intergrupala o la degradación del miembro de mayor rango y su posterior asimilación por el grupo de rango inferior. Aunque esta ha sido la norma, a lo largo de la historia y como consecuencia de un contexto social particular han surgido nuevos clanes a partir de uniones intergrupales como sería el caso de los Zhutanos. Los romaníes Zhutanos o romaníes judíos surgieron en Bulgaria durante la Segunda Guerra Mundial en los campos de concentración nazis donde estaban recluidas ambas poblaciones (Kalaydjieva et al., 2005). En definitiva, las numerosas particularidades que caracterizan a esta etnia determinan un gran interés para su estudio en el marco de la genética de poblaciones humanas.

I.2. Marcadores SNP Informativos de Ascendencia (AIM)

A lo largo del tiempo las poblaciones de una misma especie, pero de diferentes lugares geográficos, tienden a diferenciarse genéticamente. Estas diferencias pueden ser originadas bien por su adaptación a las distintas condiciones del ambiente en el que habitan, o como consecuencia de mutaciones aleatorias que se van acumulando desde que ambas poblaciones divergen, dando como resultado fluctuaciones de las frecuencias alélicas de una generación a otra. En sentido contrario, eventos de intercambio de genes debidos a mestizajes o migraciones actúan para reducir estas diferencias (Nicholson et al., 2002). Por tanto, es de gran importancia intentar comprender la diferenciación a nivel genético de las poblaciones y los mecanismos a los que están sujetas. En general, para abordar el origen de las diferencias entre poblaciones se suelen utilizar distintos tipos de marcadores como, por ejemplo, las inserciones *Alu*, los STR (*Short Tandem Repeat*) o los SNPs.

Un *Single Nucleotide Polymorphism* (polimorfismo de un solo nucleótido) o SNP es una variación de la secuencia en una sola base, en un lugar específico del genoma entre individuos (Figura 3). Los SNPs son la forma más abundante de variación genética humana; un ejemplo serían los 84,7 millones de SNPs encontrados en los 2504 individuos analizados en el proyecto 1000 genomas (1000 Genomes Project Consortium et al., 2015). Por ello, son una gran herramienta para mapear rasgos genéticos complejos (Marth et al., 1999). Se estima que aproximadamente el 50% de los SNPs se encuentran en regiones no codificantes del genoma, el

25% produce mutaciones de sentido erróneo y el 25% restante mutaciones silenciosas (Halushka et al., 1999). Los SNPs no silenciosos podrían habitualmente generar y/o promover ciertas patologías, o ser la base genética de las diferencias interindividuales en la respuesta a los medicamentos y, por tanto, podrían estar sujetos a selección natural; no sería el caso de los SNPs silenciosos que, en líneas generales, no sufren esta selección (Lohrer y LeVan et al., 2001; Tangen, 2000; Wang et al., 2012). Son, precisamente, estos últimos los que se han utilizado de manera extensiva en los estudios de genética de poblaciones (Bulbul et al., 2018; Hwa et al., 2016; Jiang et al., 2018; Jin et al., 2020; Jung et al., 2019; Kidd et al., 2021; Lan et al., 2020; Liu et



al., 2020; Piras et al., 2012; Wei et al., 2014).

Figura 3: Representación de un SNP. Tomado de SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs). National Human Genome Research Institute, NIH. 12/05/22)

Los SNPs presentan varias características que les dotan de altas capacidades para dichos estudios. En primer lugar, los productos derivados de un análisis PCR (Reacción en Cadena de la Polimerasa) de estos marcadores pueden presentar tamaños muy reducidos y, en consecuencia, son buenos candidatos para analizar muestras degradadas, como el ADN antiguo (Wutkeet et al., 2019), a diferencia por ejemplo de los *Alu* o los STR que necesitan secuencias mayores. Entre otras ventajas, los SNPs tienen una gran capacidad de automatización así como la posibilidad de predecir orígenes étnicos o fenotipos como consecuencia de una minuciosa selección de los mismos; de ahí su gran uso también en aplicaciones forenses (Butler et al., 2007). En líneas generales los SNPs son bialélicos, es decir, presentan dos alelos posibles y tres posibles genotipos. Debido a esta característica, un solo SNP bialélico proporciona menos información

que un marcador polialélico como los STRs, lo que hace necesario el uso de un mayor número para poder definir las muestras analizadas con suficiente discriminación. Por ello ha resultado muy útil el desarrollo de técnicas de análisis masivos de SNPs, cada vez más eficientes (Chen et al., 2016).

Los SNPs tienen una tasa de mutación de entre $1,8$ a $2,5 \times 10^{-8}$, significativamente inferior a otros tipos de marcadores como los CNV (*copy number variation* o variación en el número de copias) y los STRs, con tasas de mutación del orden de 10^{-5} (Fan et al., 2007; Huet et al., 2017). A su vez, esas tasas no son constantes a lo largo de la secuencia, haciendo que su distribución no sea heterogénea a lo largo del genoma, pudiéndose encontrar regiones con una alta concentración de SNPs, sujetas a unas mayores tasas de mutación, y otras con una muy baja densidad; además, su distribución parece no ser aleatoria (Varela et al., 2010). Es posible que la selección pueda modificar los niveles de polimorfismo de los nucleótidos en lugares del genoma con unas determinadas tasas de recombinación. En regiones con una baja recombinación o sin recombinación la densidad de los SNPs será reducida y en regiones con altos niveles la densidad de los SNPs tenderá a ser elevada (Nachman, 2001; Varela et al., 2010).

Es igualmente posible que la distribución no aleatoria de los SNPs esté influenciada por la presencia o ausencia de microsatélites en la secuencia. Los microsatélites están expuestos a altos niveles de polimorfismo en la longitud de su secuencia y estas variaciones en el tamaño de la región podrían promover tasas de mutación locales y, por lo tanto, aumentar la densidad de SNPs en la región (Tianet et al., 2008; Varela et al., 2010). Otra posible causa de la distribución no aleatoria es la presencia de repeticiones púricas-pirimidínicas en las regiones flanqueantes (es decir, a cada lado de la secuencia específica) de los microsatélites, que provocan desequilibrios en las tasas de mutación de esas zonas (Varela y Amos, 2009). Por todo ello los mecanismos que determinan la densidad de SNPs no están claros, pero sí que la presencia de los microsatélites tiene un efecto en las tasas de mutación de las zonas adyacentes (Varela et al., 2010).

En 2007 se celebró en Dinamarca un congreso de la Sociedad Internacional de Genética Forense con el objetivo de clasificar los SNPs en función de sus aplicaciones. Los SNPs se clasificaron en i) SNPs con capacidad de identificación individual, ii) informativos de linaje, iii) informativos de fenotipo y, por último, iv) informativos de ascendencia (Butler et al., 2008). Estos últimos son los que entrarían dentro del llamado “grupo de marcadores informativos de ascendencia” o AIMs y que se han usado en esta investigación. La necesidad de su uso se observó en los estudios genético-evolutivos de poblaciones continentales, especialmente en poblaciones

mixtas, a la hora de interpretar los resultados de las frecuencias alélicas. La existencia de subdivisiones dentro de la población, un mestizaje reciente entre etnias y poblaciones y variaciones en los procesos de muestreo, pueden dar lugar a falsas asociaciones entre un fenotipo y un locus o, incluso, pueden enmascarar las verdaderas asociaciones (Price et al., 2008; Tian et al., 2008). Este problema se hace más evidente en estudios de asociación genética en muestras multiétnicas (Cooper et al., 2008). El problema se reduce en los casos en que se manejen un gran número de SNPs, como los análisis de genoma completo (GWAS), donde se pueden obtener los componentes ancestrales de cada individuo dentro de la muestra poblacional. Sin embargo, en aquellos análisis donde el número de marcadores es más limitado, como en los estudios de un gen particular, o cuando el tamaño muestral es demasiado grande y el genotipado sería demasiado costoso, esta solución quedaría descartada. Todo ello muestra la necesidad de definir un número mínimo de marcadores capaces de evaluar las diferencias a nivel individual dentro de una población continental (Santos et al., 2016).

Los AIMs son, en definitiva, marcadores genéticos que presentan frecuencias significativamente diferentes entre poblaciones localizadas en diferentes regiones geográficas. Estos marcadores se pueden usar para estimar los orígenes geográficos de los antepasados de un individuo o población, típicamente por continente de origen (Santos et al., 2016). Por ello se han ido desarrollando sets de SNPs con características AIMs con el objetivo de identificar regiones específicas como el caso de continentes (de la Puente et al., 2016) (Figura 4).

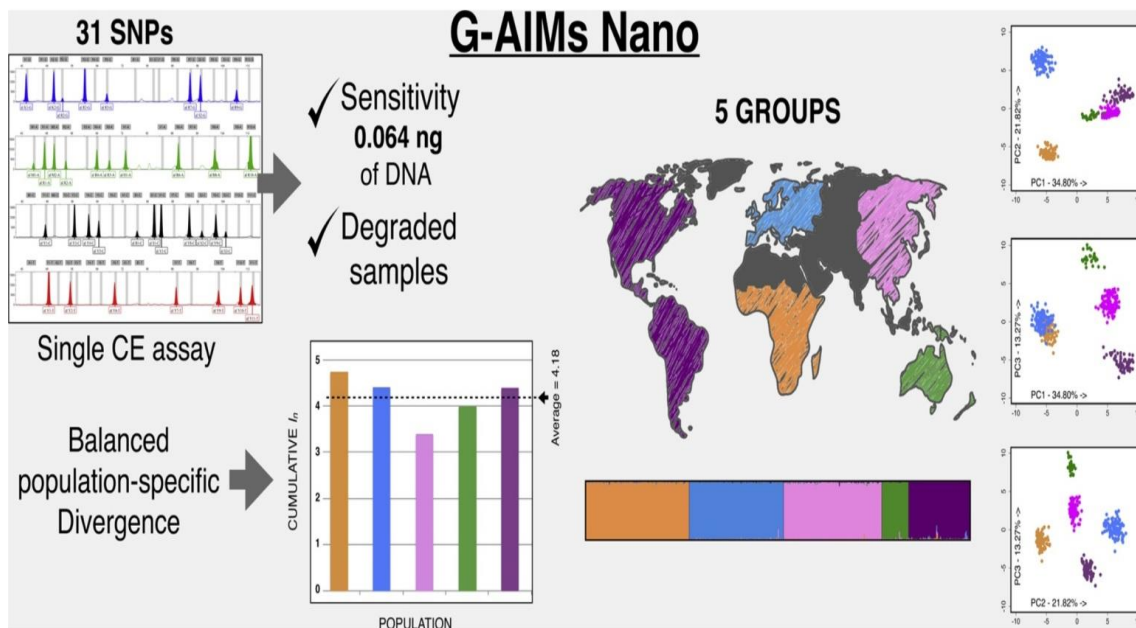


Figura 4: Representación gráfica del proceso de selección de SNPs con características AIMs para un set. Basado en de la Puente et al. (2016).

Durante las últimas décadas, diferentes estudios han demostrado la existencia de un gran número de polimorfismos de un solo nucleótido (SNPs) que presentan grandes diferencias en sus frecuencias alélicas entre varias poblaciones continentales (Mao et al., 2007; Tian et al., 2006, 2007), y que encajan en el marco definido por la característica principal de un AIM. Así, durante los últimos años, SNPs con características AIMs están siendo utilizados de forma cada vez más frecuente en los estudios de genética poblacional (Barbosa et al., 2017; He et al., 2021; Hopkins et al., 2019; Huerta-Chagoya et al., 2019; Lins et al., 2010; Wang et al., 2018, 2019; Yahya et al., 2017, 2020; Zeng et al., 2016), lo que justifica su uso en la presente investigación.

1.3 Elementos *Alu*

El genoma es el conjunto completo de instrucciones genéticas que se encuentran en una célula. En los seres humanos, el genoma consta de 23 pares de cromosomas que se encuentran en el núcleo, así como un pequeño cromosoma circular que se encuentra en las mitocondrias de las células. Cada conjunto de 23 cromosomas contiene aproximadamente 3,1 mil millones de bases de secuencia de ADN (Genome. National Human Genome Research Institute, NIH 27/01/2022). A su vez, en el ADN es posible distinguir dos tipos de secuencias: codificantes y no codificantes. Las primeras contienen la información para codificar un producto génico final; en cuanto a las segundas no generan ningún producto génico pero realizan otras funciones como la regulación de la expresión génica (ADN no codificante. National Human Genome Research Institute, NIH 27/01/2022). Dentro de las secuencias no codificantes es posible diferenciar entre secuencias únicas y repetidas. Estas últimas se pueden encontrar dispersas por el genoma o en tándem, como los STR (*Short Tandem Repeats*). En el caso de las secuencias no codificantes repetidas dispersas, se clasifican en función de su tamaño: por una parte están los LINEs (*Long INterspersed Elements*), con una longitud media de 6000 pb (pares de bases), y por otra los SINEs (*Short INterspersed Elements*), de tamaños inferiores a 500 pb. Las inserciones *Alu* conforman la familia más grande de SINEs, representando más del 10% del ADN humano (Figura 5) (Aktas et al., 2017; Cordaux et al., 2009). Fueron descubiertas en 1979 como parte de las curvas de renaturalización del ADN humano y se bautizaron con ese nombre debido a que en su secuencia presentaban un lugar de reconocimiento de la enzima de restricción *AluI* (Houck et al., 1979).

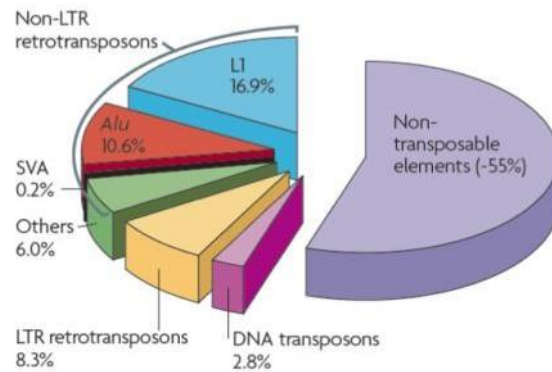


Figura 5: Representación del porcentaje de elementos *Alu* contenido en el genoma humano. Modificado de Cordaux et al. (2009).

Estructura de los elementos *Alu*

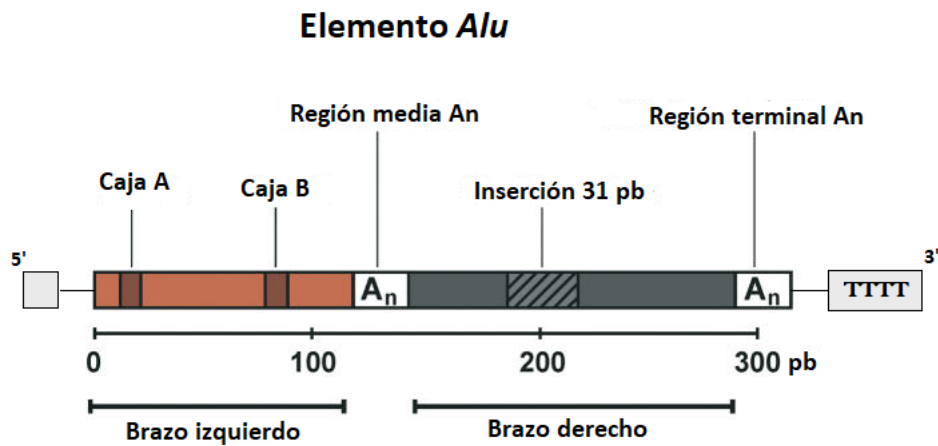


Figura 6: Estructura heterodimérica del elemento *Alu*. Modificado de Häsler y Strub (2006).

Los elementos *Alu* presentan una longitud aproximada de 300 pb. Están conformados por una estructura heterodimérica compuesta por dos monómeros relacionados, pero no equivalentes (brazos izquierdo y derecho). El brazo derecho contiene una inserción específica de 31 pb; el izquierdo contiene las denominadas cajas A y B del promotor interno de la ARN polimerasa III, funcionales pero débiles. Ambos brazos están separados por una región rica en Adenina (Región media A_n), seguidos por una cola corta de poli (A) (Región terminal A_n) (Häsler y Strub, 2006) (Figura 6).

Origen y evolución de los elementos *Alu*

Se considera el gen 7SL RNA como el más probable origen de los elementos *Alu* (Ullu y Tschudi, 1984; Gal-Mark et al., 2008). Dicho gen forma parte del complejo ribosómico y contiene la información para una partícula de reconocimiento del retículo endoplasmático (Ullu y Tschudi, 1984). Actualmente se estudia, entre otras cosas, por su relación con procesos carcinogénicos (Abdelmohsen et al., 2014) o su implicación con enfermedades retrovíticas como el VIH-1 (Itano et al., 2018; Keene et al., 2012). El origen y multiplicación de los elementos *Alu* se divide en dos etapas. Una primera etapa monomérica, basada en un monómero ancestral *Alu* FAM (*Fossil Alu Monomer*), que probablemente se formó a partir de un retrotransposón ineficiente, debido a una versión eliminada del gen 7SL RNA, en algún momento antes de la divergencia evolutiva entre primates y roedores (Batzer y Deininger, 2002). Y una segunda etapa dimérica, que se originó a partir de la evolución del monómero ancestral en las secuencias FLAM (*Free Left Alu Monomer*) y FRAM (*Free Right Alu Monomer*). Posteriormente se formó el dímero de FLAM y FRAM dando como resultado la estructura heterodimérica de los elementos *Alu* actuales (Figura 7). Este evento se produjo en algún momento del inicio de la radiación de los primates (65-57 Ma), lo que hace que sean específicos de este linaje de mamíferos (Quentin et al., 1992; Kojima et al., 2020; Kriegs et al., 2007) (Figura 8).

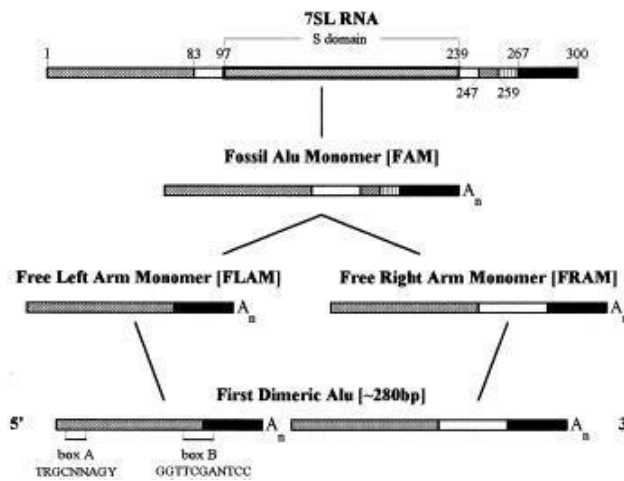


Figura 7: Representación del origen y formación de los elementos *Alu*. Tomado de Mighell et al. (1997).

La evolución de los elementos *Alu* ha sido progresiva debido a la acumulación de nuevas inserciones ya que carecen de un mecanismo específico para su eliminación. A lo largo del tiempo, diferentes períodos han dado como resultado evolutivo un grupo variado, aunque limitado, de subfamilias. La primera subfamilia de los elementos *Alu* fue la subfamilia J, seguida de una serie muy activa de subfamilias S. Las subfamilias S dominantes incluyeron a las Sx, Sq, Sp y Sc. Más recientemente, apareció la subfamilia Y, con las subfamilias Ya5 e Yb8, que es la

dominante en humanos (Shen et al., 1991) (Figura 8). Dicha evolución no ha ocurrido a la misma velocidad entre todos los linajes de primates. Si se comparan chimpancés y humanos, desde su divergencia hace unos 6 millones de años se han fijado unas 2.300 y 5.500 inserciones específicas de cada linaje, respectivamente (Locke et al., 2011; Mills et al., 2006). En el orangután sólo se han adquirido unas 250 inserciones específicas de linaje en los últimos 12 millones de años, lo que muestra una tasa de amplificación mucho menor en ese linaje (Locke et al., 2011). Estos resultados sugieren que los eventos de cuello de botella que ocurren durante los procesos de especiación pueden dar lugar a niveles alterados de actividad *Alu*, probablemente a través de la fijación de diferentes niveles de actividad de los elementos originales (Locke et al., 2011; Mills et al., 2006).

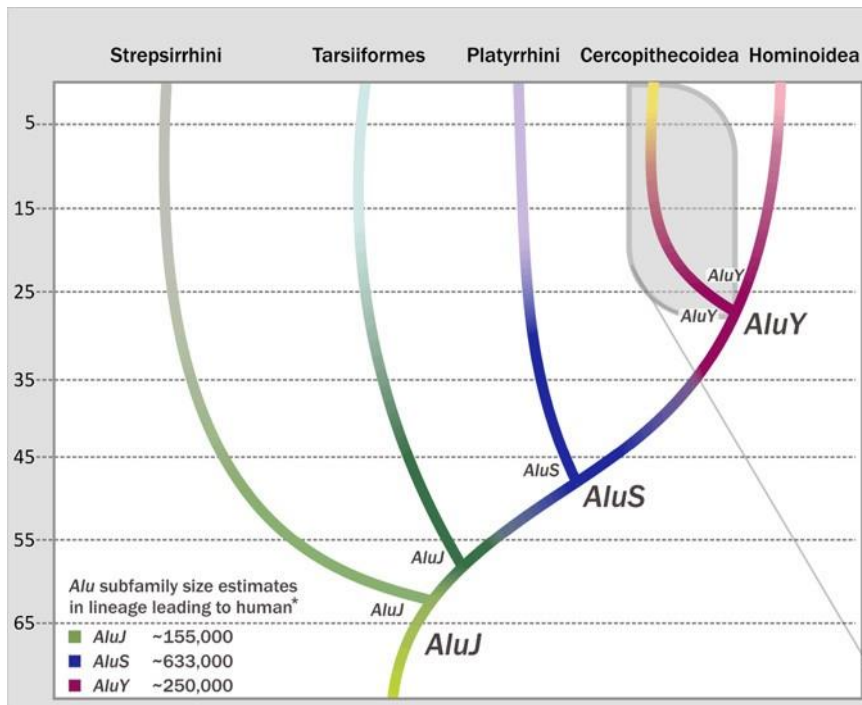


Figura 8: Evolución de las subfamilias *Alu* en Primates. Se muestran las 3 subfamilias principales (*AluJ*, *AluS* y *AluY*). El rango de su actividad y la evolución continua se indica mediante un degradado de color. Modificado de Konkel et al. (2010).

Retrotransposición

Los elementos *Alu* son elementos móviles que, por su propia naturaleza, tienen la capacidad de insertarse en diferentes regiones del ADN utilizando un intermediario de ARN; a este proceso se le denomina retrotransposición o TPRT (*Target Primed Reverse Transcription*) (Batzer y Deininger, 2002; Novick et al., 1996, 1998; Salem et al., 2003; Wu et al., 2014) (Figura 9). Este mecanismo se divide en varias fases. En primer lugar, por medio de la actividad de una RNA polimerasa III, el elemento *Alu* se transcribe a Alu-ARN (Batzer y Deininger, 2002; Salem et

al., 2003; Stoneking et al., 1997). A continuación, debido a la acción de una Transcriptasa Inversa, el ARN se utiliza como base para generar un ADN complementario y este, a su vez, es integrado en otra parte del genoma mediado por endonucleasas (Batzer y Deininger, 2002; Kajikawa et al., 2002; Salem et al., 2003). Es en este último paso en el que se generan las repeticiones cortas que flanquean al elemento *Alu* (Batzer y Deininger, 2002). Los elementos *Alu* no codifican para estas últimas enzimas por lo que posiblemente provengan de elementos LINE próximos (Batzer y Deininger, 2002; Häslér y Strub, 2006; Kajikawa et al., 2002; Salem et al., 2003).

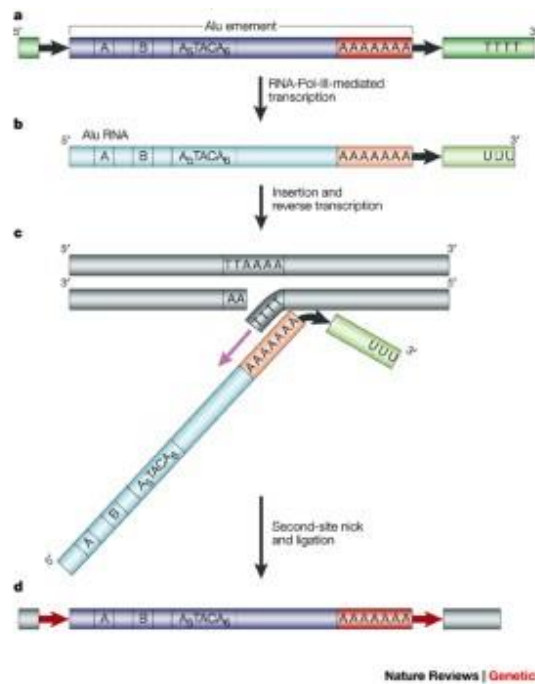


Figura 9: Modelo de retrotransposición (TPRP). a) Elemento *Alu*. b) Transcrito ARN-*Alu* mediada por ARN Polimerasa III. c) Inserción y transcripción inversa del ADN-*Alu* en el ADN huésped. d) Elemento *Alu* integrado en el ADN huésped. Tomado de Batzer y Deininger (2002).

Subfamilias de los elementos *Alu*

Los elementos *Alu* se dividen en 12 grandes familias. Se define una familia como el conjunto de elementos *Alu* que comparten una serie de mutaciones en posiciones diagnósticas de su secuencia (Batzer y Deininger, 2002; Batzer et al., 1994; Häslér y Strub, 2006; Quentin, 1992). Estas familias se organizan en función de su antigüedad, de modo que los elementos *Alu* pertenecientes a una misma familia presentan edades de inserción similares (Quentin, 1992; Batzer et al, 1994; 1996b; Batzer y Deininger, 2002; Novick et al, 1998; Häslér y Strub, 2006; Stoneking et al., 1997). La familia J es la más antigua, con una edad aproximada de unos 81 Ma (Figura 10). Es posterior a la divergencia entre roedores y primates (hace unos 100 Ma) y anterior a la radiación de los primates, hace unos 65 Ma. Esto último encaja con la ausencia de *AluJ* en el

genoma de los roedores y la presencia de secuencias diméricas similares a las *AluJ* (*AluJ*-like) (Kapitonov y Jurka, 1996).

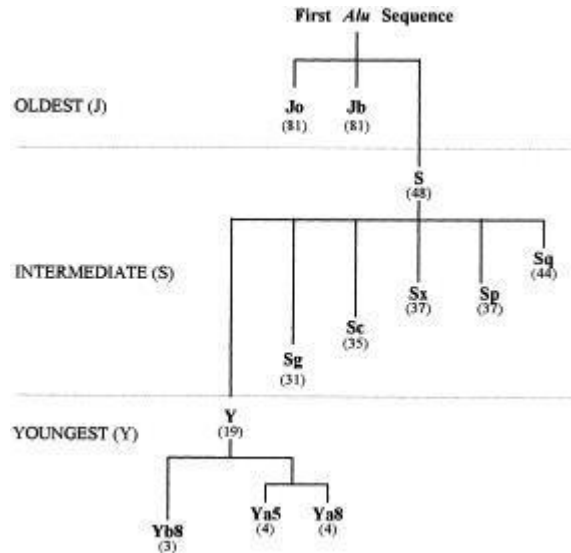
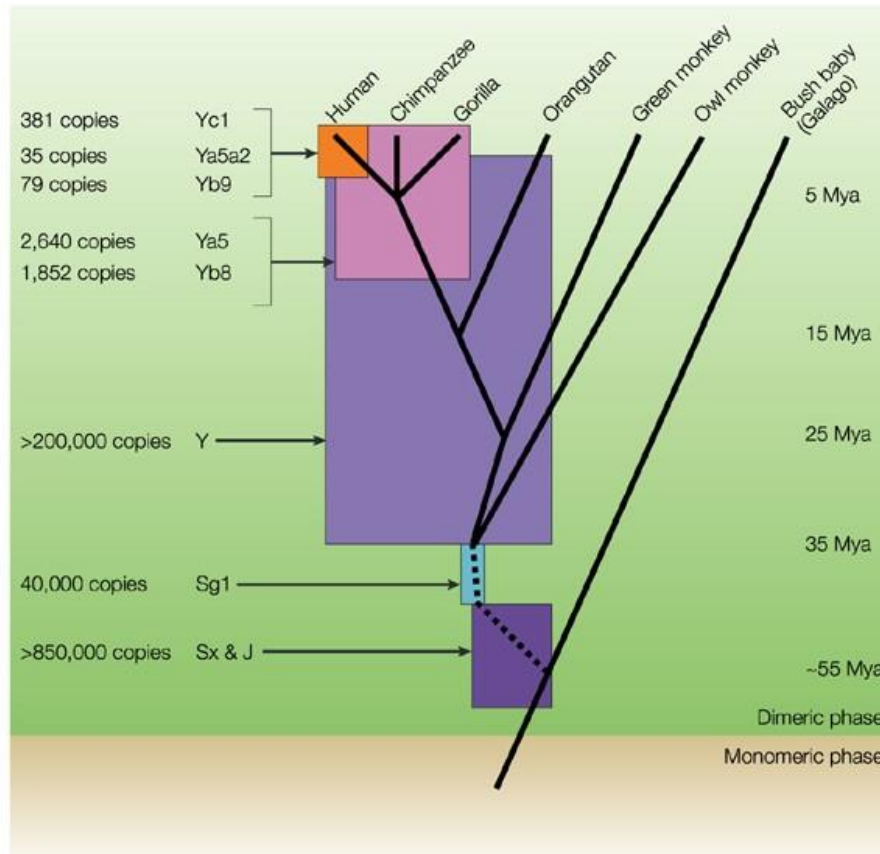


Figura 10: Representación de las 12 subfamilias de los elementos *Alu* organizadas en antiguas (J), intermedias (S) y jóvenes (Y). Los valores entre paréntesis corresponden a su antigüedad en millones de años. Según Mighell et al. (1997).

En segundo lugar, se encontraría la familia S, de una edad intermedia, entre 44 y 31 Ma (Figura 10); esta familia conforma más del 80% de los *loci Alu* presentes en el genoma humano como resultado de la alta tasa de retrotransposición durante el periodo comprendido entre los 60 Ma y 35 Ma (Roy-Engel et al., 2008) (Figura 11).



Nature Reviews | Genetics

Figura 11: Expansión de las subfamilias *Alu* durante la radiación de los Primates y tiempo en millones de años (Mya). Tomado de Batzer y Deininger (2002).

Por último, se encuentra la familia más reciente de los elementos *Alu*, la *Y*, que presenta una antigüedad de entre 19 Ma y 3 Ma (Figura 10). Aunque de manera general los elementos *Alu* se encuentran en estado monomórfico, se calcula que el 25% se encuentra en estado polimórfico, debido a su inserción reciente (Batzer y Deininger, 2002). Dentro de la familia *Y*, las subfamilias *Yb8*, *Y* e *Ya5* son las mayoritarias, con una frecuencia del 34%, 30% y 24%, respectivamente (Hedges et al., 2004) (Figura 12).

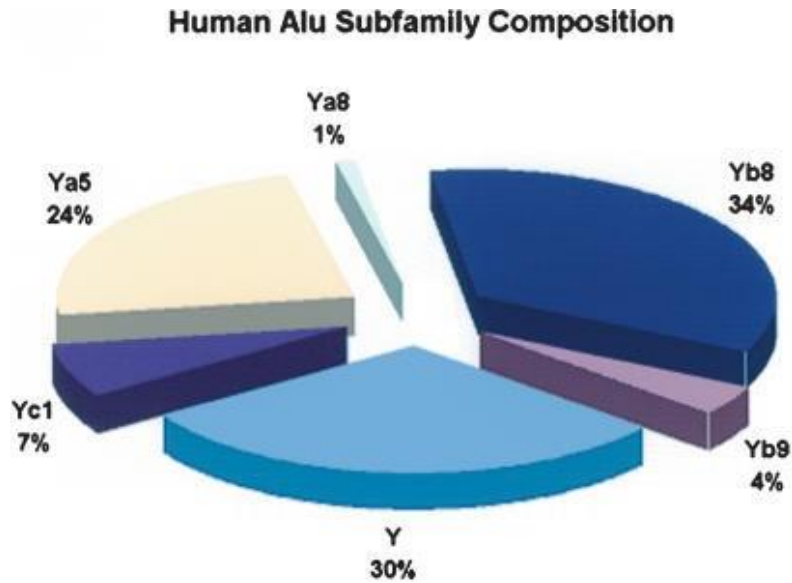


Figura 12: Composición de la subfamilia Y de los elementos *Alu* en el linaje humano. Modificado de Hedges et al. (2004).

Función de las inserciones Alu en estudios genético-evolutivos

Las inserciones *Alu* presentan diversas características que las permiten ser marcadores útiles para estudios genético-evolutivos. En primer lugar, son marcadores altamente estables. Esto significa que la probabilidad de que una inserción *Alu* sufra una segunda inserción es mínima (Batzer et al., 1994, 2000). En consecuencia, todos los individuos que presenten una determinada inserción estarán necesariamente emparentados entre sí. Otros marcadores polimórficos, como los SNPs no presentan esta característica, por lo que un mismo alelo puede estar localizado de manera independiente un número variable de veces; en este caso individuos que presenten un mismo alelo no estarían necesariamente emparentados (Batzer et al., 1994, 2000; Novick et al., 1998; Salem et al., 2005). Además, dicha estabilidad también se muestra en la baja probabilidad de que una inserción *Alu* sufra una delección. En el caso poco frecuente de que ocurriera esta no sería completa, por lo que los restos del polimorfismo en el ADN permitirían reconocer que hubo una inserción en esa posición de la secuencia. Ambas características ejemplifican la baja tasa de cambio de estas inserciones a lo largo del tiempo (Batzer et al., 1994, 2002; Novick et al., 1998; Salem et al., 2005).

Las inserciones *Alu* presentan asimismo la cualidad de tener un estado ancestral conocido. Esto es debido a que la dirección de la mutación se dirige forzosamente hacia la inserción del elemento, por lo que su ausencia es, precisamente, el estado ancestral (Batzer et al., 1994, 2002; Novick et al., 1998; Salem et al., 2005). Esta propiedad permite, mediante análisis estadísticos, poder simular una hipotética población ancestral (población con ausencia

total de inserción), que podría ser usada a la hora de analizar relaciones poblacionales o para generar árboles filogenéticos o dendrogramas. En relación al párrafo anterior hay excepciones a la norma como el caso del *Alu* CD4, que sería la presencia de inserción su estado ancestral. Esto es debido a que su polimorfismo actual está causado por una delección parcial posteriormente de la fijación de un elemento *Alu* (Majumder et al., 1999). De cualquier manera, sigue siendo posible deducir su estado ancestral. Por último, los elementos *Alu* se consideran generalmente como marcadores selectivamente neutros, por lo que las frecuencias de sus polimorfismos dependerán principalmente de las interacciones entre el flujo génico y la deriva genética (Comas et al., 2000).

Elementos Alu analizados

A continuación se describe el significado y función de los elementos *Alu* analizados en esta Tesis Doctoral, así como su interés y validez en los estudios sobre genética de poblaciones.

ACE

Este polimorfismo *Alu* recibe el nombre del gen de la enzima convertidora de angiotensina (*angiotensin converting enzyme* o ACE), que pertenece a la familia *Alu* Ya5 (Batzer et al., 1996 a, b) y es un importante marcador genético utilizado en numerosos estudios (Di Pasquale et al., 2004; Dmitrenko et al., 2020; Eisenmann et al., 2009; Feng et al., 2002; Liu et al., 2011; Rigat et al., 1990; Schunkert et al., 1994). En *Homo sapiens*, el gen que codifica la ACE se encuentra localizado en la región 17q23, en el brazo largo del cromosoma 17, y comprende 26 exones y 25 intrones (Dmitrenko et al., 2020). Este polimorfismo se caracteriza por la presencia o ausencia de una secuencia de 287 pb dentro del intrón 16 (Dmitrenko et al., 2020; Rigat et al., 1990). La ACE es una enzima clave del sistema renina-angiotensina-aldosterona (RAAS) (Bouzekri et al., 2004; Zhu et al., 2001), que juega un papel vital en la regulación de la presión arterial sistémica y la homeostasis de los electrolitos renales, al convertir la angiotensina I inactiva en la angiotensina II vasoconstrictora, e inactivando el vasodilatador proinflamatorio bradicinina (Zhu et al., 2001).

Varios trabajos de investigación han informado de una asociación significativa entre el polimorfismo *Alu* ACE y una serie de enfermedades, que incluyen hipertensión, diabetes tipo 2, enfermedades cardiovasculares, enfermedades renales, cáncer y obesidad (Di Pasquale et al., 2004; Eisenmann et al., 2009; Feng et al., 2002; Schunkert et al., 1994). No obstante, un gran número de estudios realizados sobre esta inserción sostienen que, generalmente, su comportamiento es selectivamente neutro por lo que se puede utilizar para establecer

afinidades genéticas entre poblaciones (Batzer et al., 1996a, b; Berkman et al., 2008; Cotrim et al., 2004; Deininger et al., 1999; García-Obregón et al., 2006, 2007; Hidalgo et al., 2014; Laybourn et al., 2016; Nasidze et al., 2001; Navarrete et al., 2017; Novick et al., 1998; Rupert et al., 2003; Singh et al., 2016, 20167; Stoneking et al., 1997; Vega-Requena et al., 2020).

APO

Este marcador fue descrito por primera vez por Matera et al. (1990). Su nombre proviene de su proximidad con el clúster de genes APOA1, APOC3 y APOA4 localizados en el cromosoma 11 (11q23.3) (Karathanasis, 1985). *Alu* APO se encuentra a 4,3 kb en posición telomérica del gen APOA1. Este gen interactúa con una proteína de los transportadores dependiente de ATP o transportador ABC (ABCA1) que da como resultado la unión de fosfolípidos con una apolipoproteína anfipática, formando una estructura discoidal (Yokoyama et al., 2006; Hamilton et al., 1976). Este complejo discoidal, a menudo denominado lipoproteína de alta densidad (HDL), funciona para solubilizar el exceso de colesterol presente en la membrana plasmática (Wang et al., 2014) o el colesterol que las células excretan a la matriz extracelular (Jin et al., 2018). Dicho gen ha sido estudiado por su relación con enfermedades coronarias (Muscella et al., 2020) o con enfermedades neurodegenerativas como la enfermedad de Alzheimer (AD) (Ciccione et al., 2020), entre otras. La inserción *Alu* APO es específica del ser humano (Batzer et al., 1994) y pertenece a la subfamilia *Alu*Ya5 (Batzer et al., 1996a, b). Presenta una longitud de 312 pb y es un marcador comúnmente utilizado en estudios evolutivos de poblaciones humanas (Batzer et al., 1996a, b; Berkman et al., 2008; Cotrim et al., 2004; Deininger et al., 1999; García-Obregón et al., 2006, 2007; Hidalgo et al., 2014; Kee et al., 2012; Nasidze et al., 2001; Novick et al., 1998; Singh et al., 2016, 2017; Stoneking et al., 1997; Vega-Requena et al., 2020).

B65

Este polimorfismo *Alu* fue descrito por primera vez por Arcot et al. (1995a); se localiza en el cromosoma 11 (11q14.2), en su región intergénica próxima al gen CHORDC1 (*Cysteine and Histidine-rich domain-containing zinc binding protein 1*) (Rajeevan et al., 2012), específicamente a 1,76 kb de su exón 11. Dicho gen regula la duplicación del centrómero y a su vez está presente en la supresión de procesos cancerígenos (Urabe et al., 2020). El polimorfismo presenta dos formas alternativas. Una de ellas incluye un microsatélite y una cola de poli A compuesta por 14 Adeninas. La segunda variante con ausencia tanto del microsatélite como de la cola poli A en su secuencia. Debido a ello, la inserción de este elemento *Alu* varía entre 342 pb y 313 pb (Arcot et al., 1995b). La inserción *Alu* B65 pertenece a la subfamilia *Alu*Ya5 (Batzer et al., 1996a, b). Al

igual que los marcadores anteriores ha sido extensamente utilizado en estudios de poblaciones (Berkman et al., 2008; Deininger et al., 1999; Hidalgo et al., 2014; Kee et al., 2012; Nasidze et al., 2001; Stoneking et al., 1997; Vega-Requena et al., 2020).

D1

Este polimorfismo está localizado en el cromosoma 3 (3q26.32) a 200kb aguas arriba del gen WIG1 (Geer et al., 2010; Rajeevan et al., 2012). WIG1 es un objetivo transcripcional del gen supresor de tumores p53 y codifica una proteína de dedo de zinc involucrada en la regulación génica postranscripcional (Vilborg et al., 2011). Dicha relación ha sido analizada por su implicación en varios tipos de enfermedades como la de Huntington (Kim et al., 2016), así como en procesos antitumorales (Madar et al., 2013). Aunque la inserción AluD1 fue descrita inicialmente por Batzer et al. (1995), fueron Arcot et al. (1995a), varios meses después, quienes publicaron las condiciones y cebadores necesarios para su amplificación. La inserción, perteneciente a la subfamilia Yb8, se compone de una secuencia de 334 pb, en cuyo extremo terminal 3' presenta el inicio de un microsatélite dinucleótido AT (Batzer et al., 1995). Como en los casos previos, esta inserción ha sido analizada en diversos estudios poblacionales con el fin de comprender sus relaciones filogenéticas (Berkman et al., 2008; Deininger et al., 1999; Hidalgo et al., 2014; Nasidze et al., 2001; Singh et al., 2016; Stoneking et al., 1997; Vega-Requena et al., 2020).

FXIIIB

La inserción *Alu* FXIIIB está localizada en el cromosoma 1 (1q31-q31), específicamente en el intrón 10 del gen FXIIIB. Dicho gen es altamente polimórfico con más de 200 variantes registradas en el *National Center of Biotechnology Information* (NCBI). Está conformado por 12 exones y 11 intrones y codifica una proteína que forma parte del factor de coagulación XIII (Bottenus et al., 1990). Este factor presenta una estructura heterotetramérica compuesta por dos subunidades FXIIIA y FXIIIB (Mangla et al., 2021) y juega un papel importante en la estabilización del coágulo de fibrina, al reticular la fibrina y hacer que el coágulo sea más denso y rígido. También juega un papel fundamental en la retracción del coágulo dependiente de las plaquetas, la cicatrización de heridas y la reparación de tejidos (Mangla et al., 2021). Este polimorfismo fue descubierto por Kass et al. (1994), presenta una extensión de 290 pb y pertenece a la subfamilia Ya5 (Batzer et al., 1996a). Se trata de una buena herramienta para caracterizar a las poblaciones humanas desde el punto de vista genético-evolutivo (Batzer et al., 1996a, b; Berkman et al., 2008; Cotrim et al., 2004; Deininger et al., 1999; García-Obregón et al.,

2006, 2007; Hidalgo et al., 2014; Nasidze et al., 2001; Novick et al., 1998; Stoneking et al., 1997; Vega-Requena et al., 2020).

PV92

Este polimorfismo se ubica en un intrón del gen que codifica para la Cadherina 13 (CDH13), en el cromosoma 16 (Cadherina H, de *Heart* en inglés, por su vínculo con el corazón) (Geer et al., 2010; Rajeevan et al., 2012). La CDH13 es un glicosilfosfatidilinositol de anclado atípico, miembro de la superfamilia de las cadherinas, que carece de los dominios transmembrana y citoplasmático pese a pertenecer a dicha superfamilia (Angst et al., 2001). Su interés en los procesos de neoplasias malignas ha ido en aumento a lo largo de los últimos años. Así, se ha observado que está infraexpresado en muchos tipos de cáncer como el carcinoma de ovario (Andreeva et al., 2010; Makarla et al., 2005), cáncer gástrico (Tang et al., 2012), cáncer colorrectal (Millar et al., 2002) cáncer de pulmón (Andreeva et al., 2010), cáncer de próstata (Andreeva et al., 2010) y cáncer de mama (Kong et al., 2015), entre otros, y generalmente se asocia con un mal pronóstico (Kong et al., 2017).

La inserción *Alu* PV92 presenta una longitud de 314 pb y pertenece a la subfamilia Ya5, subfamilia de elementos *Alu* más reciente (Batzer et al., 1996b); además es exclusiva de la especie humana (Batzer et al., 1994), por lo que viene siendo ampliamente utilizada en estudios poblacionales desde el punto de vista genético-evolutivo (Batzer et al., 1996a, b; Berkman et al., 2008; Deininger et al., 1999; García-Obregón et al., 2006, 2007; Hidalgo et al., 2014; Kee et al., 2012; Nasidze et al., 2001; Novick et al., 1998; Stoneking et al., 1997; Vega-Requena et al., 2020).

A25

Esta inserción fue descrita por primera vez por Arcot et al. (1995a) y tiene una longitud de 284pb. Pertenece a la subfamilia Ya5 y está localizada en el cromosoma 8 (8q21.3), a 500 kb (*aguas abajo*) en posición centromérica del gen MMP16 (Rajeevan et al., 2012). Se trata de un gen que codifica una metaloproteinasa de matriz (MMP); se trata de una familia de enzimas capaces de descomponer colágeno, entre otras proteínas que se localizan en el espacio que existe entre los tejidos, y su función está vinculada con procesos de deterioro de la matriz celular tanto a nivel fisiológico como patológico (Sato et al., 1997). A nivel patológico se ha estudiado su implicación en procesos oncológicos (Cao et al., 2016; Li et al., 2011; Shen et al., 2017) y su relación con otras patologías neurológicas como la enfermedad de Parkinson (Chen et al., 2021). Este polimorfismo se ha analizado en un gran número de poblaciones humanas con el fin de evaluar su variabilidad intra e interpoblacional (Arcot et al., 1995a; Berkman et al., 2008;

Deininger et al., 1999; García-Obregón et al., 2006, 2007; Hidalgo et al., 2014; Nasidze et al., 2001; Stoneking et al., 1997; Vega-Requena et al., 2020).

TPA25

El elemento *Alu* TPA25 está localizado en el intrón 8 del gen PLAT (PLasminogen Activator Tissue) (Tishkoff et al., 2000). Dicho gen se encuentra en el brazo corto del cromosoma 8 (8p11.2) y presenta una longitud de 36.594 pb, comprendiendo 28 inserciones *Alu* entre el gen y la región 5' que le flanquea (Degen et al., 1986). Codifica un activador del plasminógeno (PA), una serina proteinasa que convierten la proenzima plasminógena inactiva en plasmina (una proteinasa activa); dicha proteinasa está relacionada con la fibrinólisis y la remodelación tisular. Existen dos tipos de PA inmunológicamente distintos: el tipo tisular (t-PA), la forma principal involucrada en la trombólisis, y el tipo uroquinasa (u-PA), involucrado principalmente en la degradación tisular (Benham et al., 1984). Aunque TPA25 se ha asociado con la trombosis coronaria (Álvarez et al., 2015; Raslan et al., 2020; Santiago et al., 2012), varios estudios han demostrado que no hay relación entre la presencia/ausencia de esta inserción *Alu* con el riesgo de padecer dicha patología (Ahmed et al., 2011; Ridker et al., 1997; Van der Bom et al., 1997). La inserción *Alu* TPA25 se extiende por una longitud de 311 pb y pertenece a la subfamilia Ya8 (Batzer et al., 1996a, b). Se considera un marcador neutro por lo que se ha empleado en numerosos estudios filogenéticos en poblaciones humanas (Batzer et al., 1996a, b, 2008; Cotrim et al., 2004; Deininger et al., 1999; García-Obregón et al., 2006, 2007; Hidalgo et al., 2014; Kee et al., 2012; Nasidze et al., 2001; Novick et al., 1998; Stoneking et al., 1997; Vega-Requena et al., 2020).

I.4 Marcadores STR pertenecientes al Combined DNA Index System (CODIS)

La especie humana presenta más del 99% de identidad entre dos individuos aleatorios, por lo que las diferencias que presentamos tanto a nivel de fenotipo o genotipo deben ser determinadas por el porcentaje restante (Glazko et al., 2005). Dentro de estas diferencias se encuentran un gran número de secuencias repetidas con diferentes tamaños, repartidas a lo largo del genoma y que no están sujetas a una selección negativa debido a que, habitualmente, se encuentran entre genes permitiendo la acumulación de mutaciones neutras. Estas secuencias repetidas tienden a definirse en función de su longitud y/o del número de repeticiones contiguas. A aquellas secuencias con repeticiones consecutivas y con un tamaño medio de hasta 100pb se las denomina VNTR (*Variable Number Tandem Repeats*) (Chamber et al., 2000). Las

secuencias de longitudes de las repeticiones más cortas (entre 2-7 pb) se definen como STR (*Short Tandem Repeats*) y conforman cerca del 3% del genoma humano (Butler y John, 2011; Fan et al., 2007; Gharesouran et al., 2021). Las repeticiones STR se encuentran mayoritariamente en regiones no codificantes, mientras que sólo el 8% estarían en regiones codificantes (Ellegren, 2000; Fan et al. 2007). Además, el 99% de ellas presenta secuencias con una longitud inferior a 33 pb y en el 62% se observan repeticiones imperfectas (Madsen et al., 2008) (Figura 13).

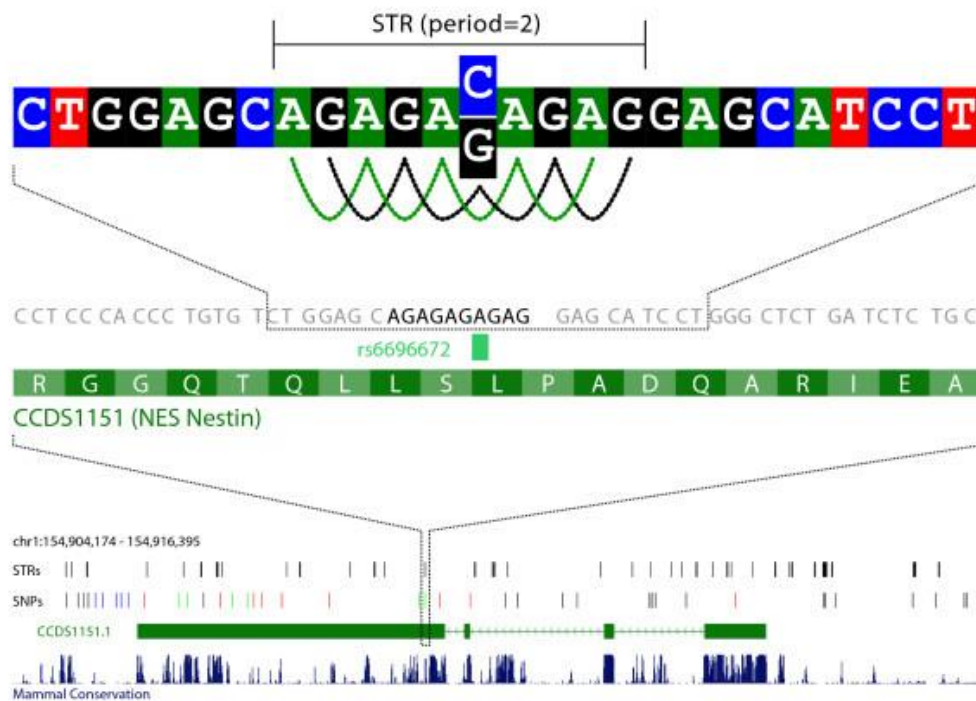


Figura 13: Ejemplo de STR localizado en el gen NES, que codifica la proteína fibrosa nestina, localizado en el cromosoma 1. Basado en Madsen et al. (2008).

A su vez, las STR se subclasifican en varias categorías en función de la longitud de cada unidad de repetición, pudiéndose clasificar en mono-, di-, tri-, tetra-, penta- y hexanucleótidos. En líneas generales la frecuencia de estas repeticiones disminuye al aumentar la longitud de las mismas, pudiendo ser dinucleótidos hasta un 67% de las STR encontradas en el genoma humano, siendo la más frecuente el par AC seguido del AT (Gharesouran et al., 2021) (Figura 14). Los di- y tetranucleótidos se localizan con una mayor frecuencia en regiones intergénicas y secuencias intrónicas frente a los trinucleótidos, que tienen una mayor presencia en los exones (Gharesouran et al., 2021; Tóth et al., 2000). Por otro lado, las STR también pueden presentar diferencias en la exactitud o el rigor en el que las repeticiones se ajustan a un patrón, pudiéndose

dividir en diferentes categorías en función del estatus de sus repeticiones. En primer lugar, se encontrarían aquellas repeticiones simples que contienen unidades de longitud y secuencia idénticas, las repeticiones compuestas que están conformadas por dos o más repeticiones simples contiguas y, finalmente, las complejas formadas por varios bloques de longitud y secuencia variables (Gharesouran et al., 2021).

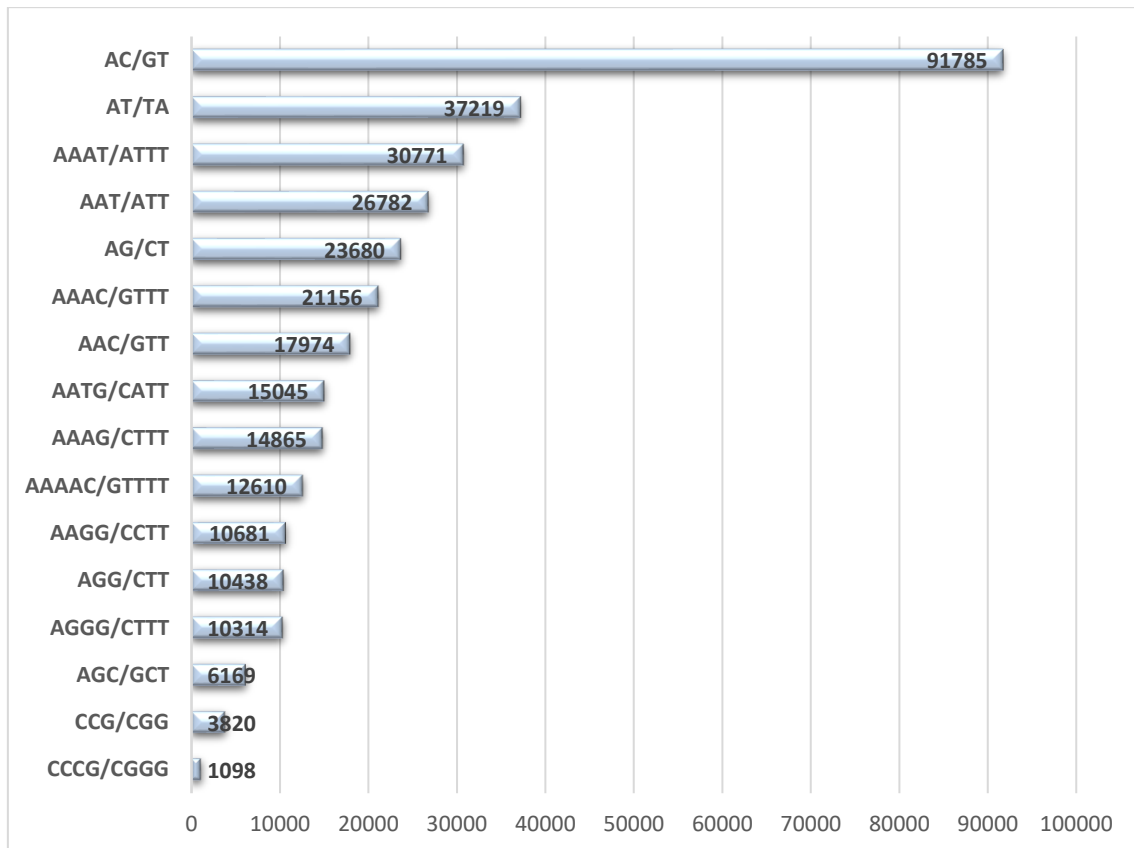


Figura 14: Frecuencias de las STR en el genoma humano. (Modificado de Gharesouran et al., 2021).

Las STR son una herramienta muy útil como marcadores de ADN debido a que el tamaño del amplificado de los alelos de un individuo heterocigoto es similar, lo que evita el problema de una amplificación diferencial durante una PCR. Por otra parte, estos marcadores muestran una tasa de mutación mucho más elevada que otros marcadores, siendo aproximadamente de 10^{-5} en humanos, lo cual, junto a su localización en el genoma, puede promover la acumulación de mutaciones aumentando la variabilidad de repeticiones consecutivas entre individuos (Fan et al., 2007). De igual manera, estas elevadas tasas de mutación pueden promover un aumento significativo en el tamaño de las STR generando patologías, (dependiendo del lugar donde se encuentren), como enfermedades neurológicas y del desarrollo (Gharesouran et al., 2021). En definitiva, las STR presentan unos altos niveles de variación y polimorfismo dotando a esta

familia de un alto valor y versatilidad como marcadores para análisis de poblaciones, o incluso identificativos de individuos, y por ello han sido incluidas en la base de datos del *Combined DNA IndexSystem* (CODIS).

A inicios de 1996 el *Federal Bureau of Investigation*, o más conocido por sus siglas FBI, patrocinó un proyecto para establecer una serie de STR centrales con el fin de incluirlos en la base de datos nacional de ADN, conocida como CODIS (Butler, 2011). El proyecto se inició en abril de 1996 y terminó en noviembre del año siguiente involucrando a 22 laboratorios de tipificado de ADN y el análisis de 17 marcadores candidatos: CSF1PO, F13A01, F13B, FES/FPS, FGA, LPL, TH01, TPOX, vWa, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 y D21S11), de los cuales 13 fueron finalmente seleccionados como la base del proyecto (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, vWA, TPOX, D18S51, D5S818 y FGA) (Butler, 2011; Budowle et al., 1998) (Figura 15).

Information on 13 CODIS STRs

Locus Name	Chromosomal Location	GenBank Accession	Allele in GenBank	Allele Range
CSF1PO	5q33.3-34	X14720	12	6-16
FGA	4q28	M64982	21	15-51.2
TH01	11p15.5	D00269	9	3-14
TPOX	2p23-pter	M68651	11	6-13
VWA	12p12-pter	M25858	18	10-24
D3S1358	3p	NT_005997	18	9-20
D5S818	5q21-31	G08446	11	7-16
D7S820	7q11.21-22	G08616	12	6-15
D8S1179	8	G08710	12	8-19
D13S317	13q22-31	G09017	13	5-15
D16S539	16q24-qter	G07925	11	5-15
D18S51	18q21.3	L18333	13	7-27
D21S11	21q21	AP000433	29	24-38

Figura 15: CODIS STRs. Denominación del locus, localización en el cromosoma, código de acceso del GenBank, código de referencia en el GenBank y su rango. Modificado de Butler (2004).

Cabe destacar que no todos los marcadores seleccionados en CODIS presentan el mismo grado de polimorfismo. Así, D21S11, D18S51 y FGA son los más polimórficos frente a CSF1PO, TH01, D16S539 y TPOX, con un menor número de variaciones entre individuos (Butler, 2011)

(Figura 16). La razón por la cual estos últimos fueron seleccionados se debió a que en ese momento la mayor base de datos de STR de Estados Unidos se encontraba en el Estado de Alabama, y contenía dichos marcadores (Butler, 2011).

Locus	Alleles		Range
D8S1179	14	14	<9 - >17
D21S11	30	30	<24.2 - >36
D7S820	13	10	6 - >14
CSFIPO	11	11	<6 - 15
D3S1358	17	17	<12 - >19
THO1	6	9.3	<5 - >10
D13S317	9	13	<8 - >15
D16S539	11	9	<8 - 15
VWA	18	14	11 - >22
TPOX	9	10	<6 - >13
D18S51	12	14	<11 - >22
D5S818	12	12	<7 - >15
FGA	23	23.2	<18 - >30

Figura 16: Número de alelos y rango de los 13 marcadores pertenecientes al CODIS. Modificado a partir de <https://dnaconsultants.com/codis-markers/10/03/2022>.

En 2010, tras años de aumento del número de muestras en la base del FBI, este organismo creó el Grupo de trabajo *CODIS Core Loci*, con el objetivo de sopesar la necesidad de una ampliación del número de marcadores en dicho sistema, con el fin de reducir la posibilidad de coincidencias erróneas, promover la compatibilidad internacional para compartir datos y mejorar la asistencia para búsqueda familiar y de personas desaparecidas (Lee et al., 2011). En el año 2012, este mismo grupo publicó una recomendación para ampliar el número de STR en el sistema, identificándose hasta 23 marcadores separados en un grupo principal y otro secundario. El grupo principal estaría compuesto por los 12 miembros originales (menos TPOX, presumiblemente por su bajo nivel de polimorfismo), 4 *loci* recomendados por incluir bases de datos de muestras europeas (D1S1656, D2S441, D10S1248 y D12S391), el locus PentaE, el DYS391 y el marcador de la Amelogenina. Por otro lado, en el grupo secundario se incluyeron TPOX, PentaD y los *loci* D22S1045 y SE33 (Hares, 2012) (Tabla 1).

Tabla 1: Locus, localización cromosómica y código de referencia en el GenBank de los 23 STR pertenecientes al sistema CODIS, separados en grupo principal y secundario.

Principal		
Locus	Localización	GenBank Accession
D8S1179	8q24.13	G08710
D21S11	21q21.1	AP000433
D7S820	7q21.11	G08616
CSF1PO	5q33.1	X14720
D3S1358	3p21.31	AC099539
TH01	11p15.5	D00269
D13S317	13q31.1	G09017
D16S539	16q24.1	G07925
vWA	12p13.31	M25858
D18S51	18q21.33	X91254
D5S818	5q23.2	G08446
FGA	4q28	M64982
D1S1656	1q42	G07820
D2S441	2p14	AC079112
D10S1248	10q26.3	AL391869
D12S391	12p12	G08921
PentaE	15q26.2	AC027004
DYS391	Yq11.21	G09613
Amelogenina	Xp22.1	M55418/M55419
Secundario		
Locus	Localización	GenBank Accession
TPOX	2p25.3	M68651
PentaD	21q22.3	AP001752
D22S1045	22q12.3	AL022314
SE33	6q14	V00481

Además, para la mencionada selección de marcadores se consideraron varios criterios, como la ausencia de relación con algún fenotipo o condición médica. Esto es debido a que el sistema CODIS está limitado por un marco legal particular, definido por la Ley de identificación de ADN de 1994, la Ley de eliminación de la acumulación de análisis de ADN del 2000, la Ley de Justicia para Todos del 2004 y la Ley de Huellas Dactilares de ADN del 2005, todas ellas aprobadas por ambas cámaras del Congreso de Estados Unidos (Katsanis et al., 2013). Por otro lado, tanto los criterios de selección, la posible relación (y su grado) con ciertos fenotipos, así como su posible ligamiento, no estaban totalmente claros (Ge et al., 2012; Katsanis et al., 2013). Por este motivo, Katsanis et al. (2013) realizaron una caracterización de ambos grupos CODIS con el fin de esclarecer estas incógnitas. Este grupo de investigación determinó que ninguno de los marcadores analizados mostraba relación alguna con fenotipos conocidos y, aunque los

marcadores permiten la identificación de individuos, no exponen información sensible relevante o biomédica. Desde ese momento se utilizaron los STR del sistema CODIS de manera extensiva, no sólo en el contexto forense (identificación de individuos), sino para la caracterización de poblaciones (Aguiar et al., 2014; Algee-Hewitt et al., 2016; Budowle et al., 1999; Fan et al., 2019; Kumar et al., 2021; Noris et al., 2012; Rodovalho et al., 2015; Roy et al. 2008; Untoro et al., 2009; Zhou et al., 2020).

I.5 Marcadores de linajes masculinos y femeninos

Marcadores del Cromosoma Y

El cromosoma Y presenta una longitud aproximada de 50 Mb siendo uno de los cromosomas más pequeños del genoma humano, apenas superando a los cromosomas 21 y 22 de 47 Mb y 49 Mb, respectivamente. La primera vez que se publicó su secuencia completa fue el 19 de junio del 2003 en la revista *Nature*, por investigadores del Instituto Whitehead y de la Universidad de Washington. Ese mismo año, Skaletsky et al. (2003), esclarecieron que la secuencia no era realmente completa, ya que sólo 23 Mb de los aproximadamente 50Mb que componen el cromosoma habían sido analizados. Los 27 Mb restantes se encuentran en una región de heterocromatina localizada en el brazo q, con una alta cantidad de secuencias repetidas y sin capacidad de transcripción (Butler, 2011) (Figura 17).

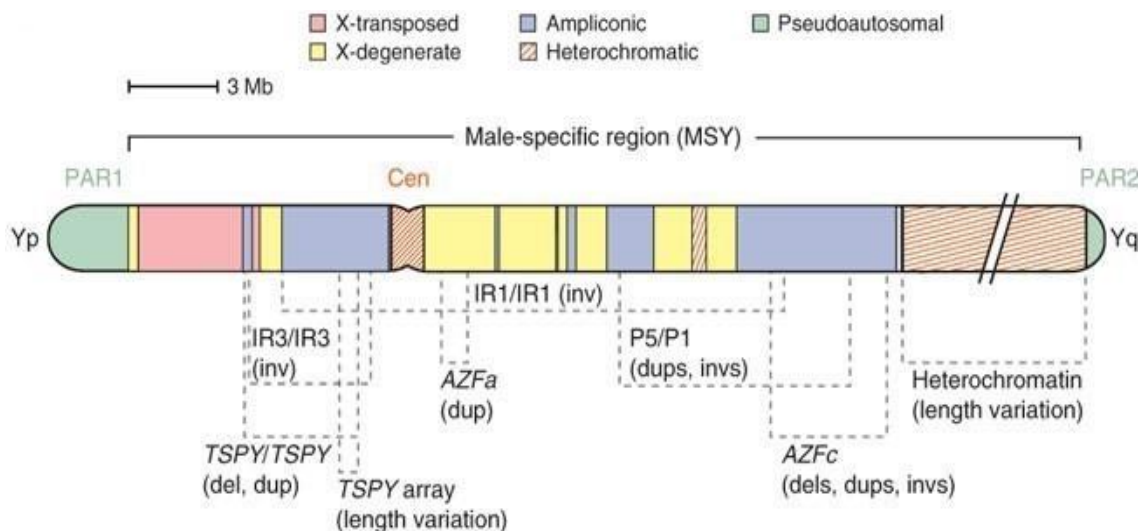


Figura 17: Estructura del cromosoma Y. Modificado de Repping et al. (2006).

La estructura del cromosoma Y está compuesta por una región principal denominada *Male-specific region* (MSY) que permanece intacta en la herencia entre padres e hijos, salvo que

se produzca una mutación; esta región presenta un grado alto de identidad que se mantiene por una frecuente conversión de genes mediante una transferencia no recíproca (Skaletsky et al., 2003). A su vez esta región central está flanqueada por dos regiones PAR o pseudoautosómicas; dichas regiones son homólogas a las encontradas en el cromosoma X. La región PAR localizada en el brazo Yp se denomina PAR1 y presenta una longitud aproximada de 2,5 Mb frente a la PAR2 (ubicada en el brazo Yq) de tan solo 1Mb (Butler, 2011; Graves et al., 1998) (Figura 17). En lo que respecta a la diversidad del cromosoma Y, suponiendo que la proporción de ambos sexos en una población es 1:1, cada pareja presentaría cuatro copias de cada autosoma, tres cromosomas X, pero sólo un cromosoma Y (Thomson et al., 2000). Asumiendo que todos los cromosomas sufren los mismos procesos de mutación se esperaría que la diversidad en el cromosoma Y fuera más baja que en el resto, hecho que ha sido contrastado empíricamente (Thomson et al., 2000). Por otro lado, la diferenciación entre los haplogrupos del cromosoma Y está acelerada por causa de la deriva genética pues es especialmente susceptible a este agente evolutivo; esta propiedad es especialmente útil para investigar eventos del pasado (Jobling y Tyler-Smith, 2003).

Los marcadores del cromosoma Y se pueden dividir en bialélicos, como los polimorfismos de un solo nucleótido del cromosoma Y (Y-SNP), o las inserciones *Alu* localizadas en este cromosoma; a su vez, estos marcadores pueden definir haplogrupos (Sahakyan et al., 2021). Por otro lado, estarían los marcadores polialélicos, como las repeticiones cortas en tándem o Y-STR, utilizadas para diferenciar haplotipos debido a que sus frecuencias de mutación son más altas, cerca de $1 \cdot 10^{-3}$, que las encontradas en los Y-SNP (Navarro-López et al., 2021; Staidig et al., 2021). Los Y-SNPs han jugado un papel muy importante en los estudios de poblaciones humanas ya que se caracterizan por tener una tasa de mutación baja, aproximadamente de $2 \cdot 10^{-8}$, siendo eventos únicos heredables por vía del linaje paterno (Navarro-López et al., 2021). Su análisis permite la asignación a un linaje en concreto, definiendo los haplogrupos mayoritarios de cada población y caracterizándola (Petrejčíková et al., 2014). Esta característica ha permitido realizar estudios de genética de poblaciones siguiendo únicamente la línea del linaje paterno, siendo de especial interés en aquellas poblaciones cuyas características, históricas o sociales, generan una divergencia entre los sexos (Martínez-Cruz et al., 2011, 2012, 2016). Por otro lado, también presentan utilidad para la realización de estudios sobre patrones migratorios de linajes paternos (Jobling y Tyler-Smith, 2003; Navarro-López et al., 2021). Su uso junto con marcadores del ADN mitocondrial (ADNmt), que mostrarían la línea del linaje materno, permite analizar la acción de los diferentes agentes microevolutivos, deriva y flujo génico, en relación a ambos sexos (Semino et al., 2000; Martínez-Cruz et al., 2016).

La nomenclatura de los haplogrupos del cromosoma Y ha sido propuesta por la Sociedad Internacional de Genealogía Genética (ISOGG), aunque la base surgió en el año 2002 en el Consorcio del Cromosoma Y y posteriormente fue actualizada por Karafet et al. (2008). En el Consorcio se propusieron dos nomenclaturas complementarias, la primera basada en el linaje y la segunda en la mutación. La primera nomenclatura se caracteriza por una letra mayúscula inicial que iría de la A a la R (S en las versiones más actuales). Además, si dos haplogrupos tienen un linaje muy próximo se les mantendría la primera letra y se les añadiría un valor numérico creciente. Posteriormente, descendiendo en la filogenia, dos haplogrupos compartirían tanto la primera letra como el número y se les añadiría una letra minúscula en orden alfabético. Siguiendo esta cadena, para casos de linajes muy cercanos, se mantendría el código anterior añadiendo un valor numérico creciente. Por último, para las filogenias más próximas, se seguiría la cadena “letra minúscula número”, sucesivamente (Figura 18).

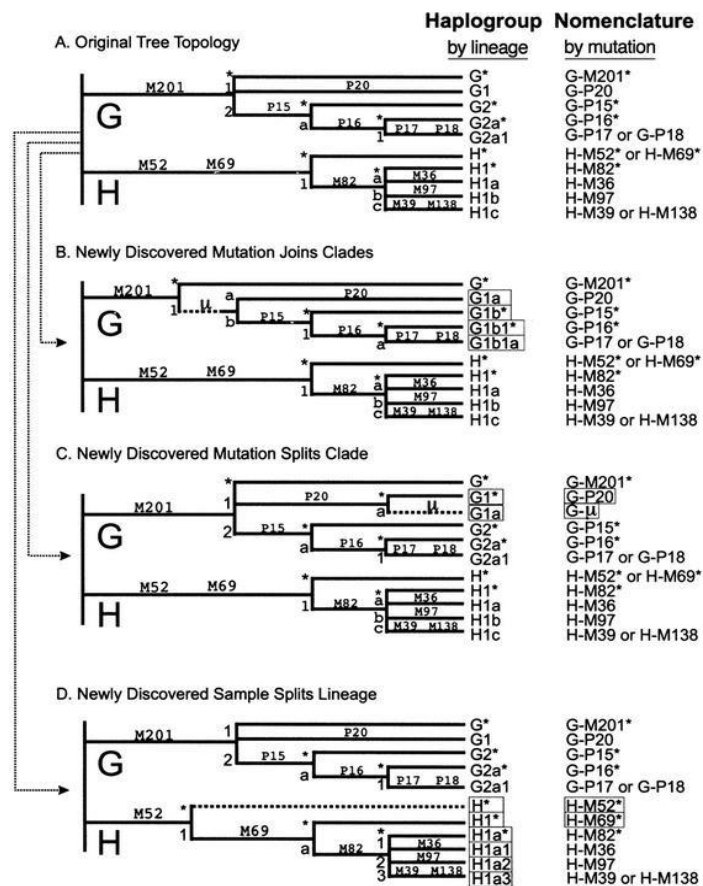


Figura 18: Ejemplos de nomenclatura de haplogrupos del cromosoma Y en función de su linaje o mutación. Tomado del Y Chromosome Consortium (2002).

La segunda nomenclatura mantiene en primer lugar, como en el caso anterior, la letra mayúscula inicial y a continuación se le adjunta el nombre de la mutación terminal que define

al haplogrupo separado por un guión (Figura 18). En caso de que un haplogrupo sea definido por más de un marcador se determina uno de ellos. Esta nomenclatura, aunque omite cierta información, también tiene menor tendencia a cambios en su topología (Y Chromosome Consortium, 2002). Los estudios basados en Y-SNP han mostrado ser útiles para los análisis de migración, mestizaje entre poblaciones y definición de etnias, todo ello bajo el prisma del linaje paterno (Larmuseau et al., 2017; Fehér et al., 2015; Petrejčíková et al., 2014; Geppert et al., 2011; Wang et al., 2021, 2022; Palencia-Madrid et al., 2022; Tamburrini et al., 2021; Tätte et al., 2021).

Por último, la distribución de los haplogrupos del cromosoma Y está limitada al clado M168 en la mayoría de continentes, con la excepción del africano, presentando un único linaje fundador hace entre 40-60 mil años coincidente con el modelo *Out of Africa* (Navarro-López et al., 2021) (Figura 19). De igual manera se acepta que las raíces de las poblaciones europeas se encuentran en poblaciones dispersas de cazadores-recolectores de hace unos 43.000 años que, posteriormente, hace unos 10.000 años, migraron a Europa tras el auge de la agricultura en Próximo Oriente. Sin embargo, estas poblaciones también muestran la presencia de subgrupos más actuales derivados del M168, lo que refleja posiblemente un flujo génico posterior a esta migración (Navarro-López et al., 2021; Roewer et al., 2005). La distribución en la actualidad de los haplogrupos del cromosoma Y en Europa se muestra en la Figura 20. También en el periodo de finales del Pleistoceno otras poblaciones migraron desde Próximo Oriente y se fueron distribuyendo por el continente asiático, asentándose en el subcontinente indio. Este asentamiento temprano en dicho subcontinente quedó registrado en su acervo genético, de forma que se observan linajes paternos con gran arraigo en las poblaciones indias, según los estudios basados en marcadores bialélicos y microsatélites del cromosoma Y (Khurana et al., 2014). A nivel mundial, la filogenia y distribución actual de los haplogrupos del cromosoma Y se puede observar en la Figura 21.

Introducción

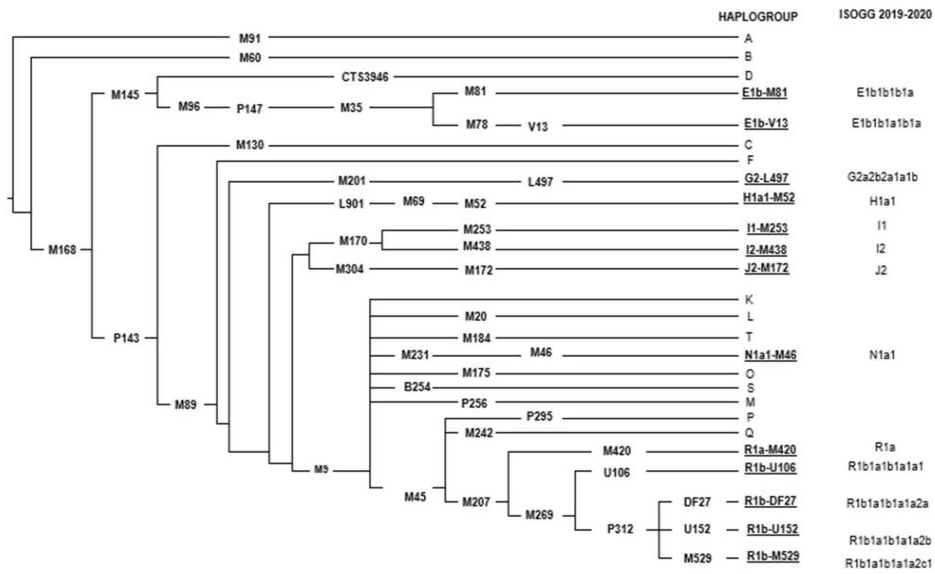


Figura 19: Árbol filogenético del cromosoma Y. Tomado de Navarro-López et al. (2021).

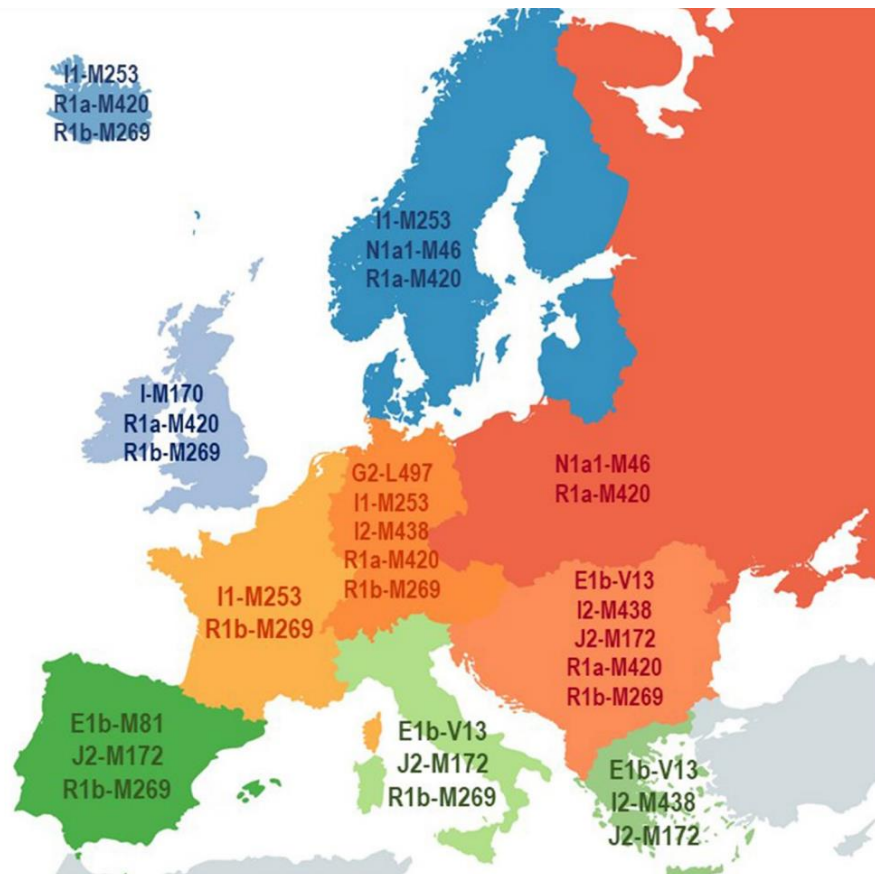


Figura 20: Distribución de los haplogrupos del cromosoma Y en el continente europeo. Tomado de Navarro-López et al. (2021).

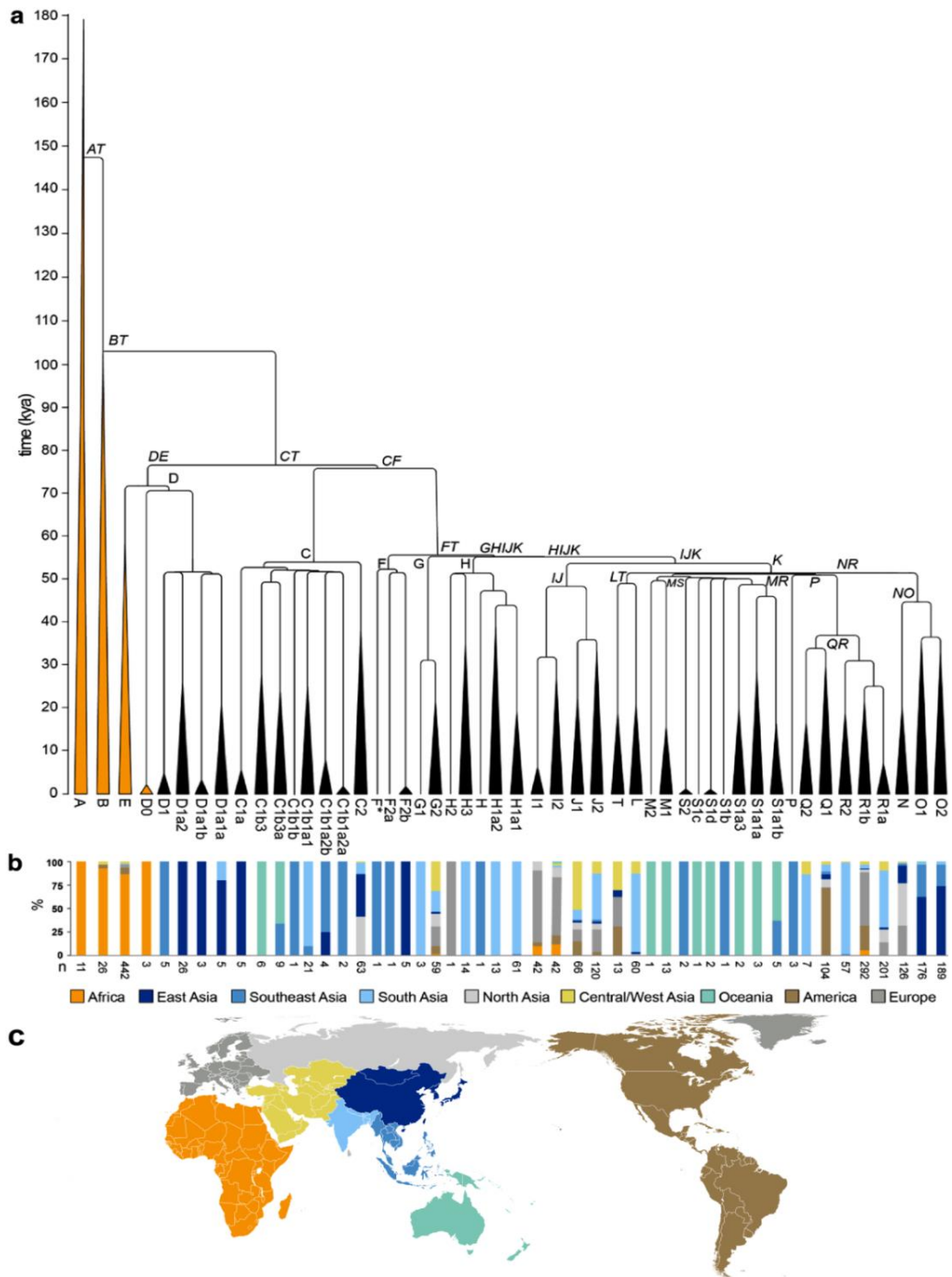


Figura 21: Distribución de haplogrupos y filogenia del cromosoma Y. a) Filogenia de máxima verosimilitud del cromosoma Y. b) Proporción de cada haplogrupo en cada continente. c) Representación de la división geográfica. Basado en Hallast et al. (2021).

Marcadores del ADN mitocondrial (ADNmt)

Las mitocondrias son un orgánulo de doble membrana localizado en el citoplasma celular. Producen el 90% de la energía celular en forma de Adenosin Trifosfato (ATP) por medio del proceso metabólico denominado fosforilación oxidativa, que oxida nutrientes, y con la energía resultante sintetiza ATP (Chen et al., 2009). Las mitocondrias también participan en diferentes vías de señalización como el ciclo de Krebs, la β oxidación de ácidos grasos y la regulación del Calcio (Mishra y Chan, 2016; Yan et al., 2019). Además, tienen un papel en la regulación del ciclo celular o en la apoptosis (Arakaki et al., 2006; Desagher y Martinou, 2000; Yan et al., 2019). Si se les compara con otros orgánulos celulares, las mitocondrias presentan una característica particular: tienen material genético propio denominado ADN mitocondrial (ADNmt). Dicho ADN tiene una longitud aproximada de unas 16.500 pb en forma circular, donde se codifican 13 subunidades del sistema de fosforilación oxidativa, 2 ARN ribosómicos (ARNr) y 22 ARN de transferencia (ARNt) (Pakendorf y Stoneking, 2005) (Figura 22). Con la excepción de cerca de 1100 pb de bases con función reguladora, el resto de la secuencia es codificante. A la región no codificante se le denomina región control (Pakendorf y Stoneking, 2005). El ADNmt se encuentra empaquetado por una variedad de proteínas, entre ellas prohibitinas, formando un complejo ADNmt-proteína denominado nucleoide (Yan et al., 2019). Los defectos en la distribución del ADNmt dentro del orgánulo suelen estar asociados a varias patologías, entre ellas procesos carcinogénicos (Stefano et al., 2016).

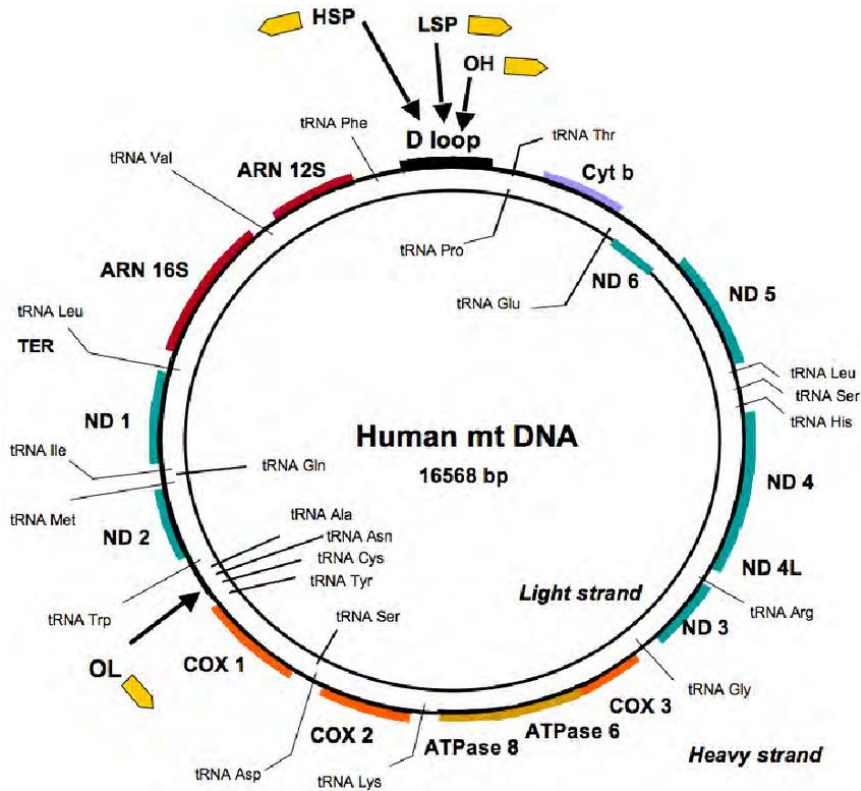


Figura 22: Representación de ADN mitocondrial humano. Tomado de Bellance et al. (2009).

En lo que respecta a la herencia de este material genético, múltiples estudios demuestran que se hereda por línea matrilineal (por ejemplo, Giles et al., 1980; Habbane et al., 2021; Mishra et al., 2020; Picard et al., 2021). Son varios los mecanismos por el cual el ADNmt paterno es eliminado. En primer lugar, la proporción de este material genético aportada por el espermatozoides y el óvulo está desbalanceada con una ratio de aproximadamente 1 por cada 15.800 moléculas. En segundo lugar, durante la fertilización del óvulo ocurre un proceso de mitofagia por el cual son eliminadas las mitocondrias paternas restantes. Todo ello tiene como consecuencia que el ADNmt paterno sea indetectable en la etapa de 4 a 8 células (Rossmann et al., 2021). Cabe destacar que, en el año 2002, Schwartz y Vissing publicaron un artículo sobre el caso de un hombre con cierta patología física, y tras el estudio de su ADNmt se descubrió que éste tenía un origen mayoritariamente paterno, rompiendo con lo establecido hasta el momento. Estudios posteriores plantearon algunos mecanismos de cómo esta herencia podría ser posible (Luo et al., 2018), pero terminaron considerando que la herencia por línea paterna era muy poco probable. Debido a ello, la herencia materna del ADNmt se sigue considerando como la regla (Luo et al., 2018; Schwartz et al., 2004; Vissing et al., 2019).

En el contexto de los estudios de genética de poblaciones esta familia de marcadores presenta de un gran interés por su capacidad de rastrear los eventos ocurridos en una población

desde el punto de vista del linaje materno. Así, su análisis conjunto con marcadores del cromosoma Y (que permiten analizar la línea paterna) en poblaciones cuyas características o historia presentan un sesgo entre los sexos, permite valorar el efecto diferencial de los agentes microevolutivos. De hecho son múltiples los estudios que los utilizan tanto de manera conjunta o individual en los análisis de poblaciones (Asari et al., 2007; Bhatti et al., 2018; Chen et al., 2015; Farha et al., 2016; Johnson et al., 2015; Martínez-Cruz et al., 2016; Mitchell et al., 2014; Park et al., 2017; Raff et al., 2011; Ren et al., 2008; Saha et al., 2021; Semino et al., 2000; Simão et al., 2018; Tran et al., 2020). Por otra parte, dado su modo de herencia, la única manera por la cual el ADNmt puede cambiar es por causa de mutaciones sucesivas (SNPs) a lo largo de la radiación de los diferentes linajes maternos. Esta característica ha permitido estimar las migraciones diferenciales por sexos en las poblaciones y superponer las afinidades genéticas entre poblaciones indígenas con su ubicación geográfica, cartografiando así la diáspora de las poblaciones humanas (Wallace, 2013) (Figura 23). Por otro lado, estas mutaciones únicas, heredadas de un ancestro común, han permitido definir haplotipos del ADNmt. Cada población presentaría un grupo limitado de haplotipos en un inicio pero, posteriormente, con el crecimiento de la población, las sucesivas generaciones presentarían variantes de estos haplotipos. Estas nuevas variantes haplotípicas estarían relacionadas entre sí al compartir un origen común, formando los haplogrupos del ADNmt (Wallace, 2013).

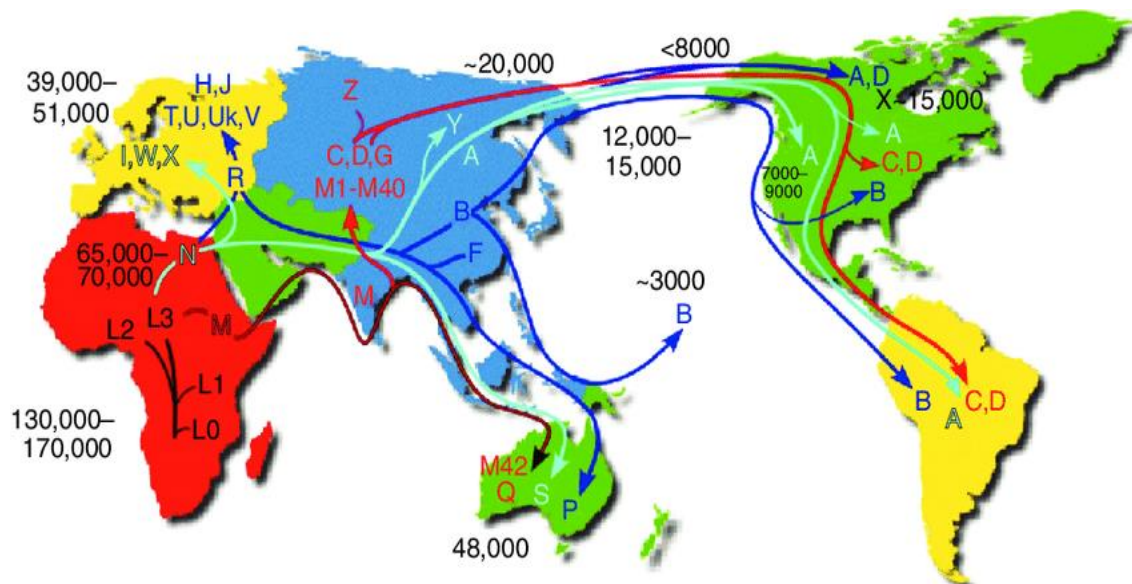


Figura 23: Mapa de la migración de las poblaciones humanas en función de sus haplogrupos del ADNmt. Según Wallace (2013).

Respecto a la evolución de los haplogrupos, entre hace 130.000 y 170.000 años surgió en África el primer haplogrupo de ADNmt humano o L0. Posteriormente surgieron el resto de haplogrupos L, previo al proceso *Out of Africa*. En el norte del continente, derivados de los haplogrupos L, surgieron los haplogrupos M y N, siendo los únicos que salieron del continente hace 65.000 años (Wallace et al., 2013; Wallace, 2013). La rama del linaje M avanzó por la costa asiática colonizando el subcontinente indio y posteriormente alcanzando Malasia y Australia, donde surgió el haplogrupo Q, hace unos 48.000, a partir del M. En lo que respecta al haplogrupo N este migró a Europa derivando durante el proceso el haplogrupo R. A su vez, del R derivaron los haplogrupos I, W y X (entre hace 39.000-51.000 años). Tanto el haplogrupo N como el R migraron hasta Australia generando los haplogrupos S y P, respectivamente. Durante esa migración surgieron los haplogrupos B y F, siendo el primero el que colonizó las islas del Pacífico hace unos 3000 años.

También, en la migración hasta Australia del haplogrupo N aparecieron los haplogrupos Y y A. Hace 20.000 años tanto el haplogrupo A derivado del N y los C y D derivados del M colonizaron Siberia y posteriormente, a través de Beringia, se difundieron por el Nuevo Mundo hace unos 15.000 años. En el continente americano se produjo una radiación Norte-Sur de dichos haplogrupos, aunque posteriormente tanto el X como el B alcanzaron las costas, pese a no estar presentes en Siberia. Cabe destacar una migración posterior de los haplogrupos A y D hace 6.000 años, a través del estrecho de Bering, producido por las migraciones de poblaciones Inuit y aleutianas. Por último, queda resaltar que en el continente americano se produjeron varias combinaciones entre los haplogrupos que lo colonizaron, generando a su vez varios subhaplogrupos. Estas observaciones fueron realizadas por el grupo de Wallace en 2013 y se muestran en la Figura 23. En la actualidad, las consecuencias de su historia evolutiva se pueden observar en la distribución continental de la Figura 24. Por un lado, en el continente africano se localizan mayoritariamente los haplogrupos L y, de igual manera, los haplogrupos N y R en Eurasia. Como última anotación la mayor diversidad de haplogrupos se encontraría en el continente africano y la menor en las poblaciones de Nativos Americanos (Kivisild, 2015).

Introducción

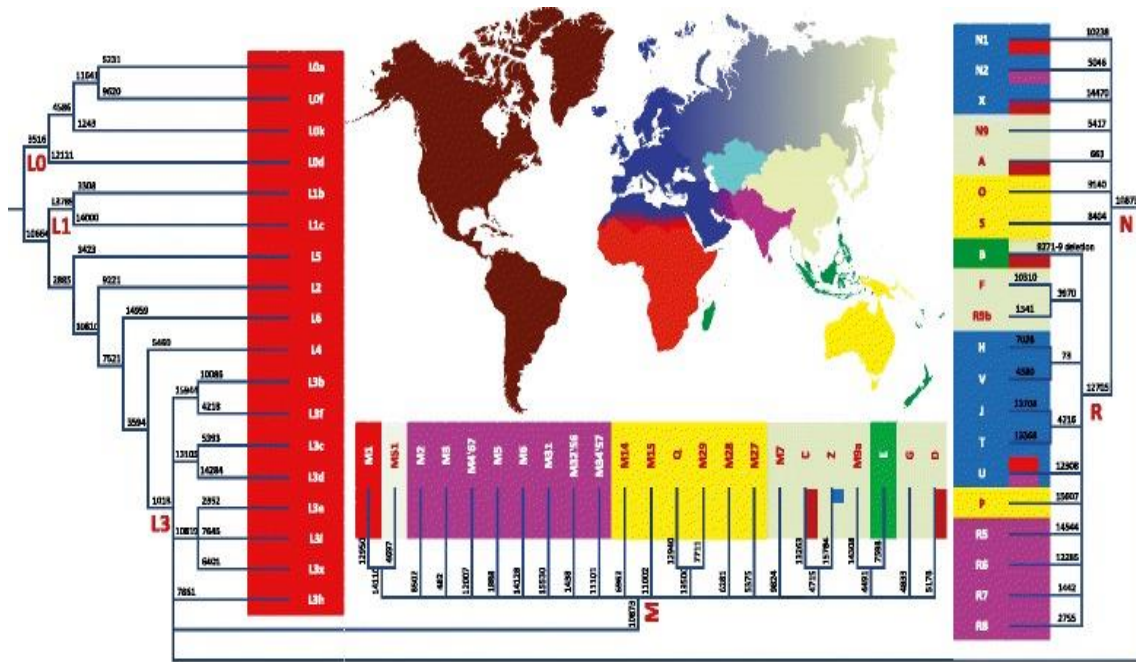


Figura 24: Árbol y distribución de los haplogrupos del ADNmt. Tomado de Kivisild (2015).



Hipótesis y Objetivos

II. Hipótesis de trabajo

La hipótesis general de esta Tesis doctoral se basa en la siguiente premisa: *“Las características socioculturales e históricas pueden afectar, a una escala microevolutiva, al patrimonio genético de una población”*.

Partimos de la base de que, en una población determinada, factores socioculturales tales como el predominio de matrimonios intra-poblacionales pueden promover la aparición de poblaciones aisladas. A su vez, eventos migratorios sucesivos en su registro histórico pueden generar efectos fundadores y cuellos de botella, afectando a su diversidad genética y desembocando en la formación de subpoblaciones (efecto de la deriva genética). El aislamiento poblacional y la formación de subpoblaciones afectarían a la acción de los principales agentes microevolutivos, la deriva genética y el flujo génico, pudiendo originar tras varias generaciones procesos de microdiferenciación entre las distintas subpoblaciones.

En el caso de la población gitana residente en el País Vasco, y en función de los marcadores utilizados e incluso del sexo, es probable la identificación de sus orígenes indios, así como del mestizaje con otras poblaciones receptoras de origen europeo. De igual modo, este grupo étnico también habrá experimentado los efectos de la deriva en su patrimonio genético.

III. Objetivos

Con el objetivo de analizar el efecto de estas presiones microevolutivas se ha estudiado a una muestra de población perteneciente a la etnia romaní, que reúne las condiciones anteriormente mencionadas, en concreto una muestra de población romaní residente en el País Vasco. Se han analizado asimismo muestras de otras poblaciones europeas de la misma etnia, a partir de una serie de bases de datos recogidas en la Tabla 1 del Anexo, así como diversas poblaciones europeas y del sudeste asiático, con el fin de situar en un determinado contexto geográfico y genético a la población objeto de estudio.

III.1. Objetivo general

El objetivo general de esta Tesis Doctoral ha sido valorar los efectos de la deriva y el flujo génico sobre la diversidad genética de una población romaní residente en el País Vasco, que han podido actuar sobre dicha población debido a sus particulares características socioculturales e históricas. Se trata por tanto de un estudio microevolutivo para el cual la etnia romaní presenta algunas características que la hace idónea para este tipo de estudios. En primer lugar, y como

ya se ha mencionado anteriormente, se considera la región del Punjab, en el Noroeste de la India, como cuna u origen de esta etnia. Debido a ello, los romaníes comparten similitudes culturales con poblaciones de la zona, tales como el sistema de castas, los matrimonios intrapoblacionales y la disparidad en la jerarquía de ambos sexos. Las dos primeras características son de interés como posibles barreras al flujo génico, mientras que la última abriría la posibilidad de analizar los linajes de manera independiente. Además, las poblaciones romaníes sufrieron sucesivos eventos migratorios desde esta región hacia el continente europeo, hecho que les dota de un gran valor para el análisis de los efectos fundadores sobre su diversidad genética. Tras su asentamiento se dividieron en subpoblaciones reduciendo su tamaño poblacional, lo que las convierte en buenas candidatas para los estudios sobre deriva genética, a la que son muy susceptibles.

III.2. Objetivos específicos

Según todo lo anterior se han propuesto una serie de objetivos específicos:

- I. Valorar el efecto de la cultura de esta etnia como posible barrera al flujo génico entre la población romaní objeto de estudio y su población huésped, mediante el análisis de dos grupos de marcadores autosómicos diferentes (SNPs e inserciones *Alu*). *Los resultados obtenidos se recogen en los apartados Análisis de la población romaní del País Vasco para AIMs y Análisis de la población romaní del País Vasco para inserciones Alu.*
- II. Medir las posibles consecuencias de la deriva sobre la diversidad genética de la población romaní del País Vasco, dado que existe posiblemente un cierto grado de aislamiento y un tamaño poblacional efectivo reducido. *Los resultados se muestran en los mismos apartados mencionados en el objetivo anterior.*
- III. Evaluar las proporciones de mestizaje entre la población romaní y su huésped, ambas residentes en una misma región geográfica pero presumiblemente separadas por una barrera socio-cultural. *Los resultados se muestran al final de los apartados anteriormente mencionados.*
- IV. Analizar los efectos del flujo génico, la deriva genética y el mestizaje sobre una serie de poblaciones romaníes europeas, mediante el análisis conjunto de diversos STRs autosómicos y valorar los resultados obtenidos en relación a la población romaní del País Vasco. *Este objetivo ha sido desarrollado en el apartado Análisis de las poblaciones romaníes europeas para CODIS.*
- V. Examinar las consecuencias del flujo génico, la deriva genética y el mestizaje en una serie de poblaciones romaníes europeas, diferenciando por linajes (masculinos y femeninos),

mediante el análisis del cromosoma Y y de haplogrupos del ADNmt, con el fin de ampliar los resultados obtenidos sobre la población romaní objeto de estudio. *Los resultados obtenidos de estos análisis se presentan en el apartado Análisis de las poblaciones romaníes europeas para marcadores de linaje.*



Materiales y Métodos

IV. Materiales y Métodos

IV.1. Materiales

IV.1.1 Poblacionales analizadas

Se han extraído una serie de muestras de ADN de una población de origen romaní asentada en el País Vasco, en la que se han analizado dos grupos de marcadores genéticos: SNPs e inserciones *Alu*. Además, se han tomado datos de la bibliografía (de distintas bases de datos genéticos poblacionales descritas en la Tabla 1 del Anexo), que se han analizado con el fin de completar el abanico de marcadores genéticos incluidos en esta memoria, incluyendo STRs, cromosoma Y y ADN mitocondrial, con el objetivo de extrapolar los resultados a los obtenidos en la población romaní objeto de estudio.

Romaníes del País Vasco

Las muestras de ADN provienen de 94 individuos no relacionados, pertenecientes a la etnia romaní y residentes en el País Vasco. En estas muestras se analizaron un conjunto de SNPs y un grupo de inserciones *Alu*. Todos los individuos fueron voluntarios sanos y debidamente informados, tanto del proceso como del uso de la información obtenida a partir de su ADN. Las muestras han sido obtenidas siguiendo un proceso de extracción no invasivo, cumpliendo las exigencias metodológicas, los estándares éticos y de seguridad definidos en el marco jurídico vigente recogidos, entre otras, tanto en la Ley Orgánica 15/1999 para la protección de datos de las personas físicas como posteriormente la Ley 14/2007 que constituye el marco jurídico de la investigación biomédica.

Inserciones Alu y AIMs en otras poblaciones

Con el fin de considerar el entorno geográfico de residencia de los romaníes actuales del País Vasco, su lugar de origen y sus posibles rutas de migración, se han incluido poblaciones tanto de Europa como de Asia, a partir de diversas bases bibliográficas. En primer lugar, se han incluido las frecuencias registradas en diferentes poblaciones de los mismos AIMs que han sido analizados en la población romaní residente en el País Vasco. De Europa se han incluido muestras de población vasca, española, francesa, británica, irlandesa, italiana, sarda, húngara, danesa, finlandesa, estonia, bielorrusa y ucraniana (Figura 25), a partir de la base de datos ALFRED (Cheung et al., 1999).



Figura 25: Mapa de Europa con la localización de las poblaciones no romaníes incluidas en la base de datos de AIMs.

En cuanto a las poblaciones asiáticas, se han incluido datos de los burusho, kalash, brahui y sindhi de Pakistán, de los balochi de Irán, de los hazara de Afganistán, de los telugu, punjabís y gujarati de la India, de los bengalíes de Bangladesh y de los tamil de Sri Lanka (Figura 26), obtenidos igualmente a partir la base de datos ALFRED (Cheung et al., 1999).



Figura 26: Mapa de Asia con la localización de las poblaciones no romaníes incluidas en la base de datos de AIMs.

En relación a las inserciones *Alu*, se han utilizado resultados obtenidos a partir de distintos estudios (Comas et al., 2000, 2004; Garcia-Obregón et al., 2006; Nasidze et al., 2001; Stoneking et al., 1997; Varzari et al., 2007). De Europa se han incluido las frecuencias de poblaciones españolas (vasco-navarra, vasco-guipuzcoana, catalana, andaluza, valenciana y canaria), francesa (bretona), suiza, ucraniana, moldava, rumana, macedonia, albanesa, griega y turca (Figura 27).

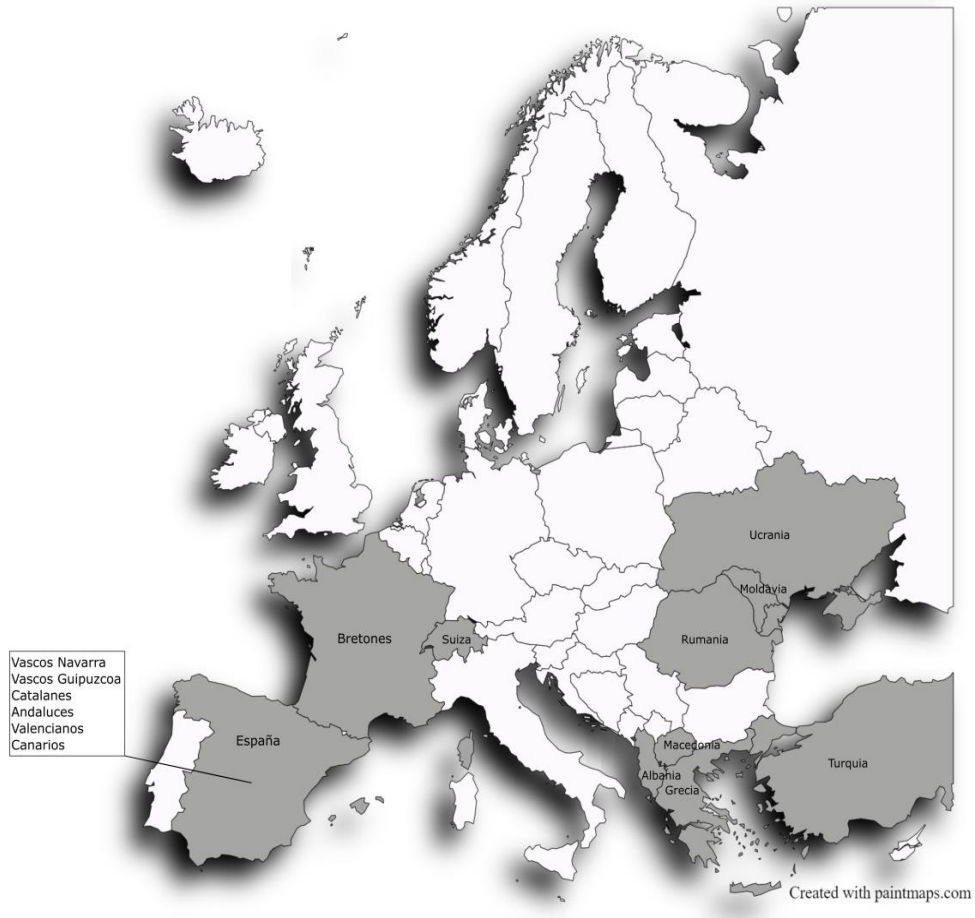


Figura 27: Poblaciones europeas no romaníes analizadas mediante inserciones *Alu*.

En lo que compete a las poblaciones asiáticas, estas incluyen las frecuencias de inserción obtenidas en los Emiratos Árabes Unidos (E.A.U.), de sirios, pakistaníes, indios (tamil, indio-cristianos, indio-hindús e indio-musulmanes) y de afganos (pashtunes) (Figura 28).



Figura 28: Poblaciones asiáticas no romaníes analizadas mediante inserciones *Alu*.

Con el fin de complementar el tamaño muestral con un mayor número de poblaciones de origen indio dada su importancia como origen de la etnia objeto de estudio, se han recopilado datos de 9 poblaciones que habían sido analizadas para 7 de las 8 inserciones *Alu* que conforman el presente estudio: TPA25, ACE, APO, PV92, FXIII B, B65 y D1. Las poblaciones procedían de la región de Uttar Pradesh en el norte del país, y estaban compuestas por siete muestras poblacionales que representan diferentes castas dentro de la India, y que incluyen a los Kayasthas, Mathurs, Rastogies, Vaish y Brahmins, además de dos subgrupos, los Bhargavas y Chaturvedis. Además, también incluían dos sectas musulmanas, la sunita y la chiita. Todas las poblaciones mencionadas fueron tomadas de la investigación realizada por Tripathi et al. (2008) (Figura 29).



Figura 29: Poblaciones indias complementarias analizadas mediante un grupo reducido (n=7) de inserciones *Alu*.

Por último, en lo que respecta a las poblaciones analizadas para inserciones *Alu*, se añadieron las frecuencias de una población del continente africano y otra del Este de Asia, pero únicamente para la realización de un dendrograma (y no para el resto de análisis), ya que dichas poblaciones no se encuentran en la ruta de la diáspora de la etnia romaní desde el subcontinente indio hasta la península Ibérica. El objetivo fue contrastar la posición de las poblaciones europeas y asiáticas en relación a la población ancestral (población con frecuencias de inserción nulas). Las poblaciones a las que se hace referencia son los !Kung del sur de Angola, muestra obtenida a partir de la base de datos de Stoneking et al. (1997) y los Han de China, a partir de la de Xiao et al. (2002) (Figura 30).

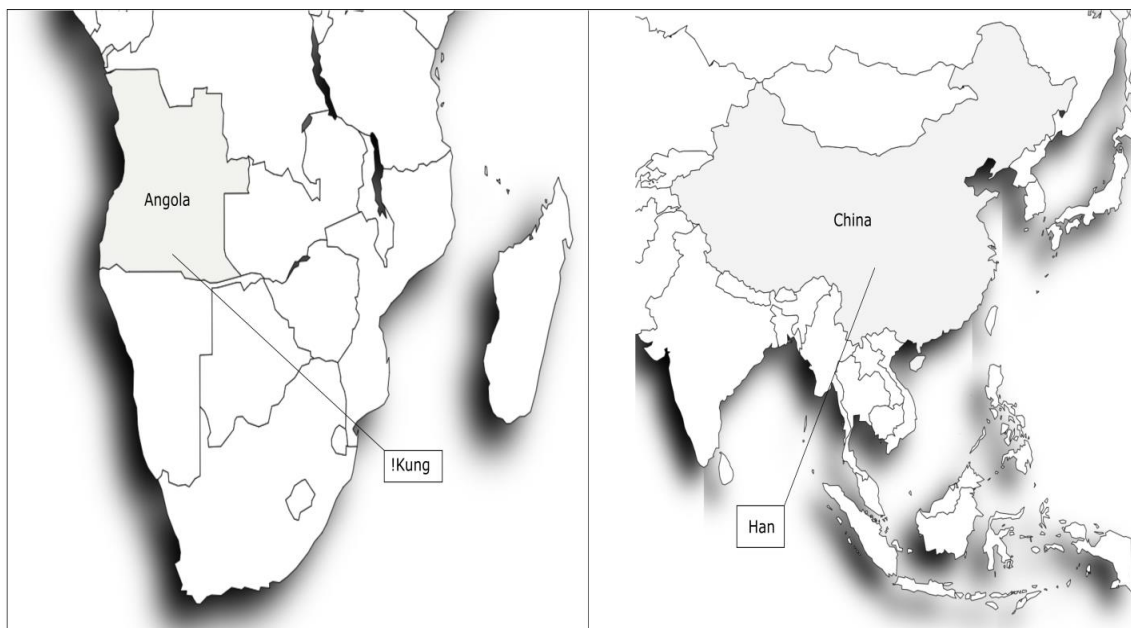


Figura 30: Poblaciones africana y del Este de Asia añadidas al conjunto de datos para la construcción de un dendrograma con las inserciones *Alu*.

Otros marcadores

CODIS

Por otra parte, se han analizado las frecuencias de los 13 STRs del sistema CODIS, también en poblaciones europeas, africanas y asiáticas. En relación a las poblaciones europeas se tomaron 10 poblaciones representativas de 5 países europeos (Macedonia del Norte, Eslovaquia, Croacia, Portugal y Hungría); cada país se compone de muestras de población general y de individuos romaníes (Havaš et al., 2007; Soták et al., 2008; Haliti et al., 2009; Novokmet et al., 2007; Cabezas Silva et al., 2016; Gusmão et al., 2010; Egyed et al., 2006; Füredi et al., 1996; Füredi et al., 1997; Egyed et al., 2000). Con el fin de ampliar la muestra comparativa se tomaron también datos de 6 poblaciones representativas de 4 países europeos, conformadas únicamente por muestras de población general (Francia, España, Italia y Polonia), con la excepción de Italia de la que se obtuvieron, además, una muestra de población general y otra Sarda, y de Francia que, además de población general, incluía población vasco-francesa (Amigo et al., 2009; Budowle et al., 2001; Kuzniar et al., 2004) (Figura 31).



Figura 31: Poblaciones europeas analizadas para 13 STRs del sistema CODIS.

De Asia se recopilaron datos genéticos de 9 poblaciones repartidas entre India, Irán y Pakistán (Figura 32). Por parte de la India se obtuvieron las frecuencias para la población teli de Maharashtra (N=69) (Badiye et al., 2021), la población rajput de Himachal Pradesh (N=179) (Sahajpal et al., 2019) y de la población general de Rajashtán (N=571) (Kumar et al., 2020). Respecto a la población iraní se obtuvieron las frecuencias de 100 individuos originarios de Fars, al suroeste del país (Hedjazi et al., 2013). Finalmente, en cuanto a las poblaciones pakistaníes se recopilaron las frecuencias obtenidas a partir del estudio de Shan et al. (2021). Dicha muestra se compone de poblaciones punjabí del Punjab Norte (N=106), saraiki del Punjab Sur (N=104), pastún de Khyber Pakhtunkhwa (N=107), sindhi de Sind (N=103), y balochis de Baluchistán (N=109).



Figura 32: Poblaciones asiáticas analizadas para 13 STRs del sistema CODIS.

Cromosoma Y y ADN mitocondrial

Se han recopilado las frecuencias en 19 haplogrupos del cromosoma Y, así como las variantes polimórficas de las regiones hipervariables HVS-I y HVS-II del ADN mitocondrial (ADNmt), de 14 poblaciones representativas de 7 países europeos (Bulgaria, Hungría, Eslovaquia, España, Grecia, Rumanía y Ucrania). En cada país se incluye una muestra de individuos romaníes y una muestra de población general. En conjunto, se han considerado un total de 1737 individuos; de ellos, 753 se definen como pertenecientes a la etnia romaní y 984 como no romaníes (Martínez-Cruz et al., 2016). Por otro lado, se han considerado 6 poblaciones generales europeas para el estudio de las frecuencias haplotípicas del cromosoma Y (Francia, Países Bajos, Alemania, Polonia, Macedonia y Albania) a partir de los trabajos de Semino et al.

(2000) y de Martínez-Cruz et al. (2016) (Figura 33), y 5 para el de las frecuencias del ADNmt (Francia, Alemania, Italia, Austria y Polonia) (Ottoni et al., 2011) (Figura 34).



Figura 33: Poblaciones europeas analizadas para los haplogrupos del cromosoma Y.



Figura 34: Poblaciones europeas analizadas para los haplogrupos del ADNmt.

En cuanto a las poblaciones asiáticas analizadas, se han recopilado frecuencias haplotípicas para el cromosoma Y de 20 poblaciones del sudoeste asiático, de las cuales 14 proceden de la India (Kivisild et al., 2003; Flores et al., 2005; Regueiro et al., 2006; Haber et al., 2012; Khurana et al., 2014) (Figura 35 y 36).



Figura 35: Poblaciones asiáticas analizadas para los haplogrupos del cromosoma Y.

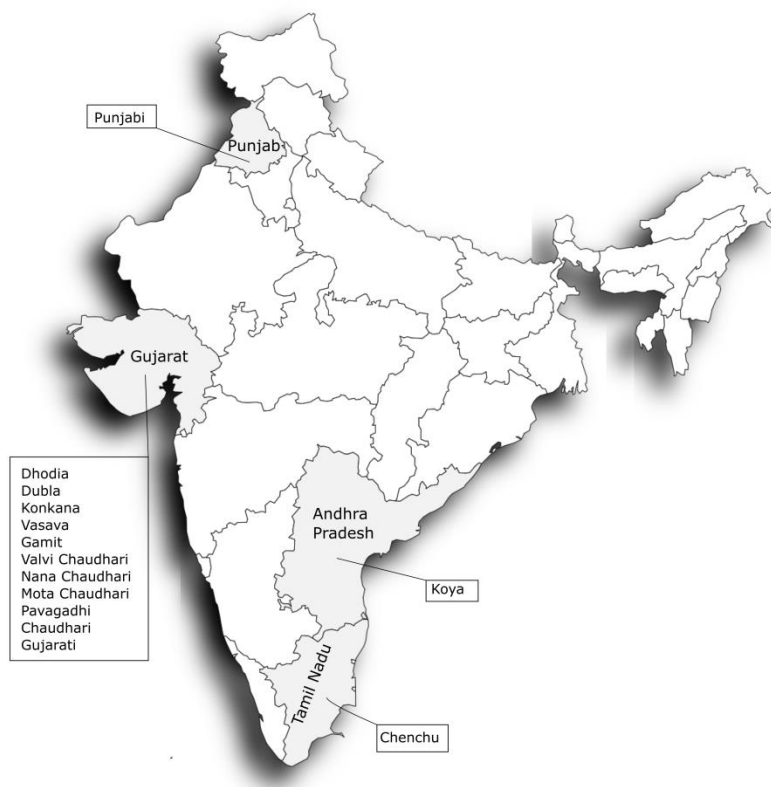


Figura 36: Poblaciones de la India analizadas para haplogrupos del cromosoma Y.

También se han obtenido las frecuencias para el segmento hipervariable HVS-1 y HVS-2 del ADNmt, a partir de los trabajos de Metspalu et al. (2004) y Ottoni et al. (2011), de 23 poblaciones repartidas en 10 países, que representan a diferentes castas y tribus, incluyendo a la población general (Figura 37). Entre los países considerados se encuentra la India, cuya muestra está compuesta por individuos de las regiones (de norte a sur) del Punjab, Rajasthán, Gujarat y Maharashtra (Figura 38).



Figura 37: Poblaciones asiáticas analizadas para los haplogrupos del ADNmt.



Figura 38: Poblaciones de la India analizadas para los haplogrupos del ADNmt.

IV.2. Métodos

IV.2.1. Extracción del ADN

El proceso de extracción de ADN de las muestras de la población romaní del País Vasco se realizó en nuestro laboratorio mediante el kit prepIT-L2P de DNA Genotek. El protocolo seguido se muestra a continuación:

- **Transferencia:** 500 µl de muestra a *Eppendorf* de 1.5 ml.
- **Incubación:** 50 °C (1h).
- **Purificación:** añadir 20 µl + vortex (3-5 seg).
- **Incubación:** hielo (10 min).
- **Centrifugación:** 13.000 rpm a 20 °C (10 min).
- **Sobrenadante:** Pipetear el sobrenadante a un nuevo *Eppendorf* de 1.5 ml y descartar el *pellet*.
- **Adición de Etanol 95-100%:** añadir 500 µl al sobrenadante y mezclar por inversión (10 inversiones).
- **Incubación:** a temperatura ambiente (10 min).
- **Centrifugación:** 13.000 rpm a 20 °C (10 min).
- **Descartar el sobrenadante:** sin tocar el *pellet* por el lado opuesto a la mancha de ADN.
- **Adición de Etanol 70%:** añadir 250 µl al *pellet* sin dañarlo. Esperar (1 min) y retirar el etanol apurando.
- **Rehidratación del ADN:** añadir 100 µl de H₂O. Vortex intenso hasta resuspender.
- **Incubación:** a 50 °C (1h) o a temperatura ambiente (1-2 días).
- **Almacenaje:** a -20 °C.

IV.2.2 Marcadores genéticos

Inserciones Alu

El análisis de los polimorfismos en inserciones *Alu* consistió inicialmente en una amplificación PCR de las muestras de estudio, para posteriormente visualizar los resultados mediante una electroforesis en gel de agarosa. Para la realización de la amplificación se usó un termociclador *Perkin Elmer GeneAmp PCR system 9700 (AppliedBiosystem)*. Las características específicas de cada PCR (concentraciones, número de ciclos y duración), *primers* de cada locus y la ubicación de cada marcador se muestran en las Tablas 2, 3 y 4.

Tabla 2: Concentraciones finales de cada reactivo empleado en cada PCR para los marcadores *Alu*.

Marcador <i>Alu</i> Reactivo	A25	PV92	B65	TPA25	FXIIIB	D1	ACE	APO
Buffer	1X	1X	1X	1X	1X	1,3X	1X	1X
MgCl₂	2 mM	1,5 mM	1,75 mM	1,25 mM	1,5 mM	2,25 mM	1,5 mM	1,5 mM
dNTPs	1,5 mM	1 mM	1,2 mM	1 mM	1,4 mM	1,5 mM	1 mM	1,5 mM
Glicerol	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
Primer F	0,6 μM	0,45 μM	0,5 μM	0,5 μM	0,6 μM	0,5 μM	0,45 μM	0,6 μM
Primer R	0,6 μM	0,45 μM	0,5 μM	0,5 μM	0,6 μM	0,5 μM	0,45 μM	0,6 μM
Taqpol	0,06 U/μl	0,05 U/μl	0,05 U/μl	0,04 U/μl	0,045 U/μl	0,06 U/μl	0,05 U/μl	0,05 U/μl

Tabla 3: Localización de cada *loci*, secuencia de los cebadores y tamaño de su amplificado.

Locus	Localización	Primers (5'-3')	Tamaño del amplificado
A25	8q21.3	F(5'-3')CCACAAATAGGCTCATGTAGAAC R(5'-3')TATAATATGGCCTGGATTATACC	Alu(+): 552pb Alu(-): 268pb
PV92	16q23.3	F(5'-3')AACTGGGAAAATTTGAAGAGAAAAGT R(5'-3')TGAGTTCTCAACTCCTGTGTGTTAG	Alu(+): 443pb Alu(-): 129pb
B65	11q14.2	F(5'-3')ATATCCTAAAAGGGACACCA R(5'-3')AAAATTTATGGCATGCGTAT	Alu(+): 423pb Alu(-): 81pb
TPA25	8p11.2	F(5'-3')GTAAGAGTTCCGTAACAGGACAGCT R(5'-3')CCCCACCCTAGGAGAACTTCTCTTT	Alu(+): 424pb Alu(-): 113pb
FXIIIB	1q31	F(5'-3')TCAACTCCATGAGATTTTCAGAAGT R(5'-3')CTGGAAAAAATGTATTCAGGTGAGT	Alu(+): 700 pb Alu(-): 410pb
D1	3q26.32	F(5'-3')TGCTGATGCCAGGGTTAGTAAA R(5'-3')TTTCTGCTATGCTCTCCCTCTC	Alu(+): 670pb Alu(-): 333pb
ACE	17q23	F(5'-3')CTGGAGACCACTCCCATCCTTTCT R(5'-3')GATGTGGCCATCACATTCGTGAGAT	Alu(+):490pb Alu(-):190pb
APO	11q23-q24	F(5'-3')AAGTGCTGTAGGCCATTTAGATTAG R(5'-3')AGTCTTCGATGACAGCGTATACAGA	Alu(+):409 Alu(-):97

Tabla 4: Programas de PCR para cada inserción *Alu*.

PROGRAMA PCR A25			
	Temperatura	Tiempo	Nº Ciclos
Desnaturalización	94°C	10 min	
Desnaturalización	94°C	1 min	x7
Anneling	60,2°C	1 min 30 s	
Extensión	72°C	1 min 30 s	
Desnaturalización	94°C	1 min	x25
Anneling	58°C	1 min	
Extensión	72°C	1 min	
Extensión Final	72°C	10 min	
PROGRAMA PCR PV92			
	Temperatura	Tiempo	Nº Ciclos
Desnaturalización	94°C	10 min	
Desnaturalización	94°C	1 min	x30
Anneling	57,8°C	1 min 30 s	
Extensión	72°C	1 min 30 s	
Extensión Final	72°C	10 min	
PROGRAMA PCR B65			
	Temperatura	Tiempo	Nº Ciclos
Desnaturalización	94°C	10 min	
Desnaturalización	94°C	1 min	x7
Anneling	60°C	1 min 30 s	
Extensión	72°C	1 min 30 s	
Desnaturalización	94°C	1 min	x28
Anneling	57,5°C	1 min	
Extensión	72°C	1 min	
Extensión Final	72°C	10 min	

PROGRAMA PCR TPA25			
	Temperatura	Tiempo	Nº Ciclos
Desnaturalización	94°C	10 min	
Desnaturalización	94°C	1 min	x7
Anneling	65°C	1 min 30 s	
Extensión	72°C	1 min 30 s	
Desnaturalización	94°C	1 min	x25
Anneling	63°C	1 min	
Extensión	72°C	1 min	
Extensión Final	72°C	10 min	
PROGRAMA PCR FIIIB			
	Temperatura	Tiempo	Nº Ciclos
Desnaturalización	94°C	10 min	
Desnaturalización	94°C	1 min	x7
Anneling	60°C	1 min 30 s	
Extensión	72°C	1 min 30 s	
Desnaturalización	94°C	1 min	x25
Anneling	58°C	1 min	
Extensión	72°C	1 min	
Extensión Final	72°C	10 min	
PROGRAMA PCR D1			
	Temperatura	Tiempo	Nº Ciclos
Desnaturalización	94°C	10 min	
Desnaturalización	94°C	30 s	x7
Anneling	59,5°C	45 s	
Extensión	72°C	1 min 30 s	
Desnaturalización	94°C	1 min	x25
Anneling	56,5°C	1 min 30 s	
Extensión	72°C	2 min	
Extensión Final	72°C	10 min	

PROGRAMA PCR ACE			
	Temperatura	Tiempo	Nº Ciclos
Desnaturalización	94°C	10 min	
Desnaturalización	94°C	1 min	x30
Annealing	55,5°C	1 min 30 s	
Extensión	72°C	1 min 30 s	
Extensión Final	72°C	10 min	
PROGRAMA PCR APO			
	Temperatura	Tiempo	Nº Ciclos
Desnaturalización	94°C	10 min	
Desnaturalización	94°C	1 min	x30
Annealing	47°C	1 min 30 s	
Extensión	72°C	1 min 30 s	
Extensión Final	72°C	10 min	

La asignación de los fenotipos de cada muestra se ha realizado mediante una electroforesis en gel de agarosa. Aunque los geles de acrilamida presentan una mayor resolución, la agarosa ha mostrado ser eficaz en la discriminación de variaciones de tamaño superiores a 15-20pb. Debido a que los tamaños de los amplificadores de las inserciones *Alu* fluctúan entre 100 y 350 pb, dicha técnica es adecuada para este objetivo, siendo además un método rápido y de bajo coste. La concentración de agarosa por gel ha sido del 1% en peso. Con el fin de visualizar las bandas de ADN en el gel se ha utilizado el *Midori Green Advance DNA Stain (NIPPON Genetics EUROPE GmbH)*, un intercalante no carcinogénico y menos mutagénico que el habitualmente utilizado Bromuro de Etidio. En todos los casos se ha utilizado una concentración para la tinción del 5% en volumen.

Para la realización de la migración de las moléculas de ADN los geles se han sometido a un tiempo y diferencia de potencial adecuados a las propiedades de cada marcador genético utilizado, que comprendían entre potenciales de 100-110 eV y un tiempo de entre 20-24 minutos. Posteriormente, la visualización de los resultados se realizó mediante un transiluminador de luz UV *Syngene (SynopticsGroup) GelVueModel No. GVM20*. Finalmente, los resultados se recogieron mediante un sistema fotográfico digital (*Kodak Digital Science EDAS*

120). La identificación de la longitud de los amplificadores se ha realizado mediante un marcador *HyperLadder 100bp (meridian BIOSCIENCE)*, que permite la identificación entre un rango de longitudes comprendidos entre 100 y 1013 pb.

Marcadores SNPs informativos de ascendencia (AIMs)

Se tomaron alícuotas del ADN extraído de las 94 muestras de estudio y se enviaron al Servicio de Genómica de la Universidad del País Vasco (UPV/EHU). Allí se analizaron por medio de espectrofotometría (nanodrop) y por fluorimetría (Qubit), con el fin de determinar si presentaban una concentración y calidad óptimas. Posteriormente, se genotipó un grupo de 48 SNPs con características de AIMs, mediante un chip 96.96 JUNO de Biomark. Los SNPs seleccionados se describen en la Tabla 5.

Tabla 5: SNPs analizados y su localización cromosómica.

SNP	Cromosoma	SNP	Cromosoma
<i>rs11102524</i>	1	<i>rs9277413</i>	6
<i>rs4908602</i>	1	<i>rs9277418</i>	6
<i>rs6429223</i>	1	<i>rs9277498</i>	6
<i>rs1019264</i>	2	<i>rs9374640</i>	6
<i>rs182549</i>	2	<i>rs7783055</i>	7
<i>rs2920059</i>	2	<i>rs917118</i>	7
<i>rs896788</i>	2	<i>rs1392121</i>	8
<i>rs10470510</i>	3	<i>rs10123957</i>	9
<i>rs1498444</i>	3	<i>rs1493489</i>	12
<i>rs4455309</i>	3	<i>rs2710293</i>	12
<i>rs9870523</i>	3	<i>rs1886510</i>	13
<i>rs1157106</i>	4	<i>rs2065982</i>	13
<i>rs116818505</i>	6	<i>rs730570</i>	14
<i>rs200789833</i>	6	<i>rs1834640</i>	15
<i>rs2071350</i>	6	<i>rs8032154</i>	15
<i>rs2770</i>	6	<i>rs881929</i>	16
<i>rs369150</i>	6	<i>rs2033719</i>	17
<i>rs3823324</i>	6	<i>rs1024116</i>	18
<i>rs72500564</i>	6	<i>rs16951664</i>	18
<i>rs72873921</i>	6	<i>rs2285513</i>	19
<i>rs72873922</i>	6	<i>rs2303798</i>	19
<i>rs79244404</i>	6	<i>rs6057059</i>	20
<i>rs9273352</i>	6	<i>rs722098</i>	21
<i>rs9277332</i>	6	<i>rs2040411</i>	22

Otros marcadores genéticos

Este apartado se describe el resto de marcadores genéticos de otras familias de marcadores, reunidos a partir de diferentes bases de datos, utilizados con el fin de extrapolar sus resultados a la población romaní del País Vasco y compararlos con otras poblaciones romaníes europeas.

CODIS

Se obtuvieron los datos de los 13 STR autosómicos incluidos en el CODIS (*Combined DNA IndexSystem*) (Figura 39). Más concretamente los microsatélites D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, vWA, TPOX, D18S51, D5S818 y FGA, a partir de las bases de datos de Amigo et al. (2009), Budowle et al. (2001), Cabezas Silva et al. (2016), Egyed et al. (2006), Egyed et al. (2000), Füredi et al. (1996), Füredi et al. (1997), Haliti et al. (2009), Havaš et al. (2007); Gusmão et al. (2010), Kuzniar et al. (2004), Soták et al. (2008) y Novokmet et al. (2007). Los STRs se seleccionaron por su gran valor en los estudios de genética de poblaciones ya que presentan una selección neutral, por lo que su evolución está únicamente determinada por la mutación, la deriva genética y el flujo génico (Hearne et al., 1992). Tradicionalmente, dichos STRs habían sido ampliamente utilizados a partir de las bases de datos de ADN del *Federal Bureau of Investigation* (FBI) para la identificación de individuos y en las pruebas de paternidad (Butler, 2006).

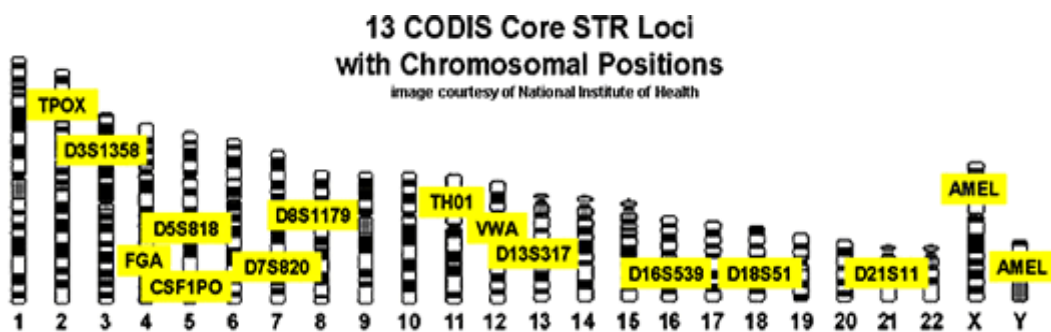


Figura 39: Posición de los 13 STRs incluidos en el CODIS junto con los loci de la amelogenina en los cromosomas X e Y. (Modificado de NIH base, [NationalInstituteofHealthhttps://www.nist.gov/image/13codiscorestrlociwithchromosomalpositions.png](https://www.nist.gov/image/13codiscorestrlociwithchromosomalpositions.png)).

Cromosoma Y

Por otra parte, se tomaron las frecuencias de haplogrupos del cromosoma Y a partir de la base de datos de Martínez-Cruz et al. (2016), en la que se habían genotipado 131 SNPs del cromosoma Y de diversas muestras de poblaciones europeas, e identificado 19 haplogrupos del cromosoma Y. Por un lado, los autores del estudio genotiparon 6 SNPs en un solo multiplex

(Multiplex-2) y 4 SNPs con ensayos TaqMan individuales (L48, M458, L2 y L20). Además, todos los individuos de la muestra se tipificaron para un conjunto de 19 STR, 17 de ellos usando el kit Yfiler (AppliedBiosystems, Foster City, CA, EE.UU.) y 2 en Multiplex-2. También se tomaron las frecuencias de la base de datos de Semino et al. (2000). Finalmente, los individuos fueron agrupados en haplogrupos del cromosoma Y, a saber A; B; C; D; E; F; G; H; I; J; K; L; M; N; O; P; Q; R y T.

ADNmt

En lo que compete a las frecuencias de haplogrupos del ADNmt éstas se extrajeron de diversas bases de datos. En la de Martínez-Cruz et al. (2016), con un total de 1341 individuos de los cuales 736 pertenecían a distintas poblaciones romaníes y 605 a poblaciones huésped, se secuenciaron ambos segmentos hipervariables (HVS I y II) de la región de control. Asimismo, se genotiparon 22 SNPs de la región codificante mediante el multiplex GenoCoRe22. Para 12 individuos (8 españoles, 1 romaní eslovaco y 3 rumanos no romaníes), los autores sólo pudieron secuenciar el haplotipo HVS-1. Las secuencias completas se recogen en la base de datos EMPOP (número de acceso EMP00671). Posteriormente, en base a los resultados obtenidos para HVS y la codificación de los SNPs, los individuos fueron agrupados en haplogrupos de ADNmt. Adicionalmente, la base de datos de Metspalu et al. (2004) contenía la secuencia de la región HVS-I de 796 individuos indios y 436 iraníes. La secuenciación fue realizada por medio del kit de secuenciación del ciclo terminador DYEnamic™ ET (AmershamPharmaciaBiotech) y en secuenciadores de ADN ABI 377 o Megabace2000.

Adicionalmente, los autores analizaron posiciones informativas dentro de la región codificante del ADNmt en individuos con diferentes haplotipos HVS-1 con el fin de determinar las afiliaciones de los haplogrupos que lo componen. Finalmente, Ottoni et al. (2011) secuenciaron ambas regiones hipervariables (HVS-I y HVS-II) en 85 individuos procedentes de la península de Anatolia y a su vez recapitularon una base de datos compuesta por 157 poblaciones de Eurasia y África. En total los haplogrupos analizados en el ADNmt fueron: C; D; F1B; H*; HV; I; J; K; L2,6; L2a; M*; N*; N1; R*; R0a; R10; T; U; U1; U2; U3; U4; U5; U6; U7; V; W y X.

IV.3. Análisis estadísticos

A continuación se describen los análisis realizados para tratar los datos obtenidos y poder interpretarlos, que incluyen el uso de distancias genéticas, diversos análisis estadísticos multivariados, modelos de representación gráfica, el estadístico $-\log(P)$ para estimar el efecto de la deriva genética, y los coeficientes de correlación de Pearson y Spearman.

Distancia genética de Reynolds

Es una medida que asume que los cambios en las frecuencias alélicas son originados únicamente por efecto de la deriva genética y el flujo génico (Reynolds et al., 1983). La fórmula matemática que permite el cálculo de la distancia es:

$$\Theta_w = \sqrt{\frac{\sum_{\ell} \sum_u (X_u - Y_u)^2}{2 \sum_{\ell} \left(1 - \sum_u X_u Y_u\right)}}$$

Donde ℓ es el conjunto de los *loci*, u el conjunto de los alelos, X_u es la frecuencia del u -ésimo alelo para el ℓ -ésimo locus en la población 1 y Y_u es la frecuencia del u -ésimo alelo para el ℓ -ésimo locus en la población 2.

Distancia genética de Harpending y Jenkins

Es una medida que se basa en que el parentesco entre poblaciones, que se puede expresar como *la probabilidad de que dos alelos sean idénticos por descendencia cuando se toman al azar de dos individuos*, uno de una población i y otro de una población j (Harpending y Jenkins, 1973). Para ello se utiliza la siguiente expresión de la matriz R de parentesco:

$$r_{ij} = \sum_k \frac{(P_{ik} - \bar{P}_k)(P_{jk} - \bar{P}_k)}{\bar{P}_k(1 - \bar{P}_k)}$$

Donde P_{ik} y P_{jk} son las frecuencias en las poblaciones i y j de un alelo k y \bar{P}_k es su promedio.

Análisis de escalamiento multidimensional (MDS)

Es una técnica de análisis multivariante que, a partir de una matriz de disimilitudes, realiza una representación en una escala euclídea ordinaria de una serie de poblaciones a analizar en un espacio N dimensional (Hout et al., 2013). En nuestro marco de estudio, las posiciones relativas de las poblaciones representarán el grado de proximidad genética entre ellas. Para determinar el grado de fiabilidad de la representación obtenida, ésta se evalúa mediante el denominado coeficiente de estrés. Dicho coeficiente se obtiene como resultado de:

$$stress = \sqrt{\frac{\sum_i \sum_j (d_{ij} - \hat{d}_{ij})^2}{\sum_i \sum_j (d_{ij}^2)}}$$

Donde d_{ij} es la distancia euclídea entre las poblaciones i y j y d_{ij}^g su distancia genética.

Los análisis MDS se han realizado mediante los programas Past v4.06b (Hammer et al., 2001) y el GeDisv1.8 (Peña et al., 2009) diseñado en nuestro laboratorio.

Análisis factorial de correspondencias (AFC)

Se trata de un método estadístico que consiste en buscar la mejor representación simultánea para dos conjuntos de datos (casos y variables) que constituyen las filas y las columnas de una tabla de contingencia. El AFC se utiliza con el fin de estudiar la relación entre los dos conjuntos. Para ello se ha utilizado el programa Past v4.06b (Hammer et al., 2001) mediante el cual se realizó el análisis y representación del estadístico.

Método del centroide

Es un modelo que permite valorar el grado de flujo génico de una población respecto al resto de poblaciones que conforman el análisis (Harpending y Ward, 1982). Se basa en la contraposición de la heterocigosidad observada de las poblaciones frente a la varianza que estas aportan al conjunto (r_i).

$$r_i = \frac{(p_i - P)^2}{P(1 - P)}$$

Donde r_i es la distancia al centroide, que equivale a la varianza de Wahlund que aporta cada población al conjunto analizado. Por otro lado, p_i sería la frecuencia de cada alelo en la población i y P la media de la frecuencia de ese alelo en el conjunto.

Este análisis y su representación se ha realizado mediante el programa el GeDisv1.8 (Peña et al., 2009) creado en nuestro laboratorio.

AMOVA

El análisis molecular de la varianza (AMOVA) es un método estadístico derivado del ANOVA (análisis de la varianza) aplicado a datos moleculares como las frecuencias alélicas o haplotípicas. Fue desarrollado por Excoffier et al. (1992), y obtiene estimaciones a partir de los valores de la varianza y los estadísticos F . Estos últimos reflejan la correlación de la diversidad molecular en diferentes niveles de subdivisiones jerárquicas. El objetivo primordial de este análisis es la gradación de la estructura que presentan los grupos poblacionales designados. Para ello se utilizan tres índices de fijación: F_{ST} =varianza genética entre individuos dentro de una población, F_{SC} =varianza genética entre poblaciones dentro de un grupo y F_{CT} =varianza genética

entre grupos. La preparación de los archivos para su análisis se realizó mediante el programa ya mencionado GeDisv1.8 (Peña et al., 2009) y el posterior análisis se realizó mediante el programa Arlequin v3.35 (Excoffier y Lischer, 2010).

Dendrograma

Se trata de un modelo de representación gráfica, en forma de árbol, que organiza los datos en función de una categoría que se va subdividiendo en subcategorías determinadas al pasar de un nivel al siguiente. En el caso de los análisis de genética de poblaciones permite representar las distancias genéticas de una serie de poblaciones pudiéndose deducir, a partir de él, las poblaciones de origen y las descendientes (Luizon et al., 2008; Saitou y Nei, 1987). En este caso se utilizó el método Neighbor-joining (Saitou y Nei, 1987), que produce un árbol final único bajo el principio de mínima evolución y es aplicable a cualquier tipo de distancia evolutiva. Para ello se han seguido una serie de pasos:

En primer lugar, se calculó la matriz Q a partir de la matriz de distancias:

$$Q(i, j) = (n - 2)d(i, j) - \sum_{k=1}^n d(i, k) - \sum_{k=1}^n d(j, k)$$

Donde d(i,j) es la distancia entre los taxones i y j y n el número de taxones analizados.

Posteriormente, se unen los taxones cuyo valor de Q(i,j) fuera el menor y se calculó la distancia de esta pareja respecto al nodo mediante las ecuaciones:

$$d(f, u) = \frac{1}{2}d(f, g) + \frac{1}{2(r - 2)} \left[\sum_{k=1}^r d(f, k) - \sum_{k=1}^r d(g, k) \right]$$

y

$$d(g, u) = d(f, g) - d(f, u)$$

Donde f y g son los taxones emparejados por presentar un valor Q mínimo y u el nuevo nodo.

A continuación, se determinó la distancia de los otros taxones con el nuevo nodo por medio de:

$$d(u, k) = \frac{1}{2}[d(f, k) + d(g, k) - d(f, g)]$$

Donde, k es el nodo del que se quiere calcular su distancia.

Finalmente, se repitieron sucesivamente los pasos anteriores considerando cada par emparejado como uno solo hasta que se determinó la distancia del último par de taxones. La matriz de distancias F_{ST} y el árbol se construyeron con el programa Populations (Langella, 1999) y el diagrama se dibujó con el programa Forester (Zmasek, 2018).

Análisis de clinas

Se utilizó el programa GenoCline (Peña et al., 2022) para la detección de clinas a partir de las bases de datos de frecuencias alélicas. GenoCline proyecta la posición geográfica de cada población (establecida a partir de sus coordenadas geográficas) sobre un eje de coordenadas que rota 360 grados, en iteraciones de 1 grado. En cada iteración, se calcula el coeficiente de correlación entre el vector de coordenadas y el vector de frecuencias alélicas. Finalmente, si existe correlación significativa para una o varias orientaciones, se selecciona aquella que muestra una significación mayor. Puesto que se han realizado un cierto número de pruebas, una parte de las clinas significativas podrían deberse al azar, por lo que se aplicó la corrección de Bonferroni. Así, en lugar de valorar una significación de $P < 0,01$, se consideraron sólo aquellas con una $P < (0,01/n)$, donde n es el número de pruebas realizadas. Como resultado se obtuvo una hoja de cálculo que contenía los nombres de los marcadores, la orientación de las clinas y su significación.

Posible acción de la deriva: $-\log(P)$

Generalmente se menciona el tamaño efectivo de una población (N_e) como una evidencia de la acción potencial de la deriva. En efecto, cuanto menor sea N_e , mayor será la acción de la deriva. Por ello, tradicionalmente se han desarrollado métodos de estimación de dicha acción potencial, cuyo base es una estimación de N_e (Jorde y Ryman, 2007; Nei y Tajima, 1981).

Sin embargo, uno de los objetivos del presente trabajo es identificar aquellos alelos afectados por la acción de la deriva para evitar su uso en el cálculo de las tasas de mestizaje y mejorar de ese modo la precisión y fiabilidad de los cálculos. Por ello, se ha desarrollado una metodología diferente, partiendo del estadístico $-\log(P)$ (Peña et al., 1997) el cual considera que dada una población mestiza, fruto del flujo génico entre dos poblaciones parentales, podremos

calcular una “varianza estandarizada esperada” de las frecuencias alélicas para estas tres poblaciones (la mestiza y las dos parentales) en ausencia de deriva, bajo dos supuestos. En el primero, con ausencia también de flujo génico, las frecuencias de la población mestiza serán idénticas a las de la población parental original. En el segundo, con flujo génico, las frecuencias de la población mestiza tendrán un valor intermedio entre las frecuencias de ambas poblaciones. Para simplificar, en nuestro estudio se ha considerado una tasa de mestizaje equivalente al 50%, y se puede calcular el cociente entre la varianza observada y la esperada bajo ambos supuestos:

$$\frac{F_{ST\ obs}}{F_{ST\ esp}}$$

Este cociente se distribuye según una función

$$F_{(r-1)(k-1),\infty}$$

Donde r es el número de subpoblaciones y k el número de alelos.

Con el fin de que la probabilidad resultante pueda ser más fácilmente interpretable, se hace que siga una función lineal mediante la transformación

$$-\log(P)$$

Así, un valor de $-\log(P)$ mayor que 1,3 implicará significación al nivel $p < 0,05$.

Coefficientes de correlación de Pearson y Spearman

Son medidas de dependencia lineal entre dos variables aleatorias cuantitativas que permite conocer el grado de relación entre ellas. A diferencia de la covarianza, los coeficientes son independientes de la escala de medida de las variables. Ambos coeficientes de correlación tienen una escala que va desde -1 a +1, donde 0 indica que no hay una asociación lineal o monótona, y cuanto mayor sea la correlación entre ambas variables más se aproxima a una línea recta (correlación de Pearson) o a una curva que aumenta o disminuye constantemente (correlación de Spearman) a medida que el coeficiente se aproxima a un valor absoluto de 1 (Schober et al., 2018).

El coeficiente de correlación de Pearson cuando se aplica a una muestra:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Donde n es el tamaño de la muestra, x_i e y_i son el par de datos a comparar y \bar{x} es la media muestral definida por:

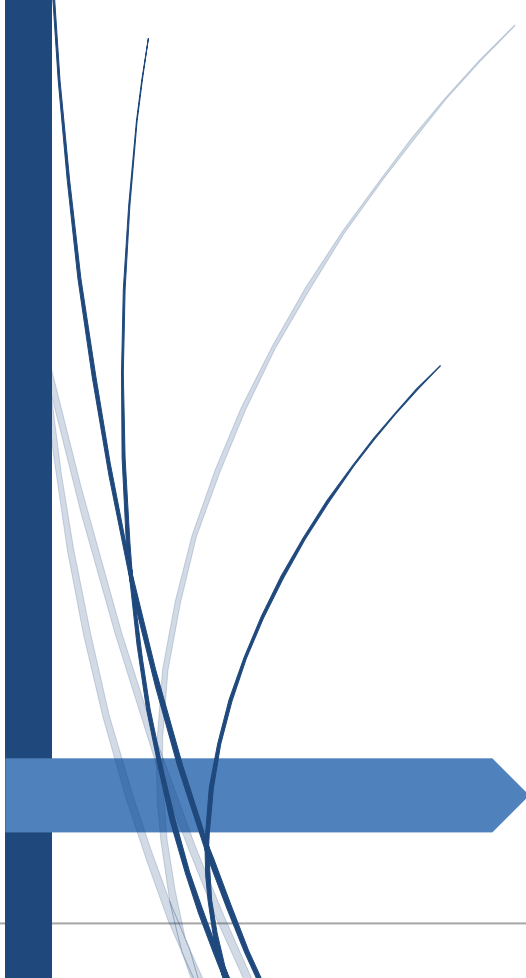
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

El coeficiente de correlación de Spearman se calcula como:

$$r_R = 1 - \frac{6\sum_i d_i^2}{n(n^2 - 1)}$$

Donde n es el número de puntos de datos de las dos variables, d_i diferencia de rango del elemento n

Ambas correlaciones se han calculado mediante del programa Past 4.0 (Hammer et al., 2001).



Resultados

V. Resultados

V.1. Análisis de la población romaní del País Vasco para AIMs

V.1.1. Frecuencias de 48 SNPs AIMs autosómicos

En la Tabla 6 se muestran las frecuencias de los 48 SNPs analizados en la población romaní objeto de estudio. Los valores de las frecuencias son muy variados (Figura 40), oscilando entre 0,000 (rs116818505) y 0,973 (rs2071350). En general, casi todos ellos muestran un carácter polimórfico, en consonancia con el criterio de ser marcadores informativos de ancestralidad, tal y como se consideró para seleccionarlos.

Tabla 6: Tamaño de muestra, alelo principal, frecuencia del alelo principal, error estándar, test Chi-2 de equilibrio Hardy-Weinberg y corrección de Bonferroni para comparaciones múltiples de los AIMs analizados.

SNP	Alelo principal	N	Frecuencias	Error estándar	Chi-2	P	P (Bonferroni)
rs11102524	C	78	0,519	0,040	0,194	0,660	n.s.
rs4908602	A	76	0,276	0,036	1,069	0,301	n.s.
rs6429223	A	72	0,576	0,041	0,859	0,354	n.s.
rs1019264	C	74	0,405	0,040	2,325	0,127	n.s.
rs182549	C	77	0,844	0,029	3,404	0,065	n.s.
rs2920059	A	71	0,451	0,042	1,349	0,246	n.s.
rs896788	A	73	0,342	0,039	0,659	0,417	n.s.
rs10470510	A	78	0,532	0,040	3,178	0,075	n.s.
rs1498444	A	76	0,625	0,039	0,023	0,878	n.s.
rs4455309	C	78	0,667	0,038	0,029	0,865	n.s.
rs9870523	A	73	0,630	0,040	1,023	0,312	n.s.
rs1157106	C	73	0,370	0,040	0,245	0,620	n.s.
rs116818505	G	76	0,000	-	-	-	-
rs200789833	G	68	0,463	0,043	43,942	0,000	p<0,01
rs2071350	C	75	0,973	0,013	0,056	0,812	n.s.
rs2770	C	72	0,222	0,035	1,125	0,289	n.s.
rs369150	A	61	0,295	0,041	1,080	0,299	n.s.
rs3823324	T	76	0,671	0,038	0,852	0,356	n.s.
rs72500564	C	71	0,838	0,031	0,569	0,451	n.s.
rs72873921	C	77	0,104	0,025	1,035	0,309	n.s.
rs72873922	A	77	0,097	0,024	0,897	0,344	n.s.
rs79244404	C	77	0,149	0,029	0,414	0,520	n.s.
rs9273352	G	76	0,638	0,039	0,272	0,602	n.s.
rs9277332	A	70	0,629	0,041	11,615	0,001	n.s.
rs9277413	A	78	0,173	0,030	0,276	0,600	n.s.
rs9277418	C	74	0,818	0,032	0,175	0,675	n.s.
rs9277498	C	78	0,160	0,029	0,712	0,399	n.s.
rs9374640	A	74	0,331	0,039	0,218	0,641	n.s.
rs7783055	C	74	0,669	0,039	2,667	0,102	n.s.
rs917118	C	78	0,679	0,037	0,263	0,608	n.s.
rs1392121	C	74	0,453	0,041	0,006	0,938	n.s.
rs10123957	C	73	0,664	0,039	0,014	0,907	n.s.
rs1493489	C	76	0,539	0,040	0,267	0,606	n.s.
rs2710293	A	71	0,634	0,040	0,605	0,437	n.s.
rs1886510	C	76	0,697	0,037	0,000	0,983	n.s.
rs2065982	C	78	0,083	0,022	0,645	0,422	n.s.
rs730570	A	77	0,870	0,027	0,091	0,763	n.s.
rs1834640	A	77	0,805	0,032	0,449	0,503	n.s.
rs8032154	A	76	0,520	0,041	0,047	0,829	n.s.
rs881929	G	75	0,480	0,041	0,017	0,897	n.s.

rs2033719	C	77	0,325	0,038	0,958	0,328	n.s.
rs1024116	A	72	0,493	0,042	0,056	0,812	n.s.
rs16951664	G	73	0,589	0,041	0,025	0,874	n.s.
rs2285513	A	77	0,383	0,039	0,394	0,530	n.s.
rs2303798	A	78	0,032	0,014	0,086	0,770	n.s.
rs6057059	A	75	0,453	0,041	6,776	0,009	n.s.
rs722098	A	74	0,831	0,031	3,057	0,080	n.s.
rs2040411	A	73	0,712	0,037	5,113	0,024	n.s.

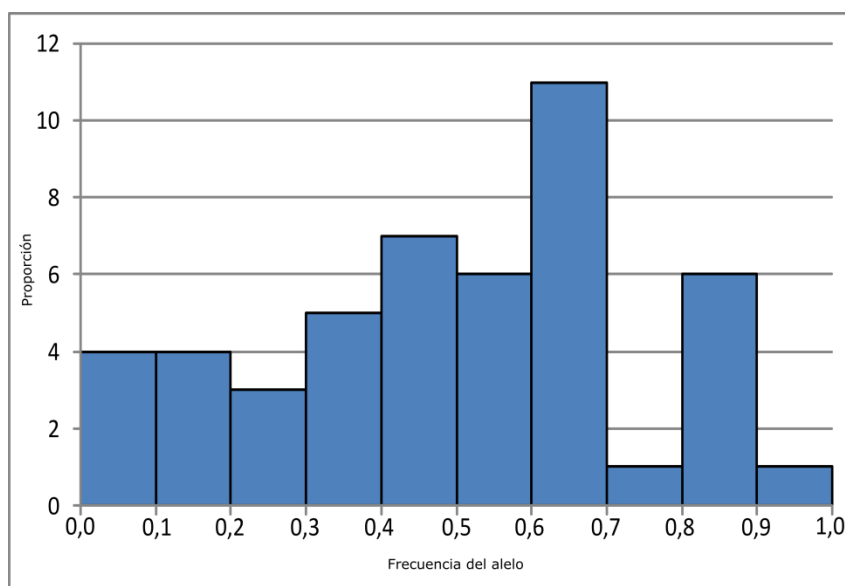


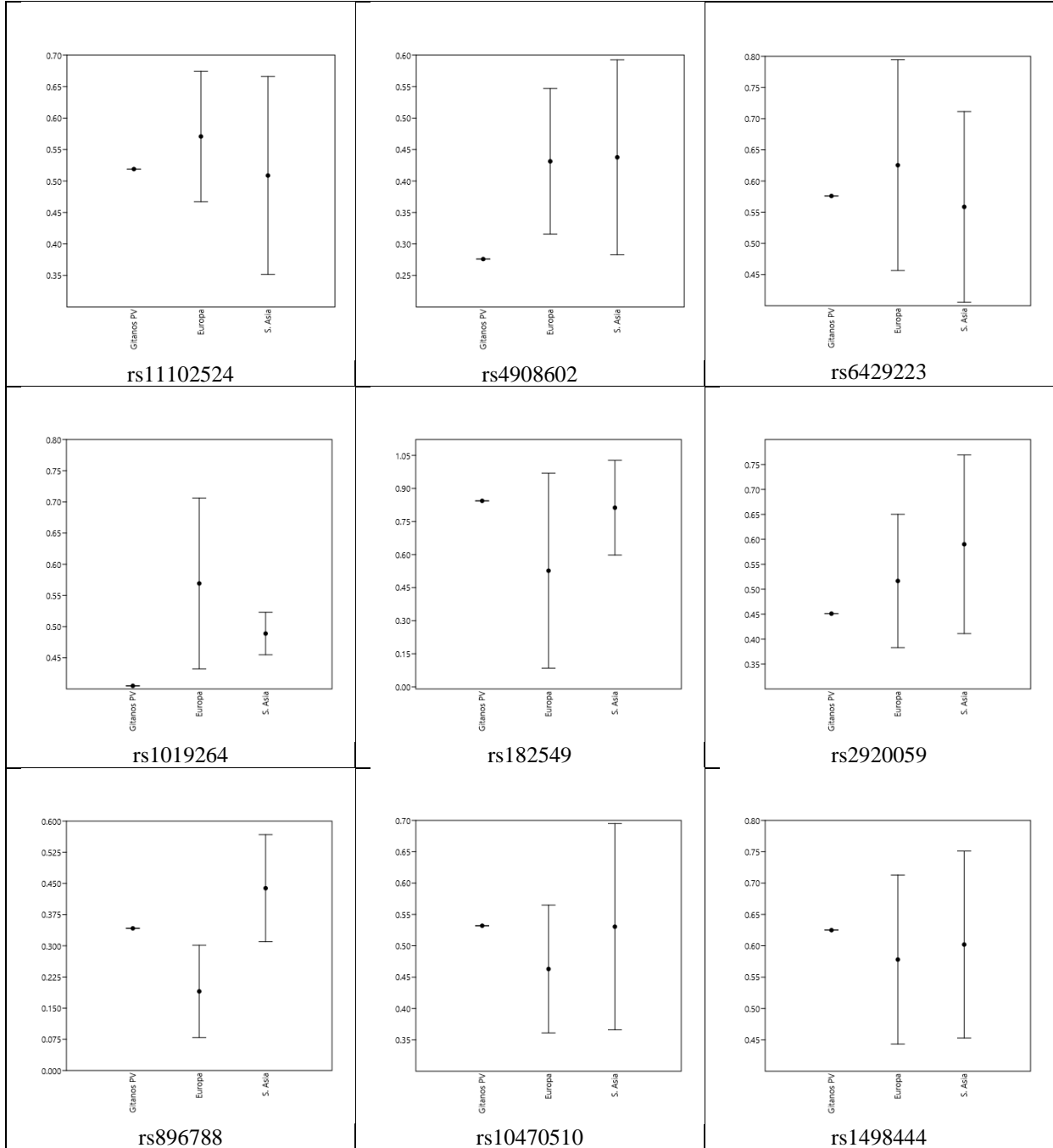
Figura 40: Distribución de las frecuencias absolutas de los AIMs analizados en la población romaní del País Vasco.

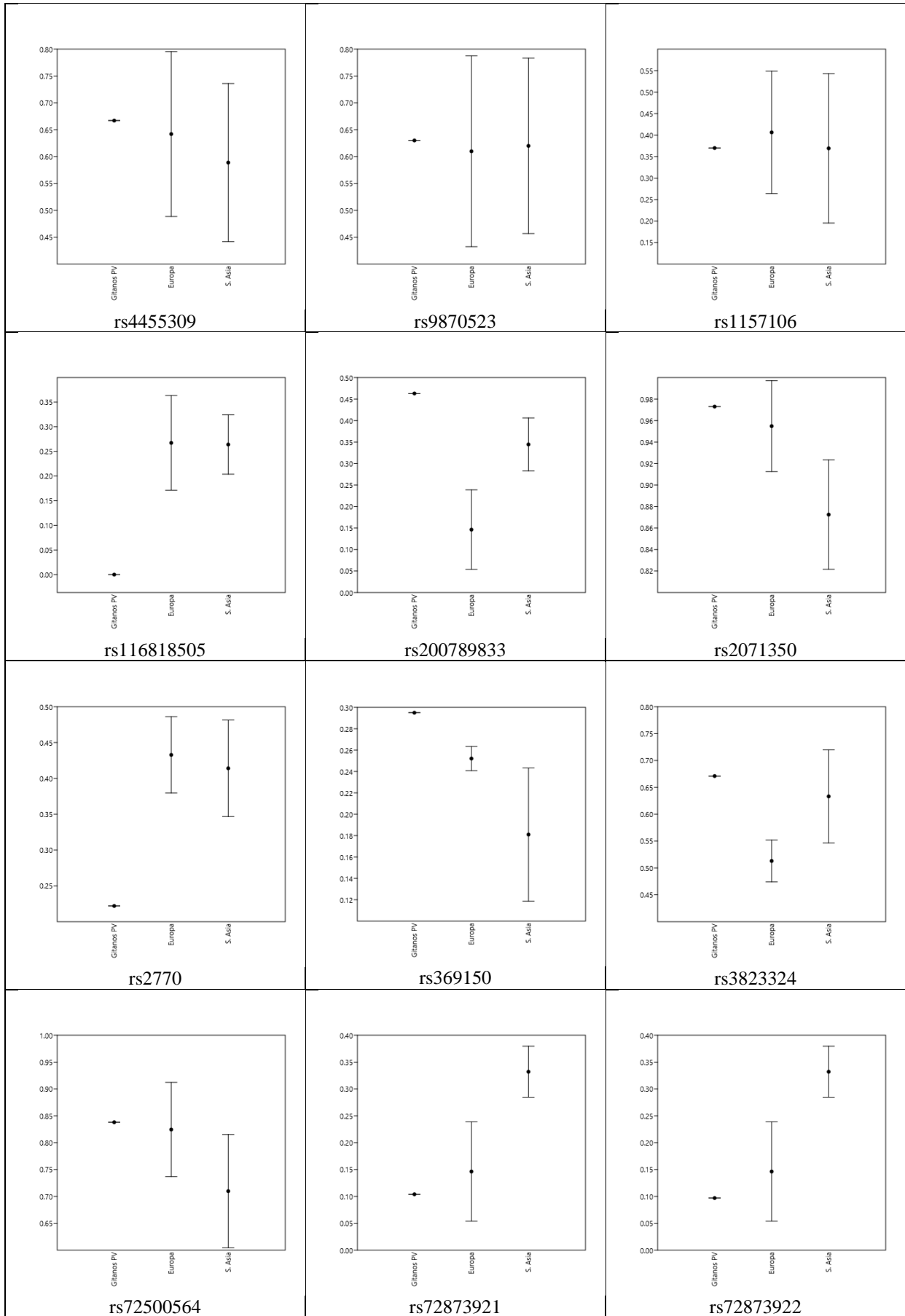
Cabe destacar que tres SNPs han mostrado valores de P significativos para el test de equilibrio Hardy-Weinberg (rs200789833, rs9277332 y rs6057059), si bien, tras aplicar la corrección de Bonferroni sólo ha mostrado significación el SNP rs200789833 ($P < 0,01$), que además muestra un notable exceso de heterocigotos (Tabla 6). La causa pudiera estar en su ubicación en el Complejo mayor de Histocompatibilidad, próximo al extremo 5' del gen HLA-DPA1 y al extremo 3' del gen HLA-DPB1.

Se han recopilado las frecuencias de estos marcadores en diferentes poblaciones, con el objetivo de analizar el patrimonio genético romaní en su contexto histórico-geográfico (Tabla 2 del Anexo). Teniendo en cuenta la historia de este grupo étnico, se han seleccionado aquellas poblaciones de Europa y del centro y Sur de Asia que tuvieran datos para los mismos SNPs analizados en este trabajo, a partir de la base de datos ALFRED (Cheung et al., 1999). En la Figura 41 se muestran, para dichos SNPs, las frecuencias en la población romaní del País Vasco, así como la media y los errores estándar en Europa y Sur de Asia. En 13 de ellos, el valor en la muestra romaní se sitúa entre las medias europeas y asiática, en 20 se encuentra por encima de ambos valores y en 15 por debajo. Además, en 11 SNPs las frecuencias obtenidas en los romaníes

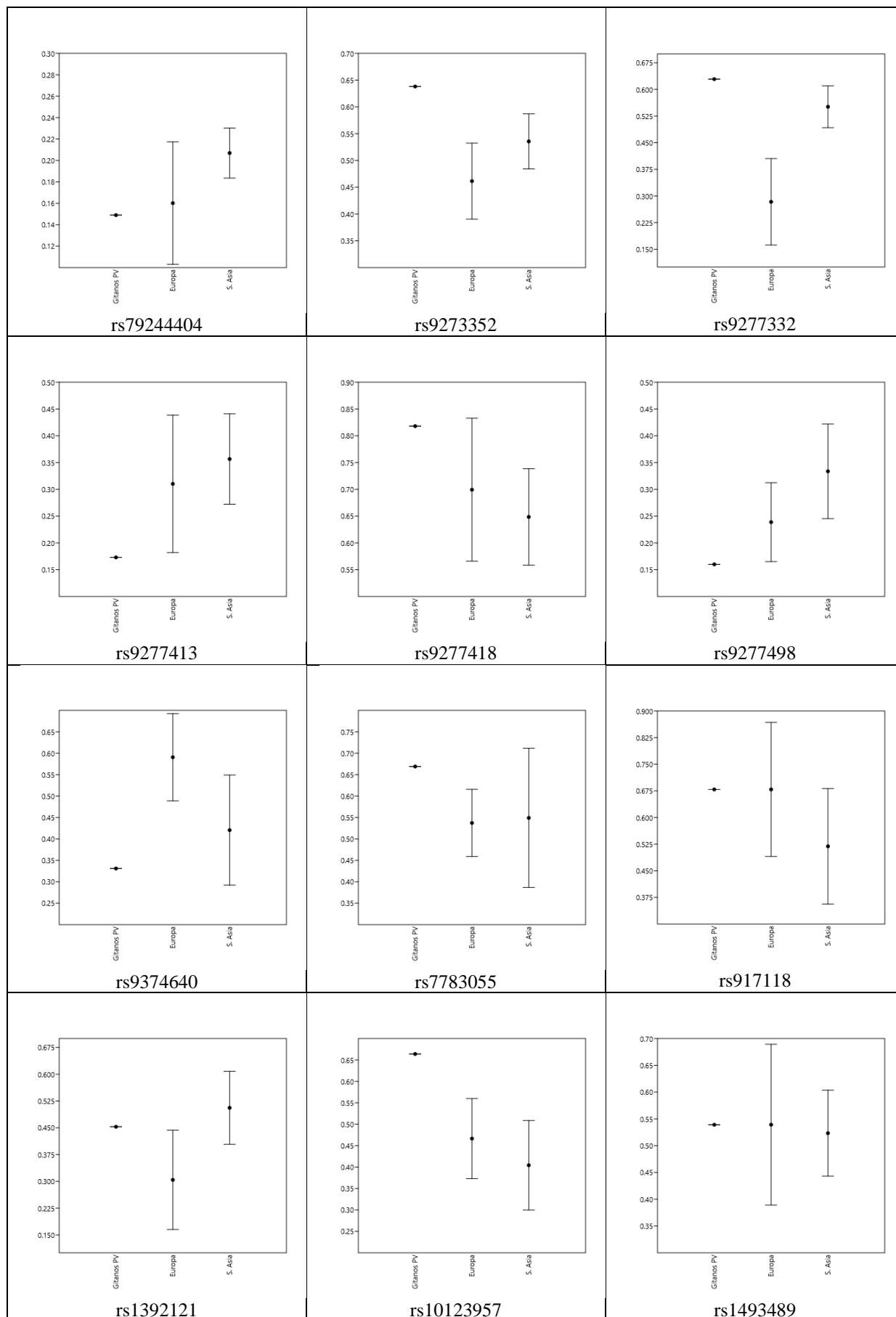
Resultados

se encuentran bien por encima o por debajo del rango de los errores estándar conjuntos de Europa y el Sur de Asia y, por tanto, muy alejados de los valores medios de ambas regiones.

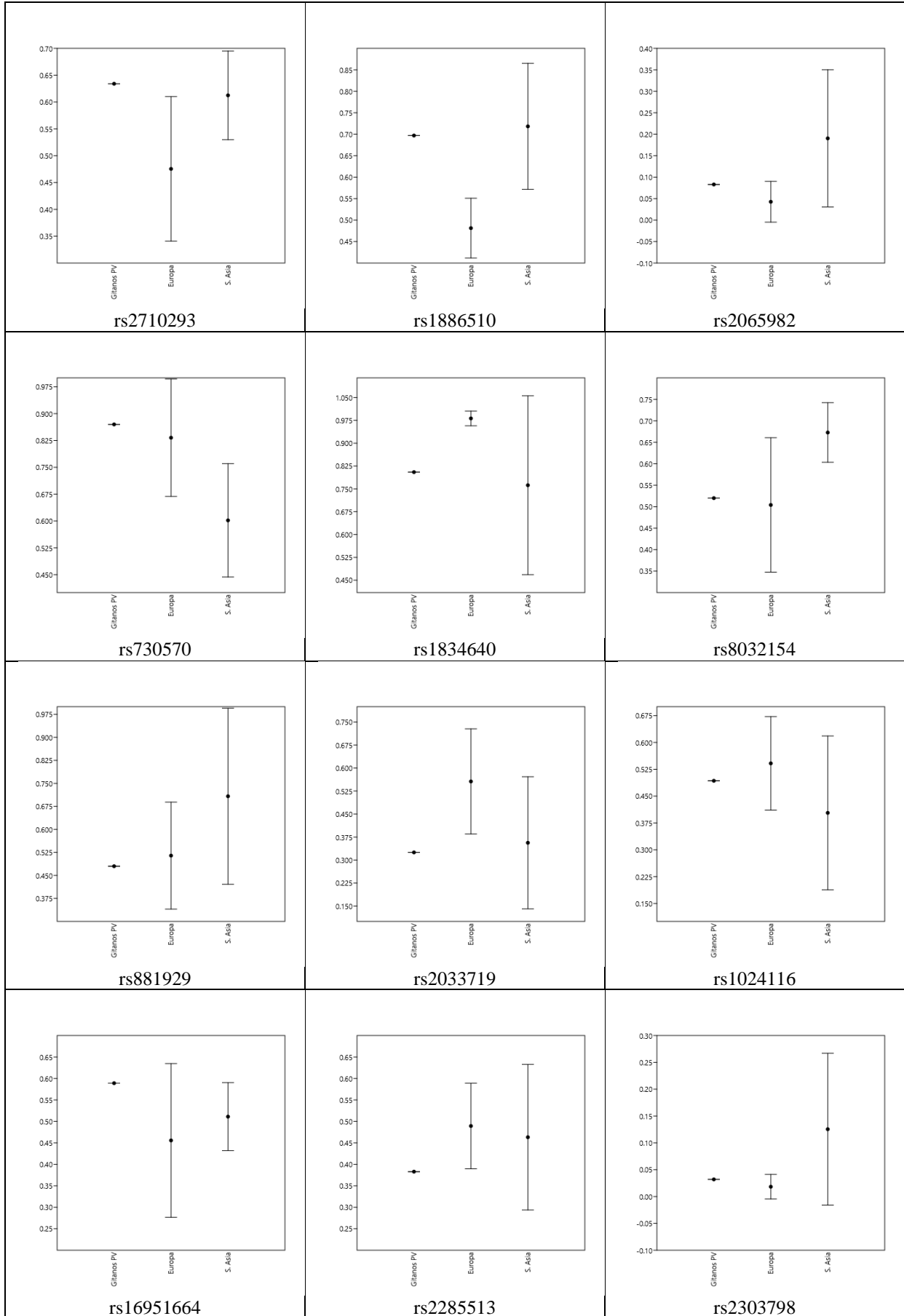




Resultados



Procesos microevolutivos en la población romaní residente en el País Vasco



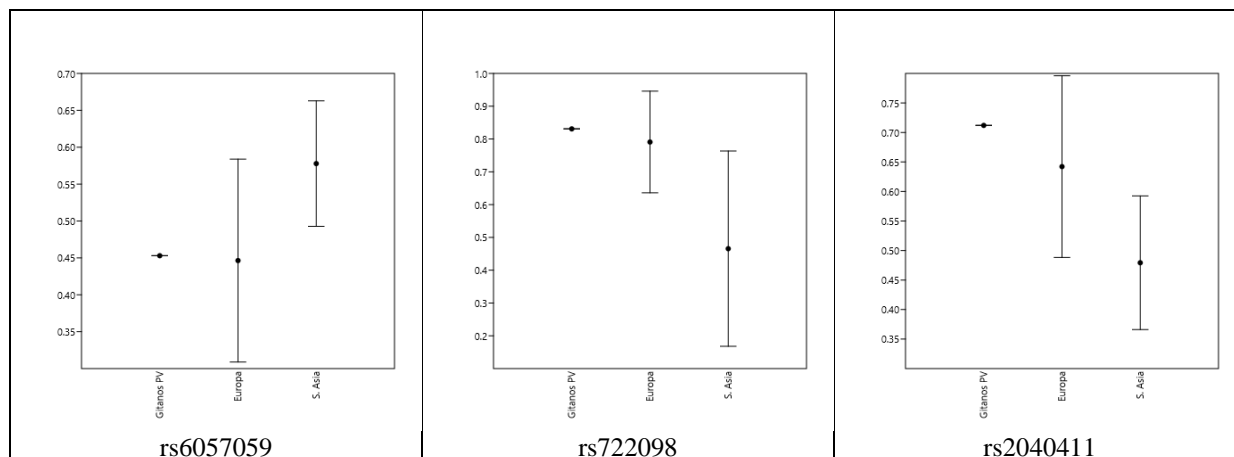


Figura 41: Frecuencias de los AIMs analizados en la población romaní del País Vasco, referidas a los valores medio y errores estándar de Europa y del Sur de Asia.

Análisis molecular de la varianza (AMOVA)

Con el fin de determinar la existencia de diferencias significativas entre las poblaciones de Europa y el Sur de Asia y determinar los marcadores más informativos para los posteriores análisis, se llevó a cabo un Análisis Molecular de la Varianza (AMOVA) para los 48 marcadores AIMs en ambos grupos continentales (Tabla 7). Junto a los valores de los índices de fijación se adjunta el número de poblaciones analizadas para cada marcador. Se ejecutaron 1000 permutaciones para este análisis, mediante el programa Arlequín. Las poblaciones utilizadas fueron tomadas de la base de datos de ALFRED (Cheung et al., 1999) y se muestran en la Tabla 2 del Anexo.

Teniendo en cuenta el índice F_{ST} se observan diferencias significativas en 46 de los 48 SNPs analizados. De entre ellos, dos presentan una significación de $P < 0,05$, uno un nivel de significación mayor ($P < 0,01$) y el resto de marcadores una significación aún más elevada ($P < 0,001$). Esto indica que hay una gran heterogeneidad genética entre los individuos dentro de cada población. En lo que respecta a la heterogeneidad entre las poblaciones dentro de los grupos seleccionados (F_{SC}), 41 de los 48 marcadores muestran diferencias con algún nivel de significación: tres un valor de $P < 0,05$, dos uno de $P < 0,01$ y el resto ha alcanzado una significación aún mayor ($P < 0,001$). Finalmente, respecto a la varianza entre grupos (F_{CT}), 30 marcadores presentan algún nivel de significación: 13 SNPs con una $P < 0,05$, tres con valores $P < 0,01$, y 14 con significación aún más elevada. En todos los casos, los valores negativos se han considerado equivalentes a 0. En definitiva, a pesar de haber seleccionado SNPs con características de AIMs, en una parte de ellos no se observa heterogeneidad significativa entre los dos grupos poblacionales considerados.

Tabla 7: Índices de fijación (F_{ST} , F_{SC} y F_{CT}) generados mediante el análisis AMOVA para las frecuencias de 48 SNPs pertenecientes a los AIMs en poblaciones de Europa y el Sur de Asia. (F_{CT} : varianza genética entre grupos. F_{SC} : varianza genética entre poblaciones dentro de grupos. F_{ST} : varianza genética entre individuos dentro de una población) (* $P < 0,05$; ** $P < 0,01$; *** $P < 0,001$).

Índices de Fijación				Nº Poblaciones
F_{ST}	F_{SC}	F_{CT}		
rs11102524	0,02591***	0,02153***	0,00447*	31
rs4908602	0,11929***	0,05297***	0,07003*	31
rs6429223	0,05530***	0,03019***	0,02589***	31
rs2920059	0,03494***	0,02579***	0,00939*	31
rs10470510	0,01356***	0,01091***	0,00268*	31
rs1498444	0,04695***	0,02563***	0,02188***	31
rs1157106	0,02918***	0,02663***	0,00263*	31
rs7783055	0,04029***	,01397***	0,02669**	31
rs917118	0,05974***	0,04207***	0,01845*	31
rs1392121	0,04739***	0,02068***	0,02728***	31
rs10123957	0,03287***	0,0093	0,02379*	31
rs2710293	0,03272***	0,02643***	0,00646*	31
rs2033719	0,03946***	0,02217***	0,01768***	31
rs6057059	0,02539***	0,00378	0,02170***	31
rs182549	0,21591***	0,16382***	0,06229	19
rs896788	0,14916***	0,02795***	0,12469***	19
rs9374640	0,06369***	0,02360***	0,04106**	19
rs1886510	0,13550***	0,03210***	0,10683***	19
rs2065982	0,21808***	0,10651***	0,12487***	19
rs730570	0,21103***	0,07666***	0,14553***	19
rs1834640	0,34381***	0,25331***	0,12120*	19
rs881929	0,24737***	0,19515***	0,06488	19
rs1024116	0,09232***	0,04166***	0,05286**	19
rs2285513	0,01538***	0,01266**	0,00275	19
rs2303798	0,10194***	0,03175***	0,07249***	19
rs722098	0,19523***	0,09296***	0,11275***	19
rs2040411	0,10196***	0,02382***	0,08005***	19
rs116818505	0,04001***	0,04545***	-0,00571	10
rs200789833	0,07054***	0,04552***	0,02622	10
rs2071350	0,04479***	0,00478	0,04021*	10
rs2770	0,01163***	0,01626***	-0,00471	10
rs369150	0,02053*	-0,0028	0,02327*	10
rs3823324	0,02458***	0,01123*	0,0135	10
rs72500564	0,03274***	0,01880***	0,0142	10
rs72873921	0,09382***	0,00723	0,08722***	10
rs72873922	0,09382***	0,00723	0,08722***	10

rs79244404	0,0001	-0,00336	0,00345	10
rs9273352	0,01323*	0,01202*	0,00123	10
rs9277332	0,01911***	0,02007***	-0,00097	10
rs9277413	0,01218***	0,02663***	-0,0012	10
rs9277418	0,02780***	0,02553***	0,00233*	10
rs9277498	0,02146***	0,02172***	-0,00026	10
rs1019264	0,02464***	0,02720***	-0,00263	15
rs4455309	0,03877***	0,01943***	0,01972*	15
rs9870523	0,03725	0,04382*	-0,00687	15
rs1493489	0,00596**	0,00750**	-0,00156	15
rs8032154	0,01668***	0,01727***	0,0006	15
rs16951664	0,02265***	0,02821***	-0,00572	15

Análisis de escalamiento multidimensional (MDS)

Con el fin obtener una primera aproximación a la posición genética de la población romaní del País Vasco respecto a las poblaciones europeas y del Sur de Asia, se realizó un Análisis de Escalamiento Multidimensional (MDS). Para ello se incluyeron aquellas poblaciones y SNPs mencionados en la Tabla 2 del Anexo y se utilizó la distancia F_{ST} de Reynolds et al. (1983). En la Figura 42 se observa la posición intermedia de la población romaní analizada entre el grupo formado por las poblaciones europeas, a la izquierda de la figura, y el formado por las poblaciones asiáticas a la derecha de la misma, siendo la población Balochi la más próxima a la romaní del País Vasco. Las posiciones de las poblaciones europeas están relativamente próximas entre sí en contraposición a las poblaciones asiáticas, que se encuentran bastante dispersas. Con el fin de comparar la variabilidad de ambos grupos poblacionales e identificar la posición de los romaníes respecto a dicha variabilidad poblacional, se ha realizado un análisis MDS mediante el programa Past, añadiendo unas elipses equiprobables del 95 % (Figura 43). Como era de esperar, el resultado muestra una elipse mucho mayor para el Sur de Asia que para Europa, con los romaníes del País Vasco en una posición intermedia, en el área de solapamiento de ambas elipses, ligeramente incluida en los grupos europeos.

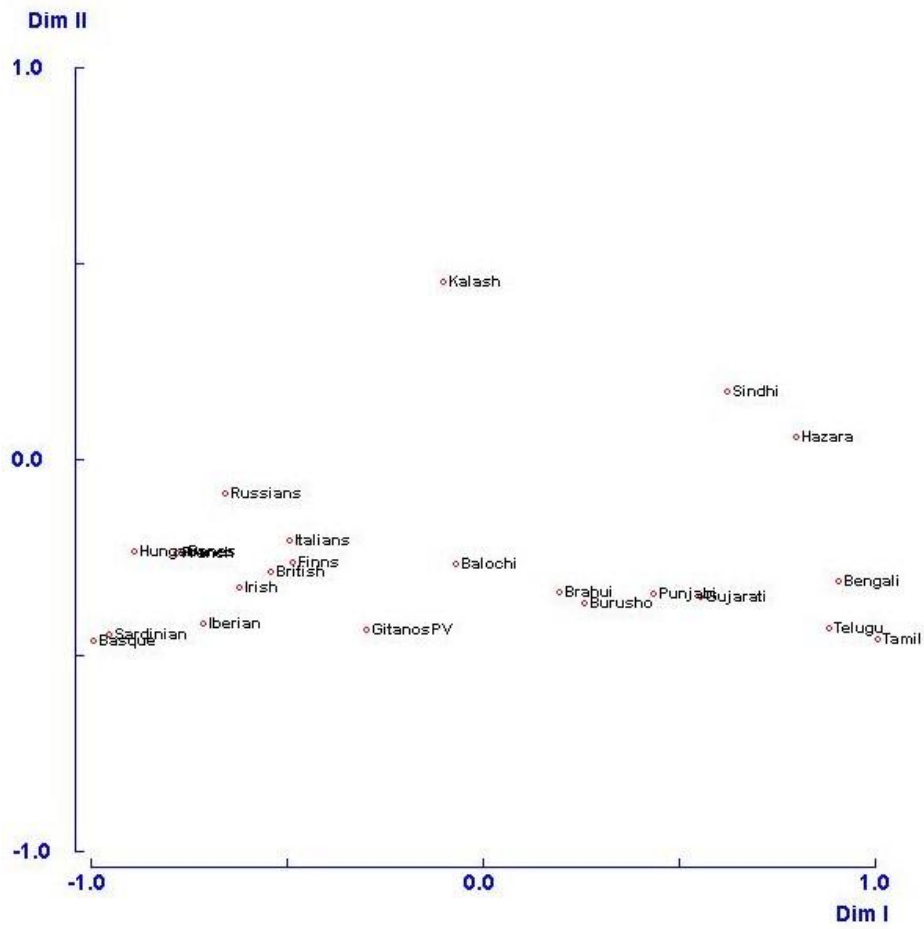


Figura 42: Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz F_{ST} de Reynolds sobre marcadores AIMS, para un grupo de poblaciones euroasiáticas. Stress: 0,016. Leyenda: GitanosPV: Población romaní del País Vasco.

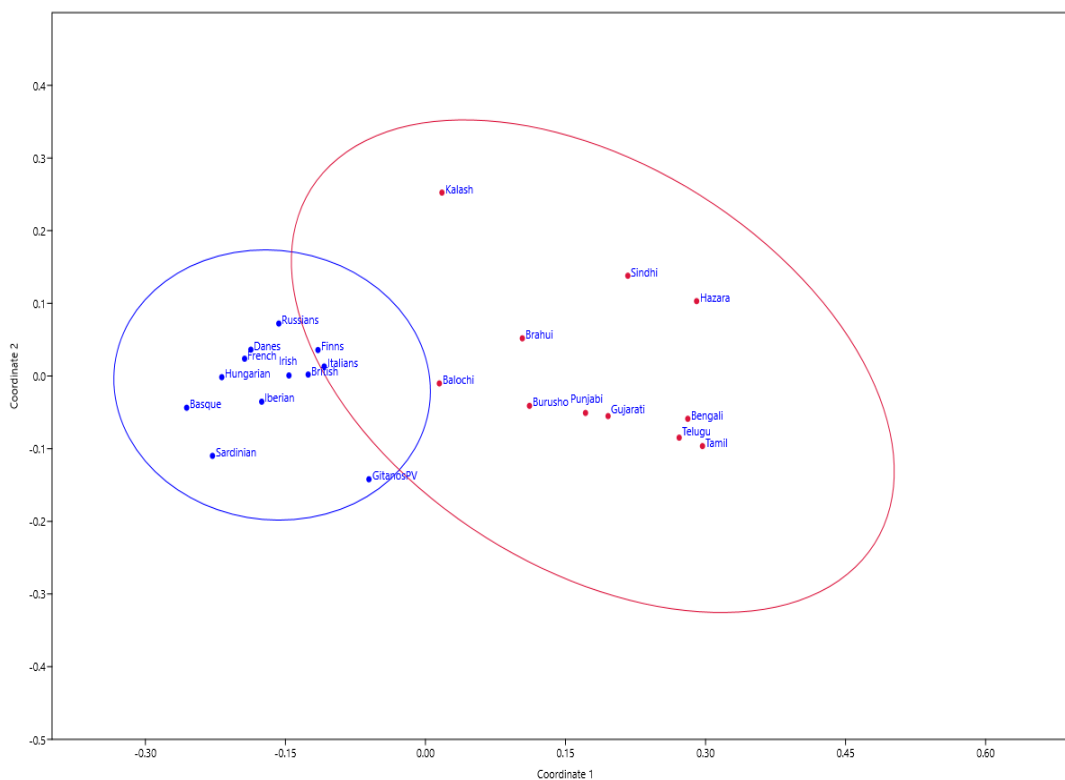


Figura 43: Análisis de Escalamiento Multidimensional (MDS): 95% elipse obtenida a partir de la matriz F_{ST} de Reynolds sobre marcadores AIMs, para un grupo de poblaciones euroasiáticas. Stress: 0,110. Leyenda: GitanosPV: Población romaní del País Vasco.

Para contrastar los resultados obtenidos mediante la distancia F_{ST} de Reynolds, se ha realizado un análisis MDS a partir de la matriz de distancias R de Harpending y Jenkins (1973), adecuada para marcadores bialélicos (Figura 44). En líneas generales los resultados obtenidos con esta nueva matriz muestran un reflejo de los resultados de la Figura 42. Así, se observa en posición central a las poblaciones pakistaníes de Burusho y Brahui, a la población iraní de Balochi y a las poblaciones indias de Punjab y Gujarati. En el extremo más alejado de las europeas se encuentran la población bengalí, las indias Tamil y Telugu, la afgana Hazara y la pakistaní Sindhi. Por último, la población pakistaní de Kalash se sitúa en una posición central para este eje, pero es la más alejada del resto. La población romaní estudiada se encuentra a una distancia similar de este grupo pakistaní, y más alejada de las poblaciones europeas, localizadas en el extremo derecho de la gráfica para dichas poblaciones.

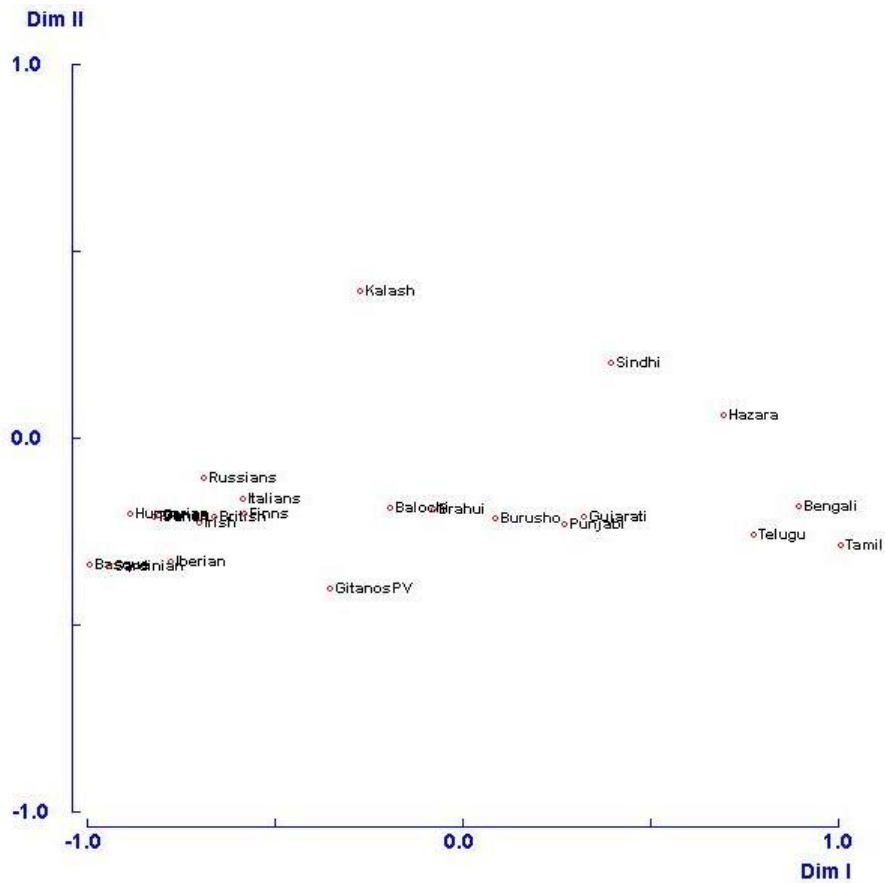


Figura 44: Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz Harpending y Jenkins'R sobre marcadores AIMs, para un grupo de poblaciones euroasiáticas. Stress: 0,019. Leyenda: GitanosPV: Población romaní del País Vasco.

A continuación, se analizó por separado a la población romaní en relación a cada grupo continental. El objetivo era doble; en primer lugar analizar con más detalle cada grupo poblacional y, en segundo lugar, interpretar la posición relativa de la población romaní respecto a las poblaciones europeas y asiáticas, que muestran una gran diferencia en cuanto a su grado de dispersión, lo que puede dificultar la comprensión de dicha localización. En la Figura 45 se observa la posición extrema de la población romaní estudiada frente al resto de poblaciones europeas, que aparecen agrupadas en el extremo negativo de la Dimensión I. Presumiblemente esto refleja una gran distancia genética de este grupo étnico frente al resto de poblaciones, las cuales se disponen a lo largo del eje de ordenadas, desde la población sarda, en el extremo positivo, a la población finlandesa, en un orden aproximadamente inverso a su latitud.

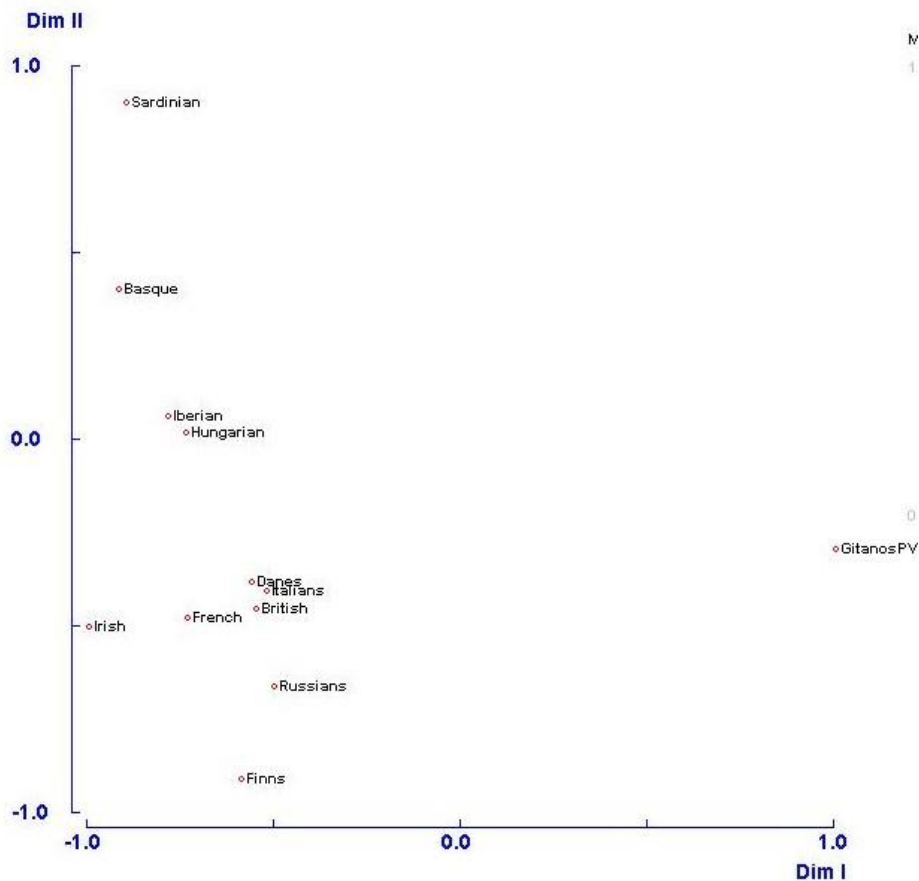


Figura 45: Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz F_{ST} de Reynolds sobre marcadores AIMS, para un grupo de poblaciones europeas y la población romaní de estudio. Stress: 0,092.
 Leyenda: GitanosPV: Población romaní del País Vasco.

En la Figura 46 se representan los resultados obtenidos al comparar la población romaní del País Vasco con las poblaciones asiáticas. La primera se localiza en un extremo de la figura (Dimensiones I y II) y el resto de poblaciones muestran diferentes grados de proximidad con ella. Así, se encuentran relativamente próximas las poblaciones Brahui y Balochi que, a su vez, están cercanas a la población de Burusho. A continuación se sitúan las poblaciones indias de Gujarati y Punjab (origen teórico de la etnia), mientras que el resto están más alejadas.

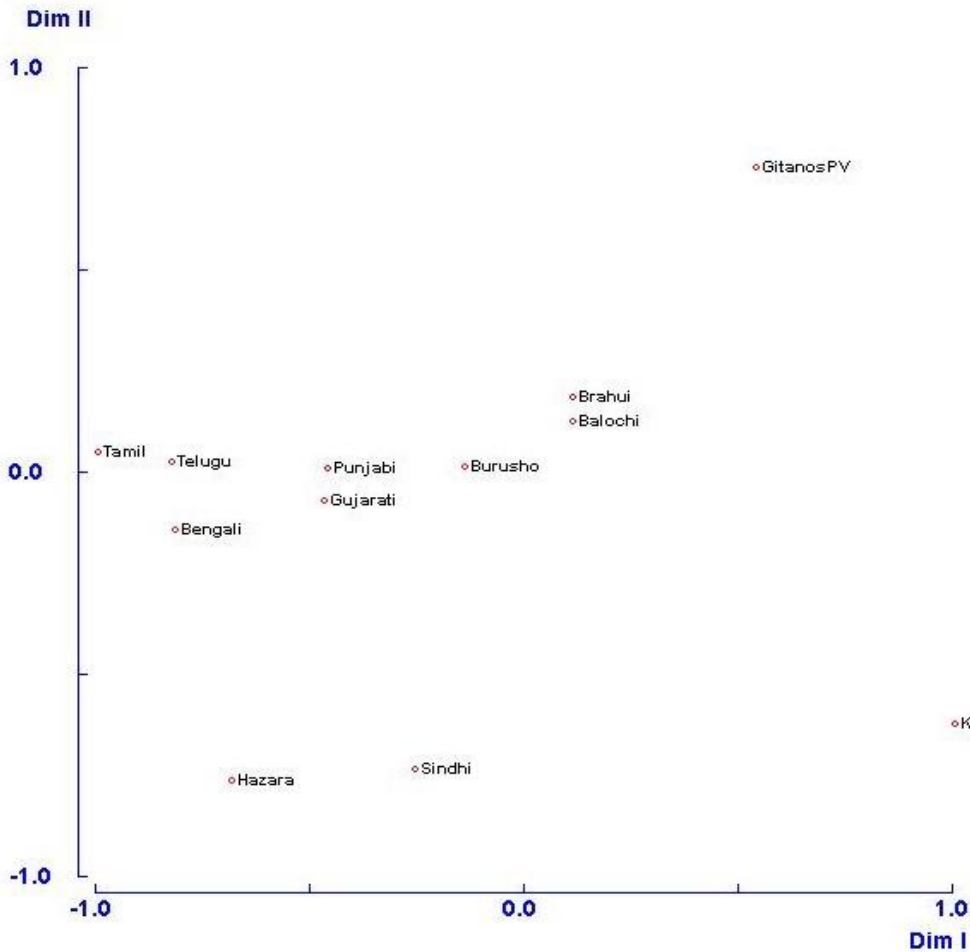


Figura 46: *Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz F_{ST} de Reynolds sobre marcadores AIMs, para un grupo de poblaciones asiáticas y la población romaní de estudio. Stress:0,022. Leyenda: GitanosPV: Población romaní del País Vasco.*

Análisis factorial de correspondencias (AFC)

Mientras que en el anterior apartado se ha comprobado la robustez de los resultados de la matriz F_{ST} mediante el análisis de la matriz R, en este se contrastan los resultados obtenidos mediante los análisis MDS con una serie de Análisis factoriales de correspondencias (AFC). Con el fin de simplificar las figuras y facilitar su interpretación, se ha omitido la representación de los SNPs, puesto que el objetivo no era interpretar la influencia de cada marcador en la posición relativa de las poblaciones. En el primer análisis se observa de nuevo la concentración de las poblaciones europeas, localizadas en el extremo negativo del eje I, frente a la mayor dispersión de las poblaciones asiáticas, situadas en la mitad positiva (Figura 47). Como en los resultados del apartado anterior, la población romaní se localiza en una posición intermedia entre los grupos europeo y asiático. Por otra parte, en ambos análisis, la población romaní está alejada de las poblaciones del País Vasco y de la Península Ibérica, aquellas precisamente con las que convive. De hecho, no se observa un patrón claro de microdiferenciación, ya que las poblaciones

europeas más próximas son, en este análisis, Finlandia e Italia. En todo caso, se confirma que la población más próxima a los romanís estudiados son los Balochi del Sur de Asia, en tanto que la población del Punjab mantiene una notable distancia con la población romaní. Al igual que en los análisis MDS, la población Kalash tiende a desplazarse hacia una posición extrema en el eje II, lo que puede interpretarse como el reflejo de una heterogeneidad notable respecto al resto de poblaciones del Sur de Asia, como se discutirá posteriormente (ver apartado de Discusión).

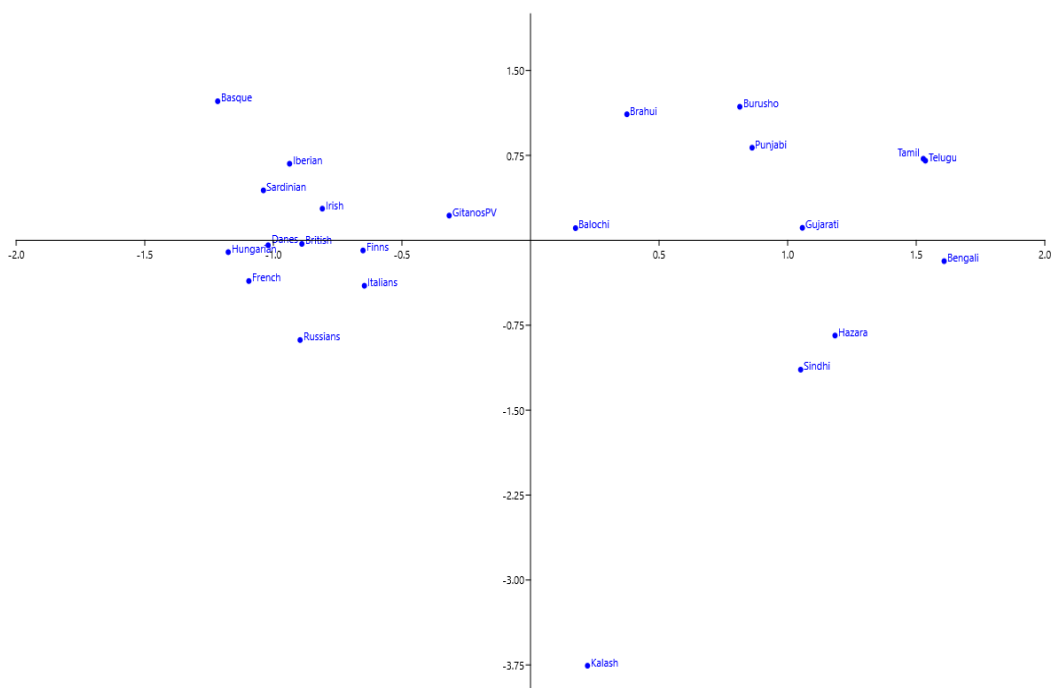


Figura 47: Análisis Factorial de Correspondencias (AFC) sobre marcadores AIMs, para un grupo de poblaciones euroasiáticas. Varianza explicada: 62,04%. Leyenda: GitanosPV: Población romaní del País Vasco.

En el siguiente análisis, realizado exclusivamente sobre las poblaciones europeas (Figura 48) los romaníes del País Vasco se sitúan en una posición extrema, alejándose del resto de poblaciones, las cuales muestran de nuevo, y a grandes rasgos, una configuración polarizada norte-sur.

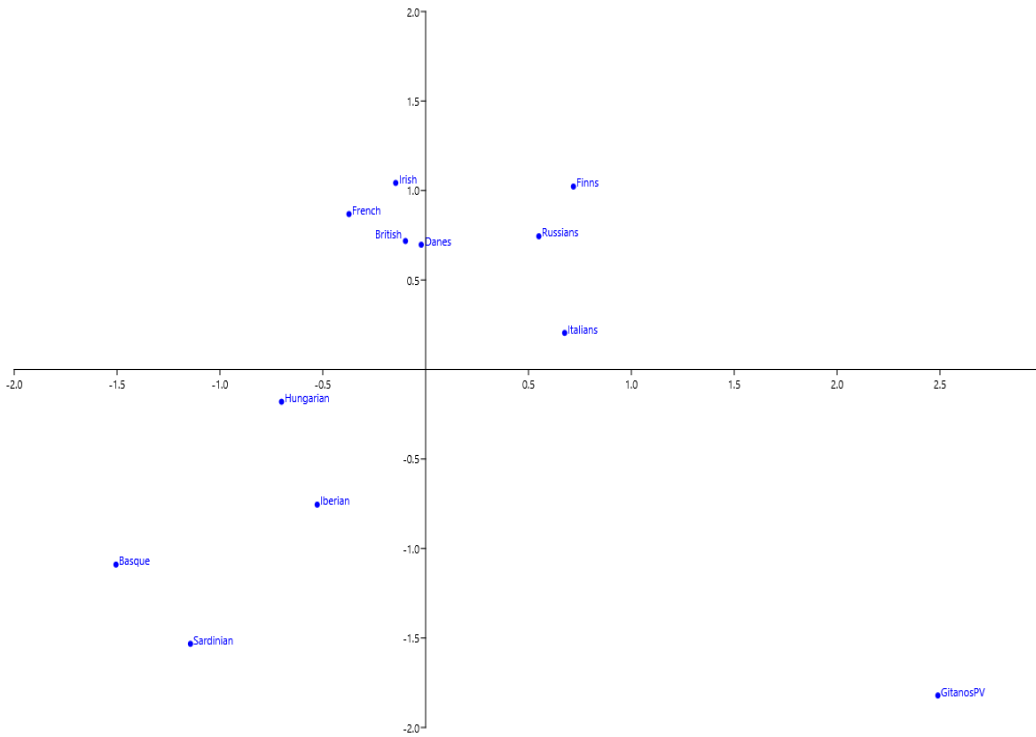


Figura 48: Análisis Factorial de Correspondencias (AFC) sobre marcadores AIMs, para un grupo de poblaciones europeas y la población romaní objeto de estudio. Varianza explicada: 26,82%. Leyenda: GitanosPV: Población romaní del País Vasco.

En el análisis realizado con las poblaciones asiáticas (Figura 49), se confirma que la población romaní no se agrupa ni con las poblaciones europeas ni con las del Sur de Asia, quedando en ambos casos desplazada hacia una posición marginal.

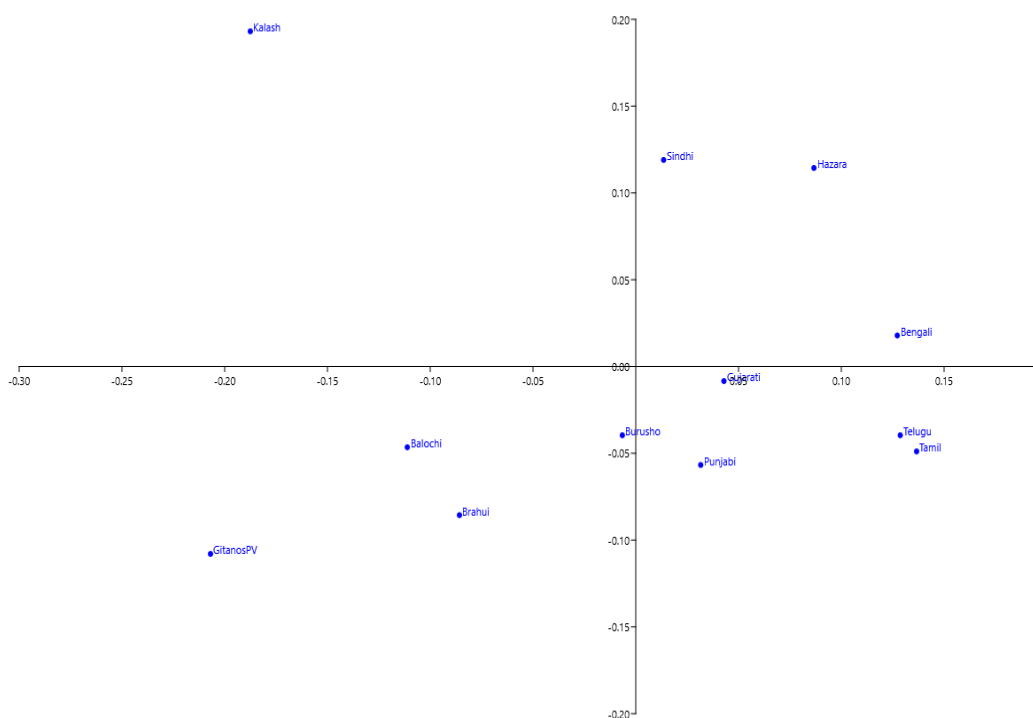


Figura 49: Análisis Factorial de Correspondencias (AFC) sobre marcadores AIMS, para un grupo de poblaciones asiáticas y la población romaní objeto de estudio. Varianza explicada: 36,50%. Leyenda: GitanosPV: Población romaní del País Vasco.

Método del centroide

Con el fin de analizar el flujo génico que ha experimentado la población romaní objeto de estudio, así como su heterocigosidad y la varianza que aporta al conjunto de poblaciones, se realizó un análisis del centroide. En este análisis se incluyeron de nuevo la población romaní del País Vasco, las poblaciones europeas y las poblaciones del Sur de Asia (Figura 50). Las poblaciones que se ubican en el cuadrante superior izquierdo son aquellas que muestran un alto grado de diversidad genética, pero aportan poca heterogeneidad al conjunto; es decir, reflejan un alto grado de mestizaje con otras poblaciones próximas. Este es el caso de las poblaciones de Italia y los Balochi, que aparecían en los análisis MDS y AFC en posiciones próximas al centroide. Las poblaciones que se ubican en el cuadrante superior derecho aportan un alto grado de heterogeneidad al conjunto, presentando también un nivel elevado de diversidad. En general, las poblaciones que se ubican en esta posición reflejan un elevado flujo génico, pero sólo en relación a poblaciones ajenas al área estudiada, como es el caso de la población Brahui. Probablemente por su vecindad con los Balochi ya mencionados, en las figuras previas se ha

observado un alto grado de parentesco genético entre ambos grupos, aunque del análisis del centroide se deduce además algún flujo génico externo, que podría corresponderse con su ubicación geográfica, en una zona que es el paso natural entre Irán y Asia Central con el subcontinente indio.

Otra parte relevante del gráfico centroide es el cuadrante inferior derecho, en el que se ubican poblaciones con poca diversidad genética, que aportan una notable heterogeneidad genética al conjunto. Es decir, poblaciones aisladas, con poco flujo génico con otras poblaciones. El resto de poblaciones se distribuyen no muy alejadas, a ambos lados de la línea de regresión que muestra los valores esperados de heterogeneidad frente a diversidad. En todo caso, las poblaciones del Sur de Asia muestran en general una heterocigosidad mayor que las europeas, fruto sin duda de la variabilidad extrema del continente indio. En lo que respecta a las poblaciones europeas que se encuentran por debajo de la recta de regresión, su situación indica cierto grado de aislamiento, pero presumiblemente este resultado es debido a la gran heterogeneidad de las poblaciones del Sur de Asia, que termina desplazando la recta y sitúa a las poblaciones europeas en valores inferiores a los esperados.

Por último, la posición en el extremo inferior de la población vasca (Basque) indica también cierto grado de aislamiento, a la par de una gran heterogeneidad, posiblemente por haber experimentado un grado alto de deriva a causa de un tamaño poblacional reducido y un aislamiento prolongado. Por su parte, la población romaní del País Vasco se sitúa muy cerca de la recta de regresión, por lo que no se observa un elevado grado de aislamiento ni tampoco un intenso flujo génico; esto podría deberse a que se han valorado ambas variables al mismo tiempo, por lo que no se puede abstraer el efecto individual de cada una en la población.

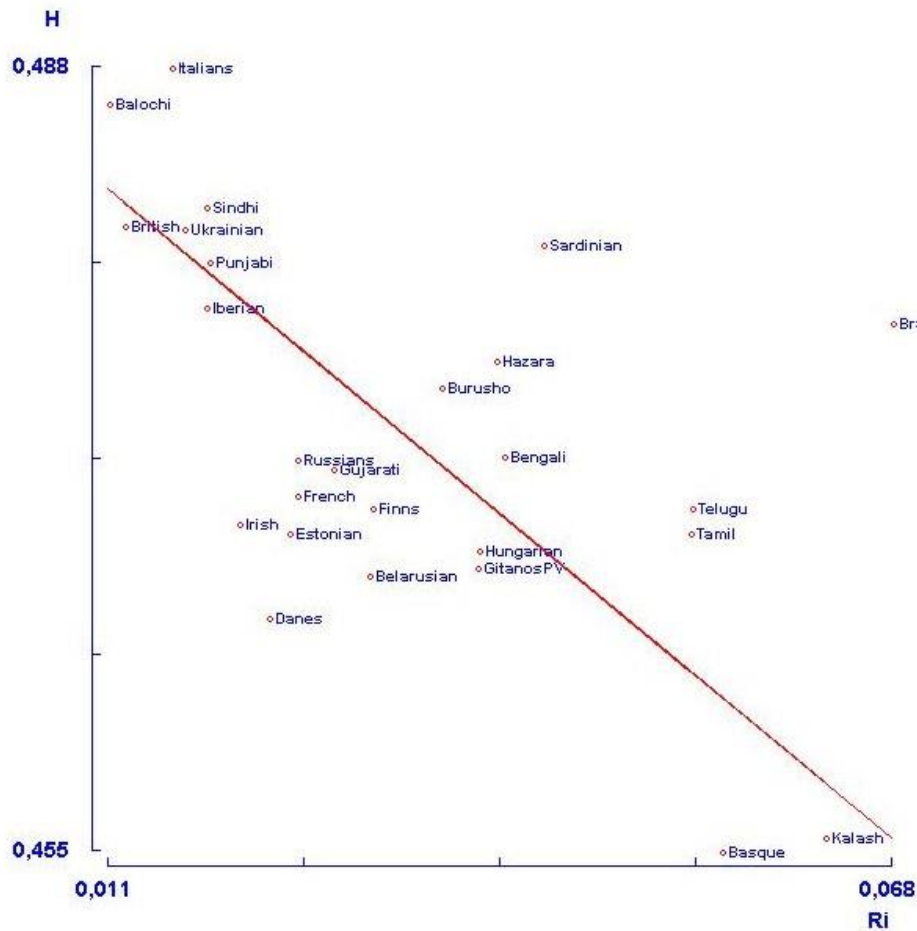


Figura 50: Método del centroide sobre marcadores AIMs para un grupo de poblaciones euroasiáticas y la población romaní objeto de estudio. Leyenda: GitanosPV: Población romaní del País Vasco.

Análisis de clinas

En varios de los SNPs analizados sólo se han podido obtener frecuencias de un grupo reducido de poblaciones (entre 9 y 16), por lo que han sido desechados para el análisis de clinas. Por ello, la base de datos utilizada para este análisis se ha visto reducida a 25 SNPs en 22 poblaciones de Europa y el Sur de Asia. Los resultados obtenidos con GenoCline se muestran en la Tabla 8 y en la Figura 51. Se observan clinas significativas en 22 de los 25 SNPs, que se reducen a 17 cuando se aplica la corrección de Bonferroni para comparaciones múltiples. El hecho de que se hayan seleccionado una serie de marcadores informativos de ancestralidad (AIMs) y, por tanto, con frecuencias *a priori* diferentes en Europa y el Sur de Asia no implica que necesariamente se observe un gradiente geográfico, pero desde luego facilita que suceda. La mayor parte de las clinas muestran una orientación nor-noroeste sur-sudeste. En principio parecería más lógica una orientación más alejada del eje latitudinal por la disposición de ambas regiones geográficas, pero cabe recordar que la tendencia mayoritaria en los gradientes de frecuencias en Europa es prácticamente latitudinal, debido a la variación geográfica de las

proporciones de un origen del Neolítico y de la Estepa Póntica en el patrimonio genético europeo (Haak et al., 2015). En la India cabe esperar una distribución similar, ya que los componentes indoeuropeo y dravídico se distribuyen de forma desigual de sur a norte del subcontinente. Si se considera además que Europa se sitúa más al norte que el Sur de Asia, puede explicarse la orientación predominante de las clinas.

Tabla 8: Características de las clinas observadas en las frecuencias de 25 SNPs, en 22 poblaciones del sudoeste de Eurasia.

SNP	Ángulo	Intersección	Pendiente	r	df	p	Test de Bonferroni
rs11102524	3	0,391763	0,233116	0,714631	20	0,000186	p<0,01
rs4908602	179	0,372976	0,142877	0,479529	20	0,023928	
rs6429223	359	0,722478	-0,28214	0,719738	20	0,000159	p<0,01
rs182549	158	0,439402	0,548441	0,676405	20	0,000548	p<0,05
rs2920059	78	0,481744	0,167995	0,569839	20	0,00563	
rs896788	159	0,156102	0,400047	0,890641	20	2,79E-08	p<0,01
rs10470510	146	0,471416	0,096648	0,430104	20	0,045721	
rs9374640	342	0,601846	-0,26726	0,701454	20	0,000275	p<0,01
rs917118	344	0,736902	-0,25551	0,713527	20	0,000193	p<0,01
rs1392121	137	0,315642	0,219654	0,718511	20	0,000165	p<0,01
rs10123957	344	0,498371	-0,14495	0,535402	20	0,010233	
rs2710293	184	0,705188	-0,25932	0,732437	20	0,000106	p<0,01
rs1886510	160	0,443587	0,410471	0,869528	20	1,49E-07	p<0,01
rs2065982	124	0,015073	0,232303	0,827544	20	2,02E-06	p<0,01
rs730570	262	0,905632	-0,36905	0,892531	20	2,36E-08	p<0,01
rs1834640	353	1,043736	-0,3824	0,692944	20	0,00035	p<0,01
rs881929	166	0,462254	0,396091	0,691254	20	0,000367	p<0,01
rs2033719	284	0,538047	-0,20391	0,614993	20	0,002319	
rs2303798	150	-0,01341	0,189408	0,835876	20	1,28E-06	p<0,01
rs6057059	137	0,438853	0,178631	0,723842	20	0,00014	p<0,01
rs722098	240	0,857053	-0,39422	0,706041	20	0,000241	p<0,01
rs2040411	255	0,697406	-0,26282	0,827017	20	2,08E-06	p<0,01

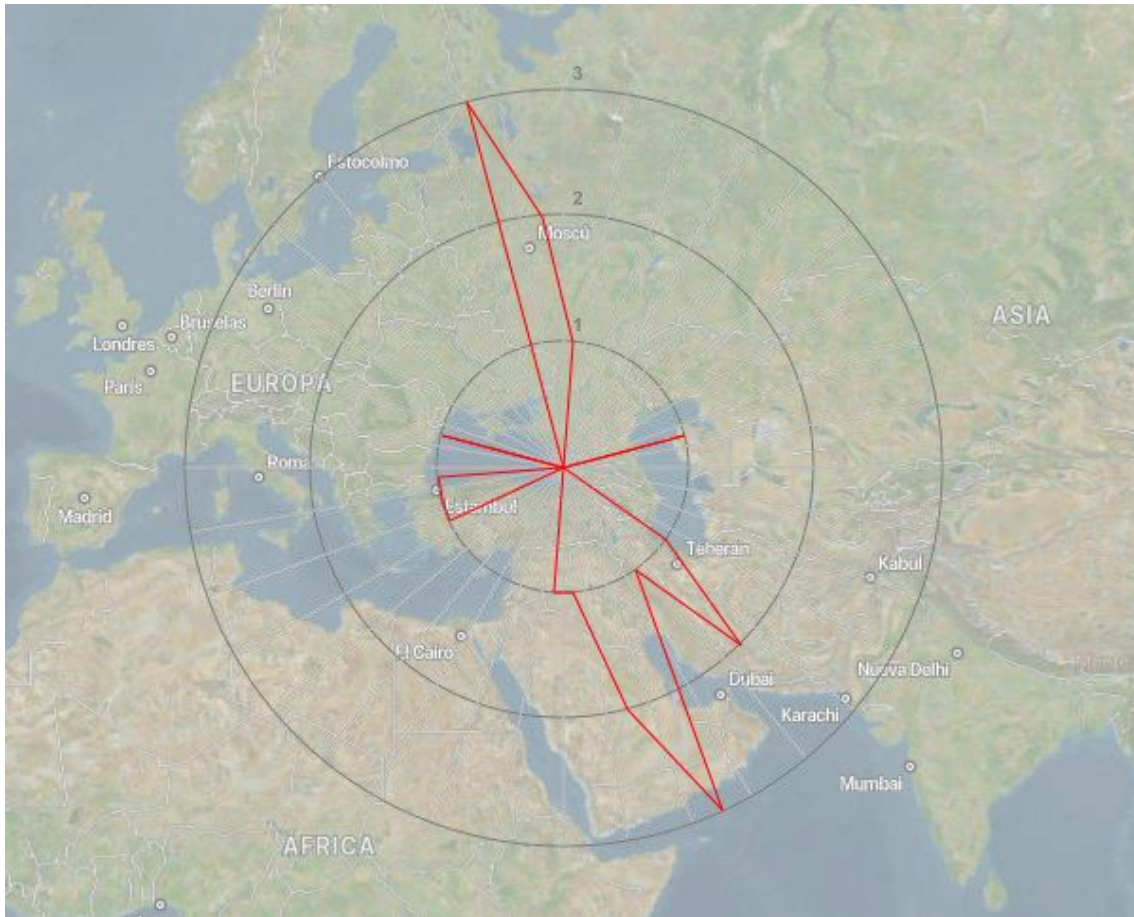


Figura 51: Orientación y frecuencia de las clinas observadas en las frecuencias de 25 SNPs, en 22 poblaciones del sudoeste de Eurasia.

Con el fin de valorar la posible relación entre un determinado modelo de clina y la posición de la población romaní del País Vasco, se han analizado los patrones de las clinas obtenidas. No se ha observado ningún patrón claro, encontrándose una gran variedad de situaciones. Así, como puede apreciarse en las dos clinas más significativas correspondientes a los SNPs rs896788 y rs730570, los romaníes se pueden situar tanto dentro del rango de frecuencias del Sur de Asia (Figura 52) como dentro del europeo (Figura 53), con independencia de la orientación de la clina. En definitiva, estos resultados indican la no influencia en el patrimonio genético romaní de un proceso masivo de flujo génico que hubiese afectado al sudoeste de Asia (ver Discusión).

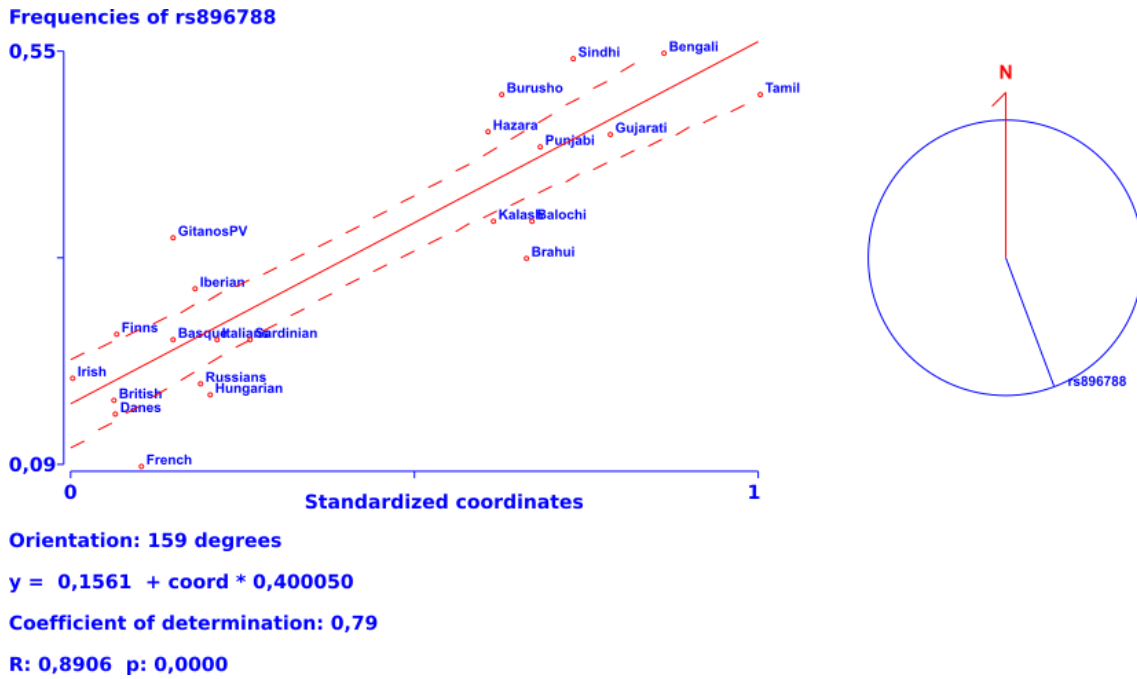


Figura 52: Características de la clina observada para el SNP rs896788.

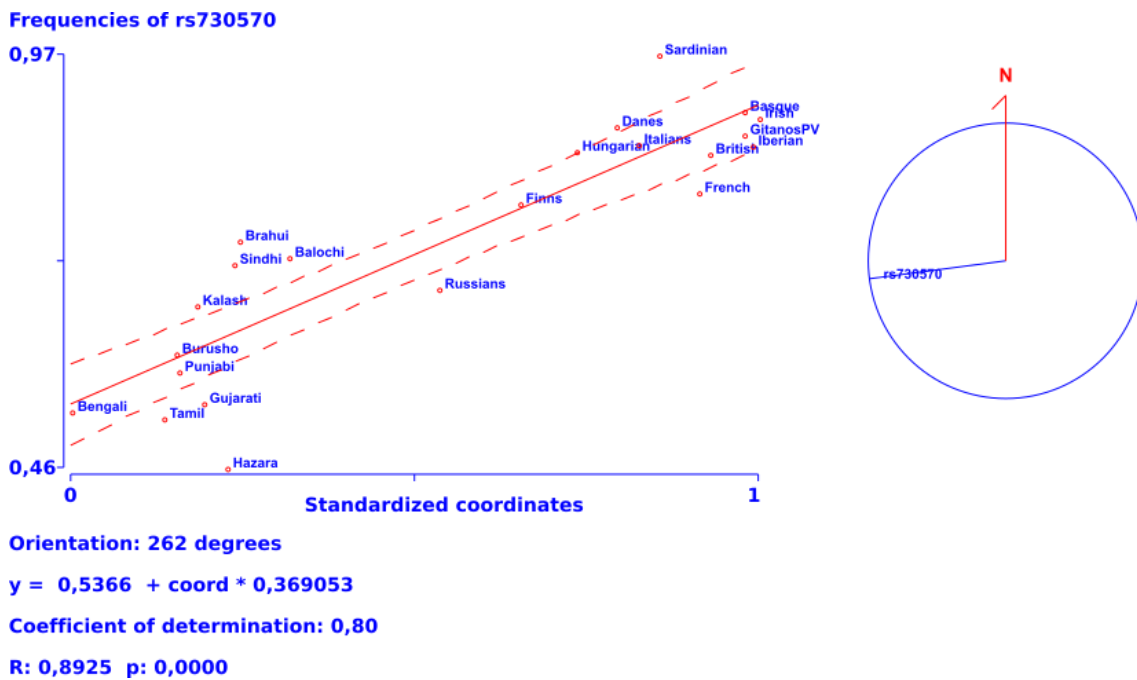


Figura 53: Características de la clina observada para el SNP rs730570.

Possible acción de la deriva genética

En la Tabla 9 se muestran los valores obtenidos de $-\log(P)$ para el cociente entre el valor de F_{ST} observado y el esperado en dos situaciones diferentes: i) considerando las frecuencias en romaníes iguales a las frecuencias en el noroeste de la India por un lado (F_{ST} min), y ii) considerando las frecuencias en romaníes intermedias entre España y el noroeste de la India por otro (F_{ST} med). Como ya se ha señalado en el apartado de Material y Métodos, en el primer caso no se considera la posible existencia de flujo génico y en el segundo se considera que hay un 50%. Podría haberse analizado un rango más amplio de valores de mestizaje en un proceso iterativo con el fin de obtener una mayor resolución, pero como se observa en la tabla, los SNPs con valores significativos de $-\log(P)$ son prácticamente los mismos en ambas situaciones, con la única excepción de rs182549, que no muestra significación para la primera pero sí para la segunda. En el SNP rs2770, el valor de $-\log(P)$ no se ha podido precisar, ya que no ha sido posible obtener más de 323 decimales en el test de la F, alcanzándose el límite del procesador del ordenador.

En conjunto, se han observado 21 SNPs con una heterogeneidad significativa entre el valor observado y el esperado de F_{ST} ; 22 en el caso en que se considera el flujo génico (F_{ST} med). Es decir, se ha identificado una posible acción significativa de la deriva en una proporción de entre el 46 y el 48% de los SNPs analizados. En la mayor parte de los casos, el valor de la frecuencia en la población romaní muestra un valor extremo, fuera del rango de frecuencias de Europa y el Sur de Asia, como puede observarse en la Figura 41. Por otra parte se observa que los valores de $-\log(P)$ son muy heterogéneos, como cabría esperar de la acción de la deriva genética. De hecho, debería esperarse, una vez ordenados los valores, que éstos siguieran una distribución exponencial negativa; así, como se observa en la Figura 54, se ajustan razonablemente bien a dicha función.

Tabla 9: Valores obtenidos del cociente de F_{ST} observado y esperado, con su correspondiente P asociada y el valor de $-\log(P)$. Los valores significativos se marcan con un asterisco.

SNP	F_{ST} obs/			F_{ST} obs/		
	F_{ST} min	p(F)	$-\log(p)$	F_{ST} med	P(F)	$-\log(P)$
rs11102524	1,7875	1,67E-01	0,7763	2,3797	9,26E-02	1,0335
rs4908602	44,2052	6,35E-20	19,1972*	58,8309	2,83E-26	25,5484*
rs6429223	1,5886	2,04E-01	0,6899	2,0928	1,23E-01	0,9089
rs1019264	2,0432	1,30E-01	0,8874	2,7110	6,65E-02	1,1774
rs182549	2,1809	1,13E-01	0,9472	3,1167	4,43E-02	1,3536*
rs2920059	1,1657	3,12E-01	0,5062	1,5664	2,09E-01	0,6803
rs896788	0,9596	3,83E-01	0,4167	1,2087	2,99E-01	0,5249
rs10470510	4,5396	1,07E-02	1,9715*	6,0555	2,34E-03	2,6299*
rs1498444	0,3721	6,89E-01	0,1616	0,5008	6,06E-01	0,2175
rs4455309	0,4253	6,54E-01	0,1847	0,5544	5,74E-01	0,2408
rs9870523	3,9241	1,98E-02	1,7042*	5,2435	5,28E-03	2,2772*
rs1157106	1,1322	3,22E-01	0,4917	1,5023	2,23E-01	0,6525
rs116818505	31,1433	2,99E-14	13,5250*	42,5851	3,21E-19	18,4937*
rs200789833	3,2485	3,88E-02	1,4108*	4,1205	1,62E-02	1,7895*
rs2071350	10,9595	1,74E-05	4,7596*	13,8343	9,82E-07	6,0081*
rs2770	619,4666	1,37E-269	268,8644*	825,4535	0,00E+00	>323*
rs369150	1,1966	3,02E-01	0,5197	1,6812	1,86E-01	0,7301
rs3823324	3,8060	2,22E-02	1,6529*	5,1110	6,03E-03	2,2197*
rs72500564	59,3573	1,67E-26	25,7770*	79,7504	2,33E-35	34,6324*
rs72873921	4,1487	1,58E-02	1,8017*	5,2847	5,07E-03	2,2951*
rs72873922	4,3828	1,25E-02	1,9034*	5,5829	3,76E-03	2,4246*
rs79244404	355,5254	4,49E-155	154,3478*	473,1505	4,08E-206	205,3895*
rs9273352	2,1148	1,21E-01	0,9184	2,8083	6,03E-02	1,2196
rs9277332	3,3319	3,57E-02	1,4470*	4,4080	1,22E-02	1,9143*
rs9277413	16,3655	7,81E-08	7,1073*	22,0975	2,53E-10	9,5966*

Resultados

rs9277418	12,0822	5,66E-06	5,2472*	16,3360	8,04E-08	7,0945*
rs9277498	63,3472	3,09E-28	27,5096*	83,7921	4,10E-37	36,3874*
rs9374640	1,4962	2,24E-01	0,6498	1,9975	1,36E-01	0,8675
rs7783055	19,5797	3,14E-09	8,5032*	26,1401	4,44E-12	11,3522*
rs917118	0,6806	5,06E-01	0,2956	0,8772	4,16E-01	0,3810
rs1392121	0,6731	5,10E-01	0,2923	0,8791	4,15E-01	0,3818
rs10123957	16,1592	9,60E-08	7,0177*	21,6678	3,89E-10	9,4100*
rs1493489	5,3986	4,52E-03	2,3446*	7,1987	7,48E-04	3,1263*
rs2710293	7,6639	4,69E-04	3,3284*	10,2658	3,48E-05	4,4583*
rs1886510	1,1564	3,15E-01	0,5022	1,6035	2,01E-01	0,6964
rs2065982	1,0008	3,68E-01	0,4346	1,1147	3,28E-01	0,4841
rs730570	1,4909	2,25E-01	0,6475	1,8289	1,61E-01	0,7943
rs1834640	1,6142	1,99E-01	0,7010	1,7074	1,81E-01	0,7415
rs8032154	0,8139	4,43E-01	0,3535	1,1071	3,31E-01	0,4808
rs881929	1,2038	3,00E-01	0,5228	1,6642	1,89E-01	0,7227
rs2033719	1,0354	3,55E-01	0,4497	1,3966	2,47E-01	0,6065
rs1024116	1,0107	3,64E-01	0,4389	1,3545	2,58E-01	0,5883
rs16951664	4,0109	1,81E-02	1,7419*	5,3450	4,77E-03	2,3213*
rs2285513	1,4343	2,38E-01	0,6229	1,9207	1,47E-01	0,8341
rs2303798	1,8565	1,56E-01	0,8063	2,0257	1,32E-01	0,8798
rs6057059	0,6779	5,08E-01	0,2944	0,9051	4,04E-01	0,3931
rs722098	1,0135	3,63E-01	0,4402	1,2762	2,79E-01	0,5543
rs2040411	1,7191	1,79E-01	0,7466	2,2676	1,04E-01	0,9848

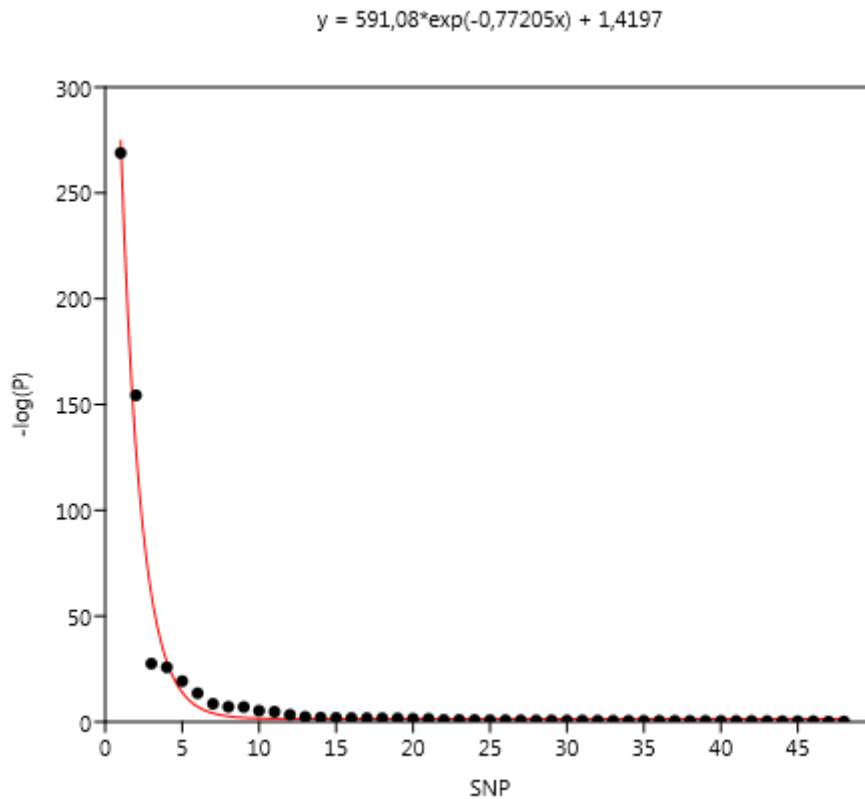


Figura 54: Regresión exponencial negativa sobre los valores ordenados de $-\log(P)$ para 48 SNPs. Los valores representados son los del modelo con flujo génico.

Estimación del grado de mestizaje

Con el fin de establecer el grado de flujo génico experimentado por la población romaní del País Vasco, se ha abordado la cuantificación del grado de mestizaje. Para ello, se han considerado dos hipotéticas poblaciones parentales. Por un lado, la población antecesora del Sur de Asia de la que se originó presumiblemente el patrimonio genético romaní y, por otro, la población europea. Se ha seleccionado a la muestra punjabí como representante de la población parental de Asia, ya que es la única disponible que se encuentra en la región supuestamente originaria de los romaníes. Por otra parte, se ha considerado al País Vasco como la población de la que ha procedido el flujo génico hacia los romaníes estudiados. Puesto que aproximadamente la mitad de su censo puede considerarse de origen vasco y la otra mitad inmigrante, sus frecuencias se han estimado como la media de las frecuencias de ambas muestras (vascos y población general española). Como ya se ha señalado anteriormente, se ha detectado una posible acción de la deriva en algunos SNPs, por lo que sus frecuencias alélicas podrían

encontrarse alteradas (lo que afectaría a los cálculos del mestizaje), por lo que se han eliminado aquellos con un valor de $-\log(P)$ significativo.

En la Tabla 10 se muestra el resultado del análisis de mestizaje. Ambos componentes, el parental del sur de la India y el originado por el mestizaje europeo, muestran unos valores similares, aunque se encuentra ligeramente por encima el predicho para el componente europeo. En efecto, la población romaní del País Vasco mantendría un 47,3% de su patrimonio original proveniente del noroeste de la India y habría experimentado un flujo génico posterior que habría afectado al 52,7% de dicho patrimonio.

Tabla 10: Estimación de la proporción del patrimonio genético de los romaníes del País Vasco proveniente de dos hipotéticas poblaciones parentales, error estándar e intervalo de confianza del 95%.

Pob. parental	Coefficientes	Error estándar	Límite inferior	Límite superior
Punjab	0,473	0,109	0,246	0,7
País Vasco	0,527	0,104	0,311	0,742

V.2. Análisis de la población romaní del País Vasco para elementos *Alu*

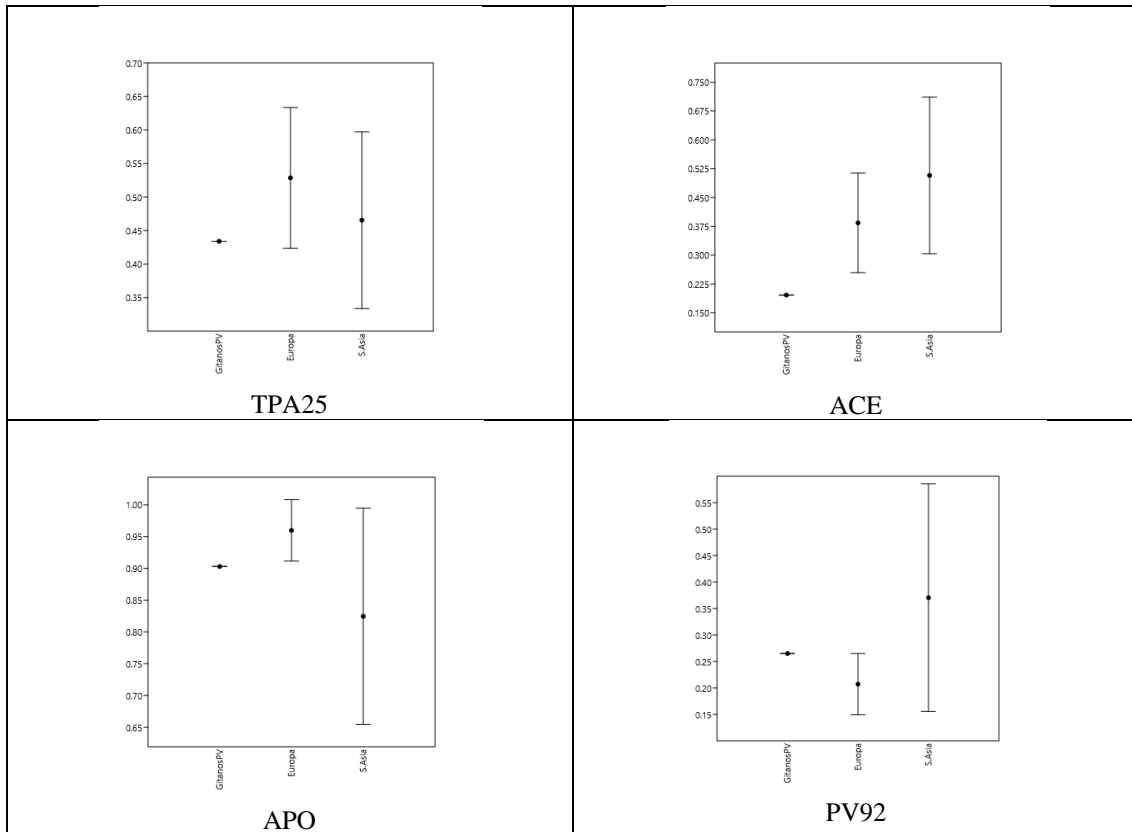
V.2.1 Frecuencias de inserción de los 8 elementos *Alu*

En la Tabla 11 se muestran las frecuencias de las 8 inserciones *Alu* analizadas en la población romaní. El rango de valores es amplio, ya que oscilan entre 0,196 (ACE) y 0,903 (APO). El test de equilibrio Hardy-Weinberg revela desequilibrio en la mitad de las inserciones después de la corrección de Bonferroni. En todos los casos el origen del desequilibrio es un exceso de homocigotos, lo que podría deberse a un cierto grado de endogamia.

Tabla 11: Tamaño de muestra, alelo principal (Inserción), frecuencia del alelo principal en la muestra romaní del País Vasco, error estándar, resultado del test Chi-2 de equilibrio Hardy-Weinberg y corrección de Bonferroni, para comparaciones múltiples de las inserciones *Alu* analizadas.

Inserción	N	Frecuencia	Error estándar	Chi-2	p	p (Bonferroni)
A25	71	0,155	0,043	1,380	2,4E-01	n.s.
PV92	66	0,265	0,052	21,618	3,3E-06	P<0,01
B65	64	0,422	0,059	15,209	9,6E-05	P<0,01
TPA25	61	0,434	0,059	5,469	1,9E-02	n.s.
FXIIIB	63	0,238	0,051	14,214	1,6E-04	P<0,01
D1	67	0,313	0,055	22,839	1,8E-06	P<0,01
ACE	74	0,196	0,047	0,385	5,3E-01	n.s.
APO	72	0,903	0,035	0,835	3,6E-01	n.s.

A continuación, se ha elaborado una base de datos con las frecuencias observadas en la población romaní estudiada junto a otras 35 poblaciones euroasiáticas adquiridas de otras fuentes (Comas et al., 2000; Garcia-Obregón et al., 2006; Stoneking et al., 1997; Comas et al., 2004; Varzari et al., 2007; Nasidze et al., 2001; Xiao et al., 2002), tal y como se muestra en la Tabla 2 del Anexo. En la Figura 55 se observa, para las 8 inserciones *Alu* analizadas, la frecuencia en la población romaní del País Vasco, así como los errores estándar y la media en Europa y el Sur de Asia. En tres de ellos, el valor en la población romaní se sitúa entre las medias europeas y asiática; sólo en un caso se encuentra por encima de ambos valores, y en cuatro por debajo. Además, en una inserción (ACE) la frecuencia en romaníes se encuentra fuera del rango de los errores estándar conjuntos de Europa y el Sur de Asia, y por tanto muy alejados de los valores medios de ambas regiones geográficas.



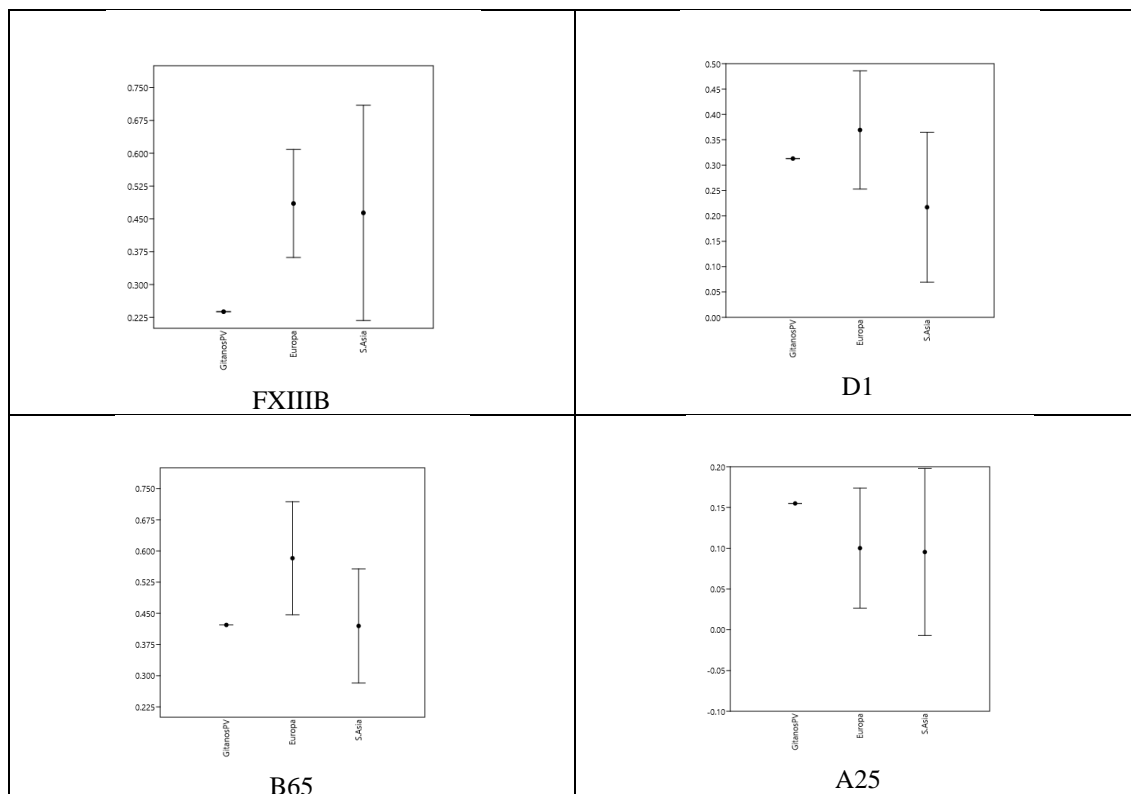


Figura 55: Frecuencia en la población romaní del País Vasco, valor medio y errores estándar en Europa y el Sur de Asia, de las inserciones *Alu* analizadas.

Análisis molecular de la varianza (AMOVA)

Al igual que con los SNPs, se han comparado las frecuencias de las inserciones *Alu* en los romaníes del País Vasco con las de otras poblaciones euroasiáticas obtenidas de diferentes fuentes (Comas et al., 2000; Comas et al., 2004; Garcia-Obregón et al., 2006; Nasidze et al., 2001; Stoneking et al., 1997; Varzari et al., 2007) (Tabla 3 del Anexo). En la Tabla 12 se muestran los resultados del AMOVA para las 8 inserciones *Alu* en los dos grupos continentales analizados (Europa y Sur de Asia). Este análisis se realizó con el objetivo de valorar las diferencias entre ambos grupos y para comprobar la capacidad diferenciadora de las inserciones *Alu* en este contexto geográfico. Los valores de F_{ST} revelan una alta significación ($P < 0,001$) en todas las inserciones, lo que indica que hay una gran heterogeneidad entre los individuos dentro de las poblaciones consideradas. En cuanto a la heterogeneidad entre poblaciones dentro de los grupos continentales (mostrado por el valor F_{SC}), todos los marcadores han mostrado diferencias altamente significativas ($P < 0,001$), lo que revela de nuevo una gran heterogeneidad entre las poblaciones. Finalmente, se observa una alta heterogeneidad, pero menor que en los niveles de organización anteriores, entre los grupos continentales (F_{CT}), mostrando tres de las ocho inserciones *Alu* un nivel de significación bajo ($P < 0,05$) y otras cinco un nivel elevado ($P < 0,001$).

En definitiva, las ocho inserciones analizadas resultan útiles para la comparación de las frecuencias de la población romaní con los grupos europeo y asiático.

Tabla 12: Índices de fijación generados mediante el análisis AMOVA para las frecuencias de 8 inserciones Alu en poblaciones euroasiáticas. (F_{CT} : varianza genética entre grupos. F_{SC} : varianza genética entre poblaciones dentro de grupos. F_{ST} : varianza genética entre individuos dentro de una población) (* $P < 0,05$; ** $P < 0,01$; * $P < 0,001$).**

	Índices de Fijación		
	F_{ST}	F_{SC}	F_{CT}
TPA25	0,01303***	0,01019***	0,00288*
ACE	0,08514***	0,05130***	0,03567***
APO	0,09928***	0,04258***	0,05922***
PV92	0,30415***	0,16631***	0,16533***
FXIIB	0,16573***	0,14313***	0,02637*
D1	0,04355***	0,02334***	0,02069***
B65	0,06023***	0,03641***	0,02472***
A25	0,02372***	0,02206***	0,00170*

Análisis de escalamiento multidimensional (MDS)

Se ha realizado un análisis MDS sobre la matriz de distancias F_{ST} de Reynolds con las 24 poblaciones de Europa y el Sur de Asia en las que han sido analizadas las 8 inserciones *Alu* (Figura 56). Las poblaciones europeas forman una densa nube de puntos localizada en el cuadrante inferior izquierdo de la figura. Las poblaciones asiáticas se encuentran mucho más dispersas por todo el resto del gráfico. En la mitad positiva del eje de abscisas (Dimensión I) se localizan las poblaciones tamil, indo-cristianos e indo-hindús. La población de indo-musulmanes presenta una posición relativa más próxima a los pastunes de Afganistán, situadas ambas próximas al centroide. En el cuadrante superior izquierdo se sitúan las poblaciones pakistaní, Siria y Emiratos, y próxima a ellas la población romaní del País Vasco (denominada en estos análisis BSR).

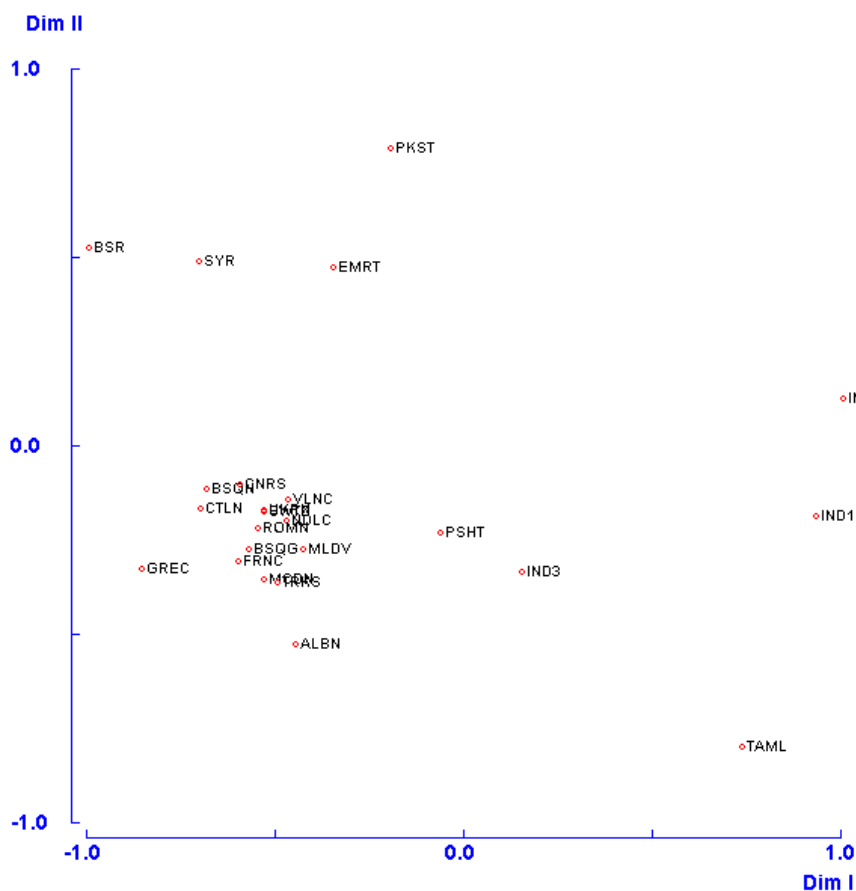


Figura 56: Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz F_{ST} de Reynolds para 8 inserciones Alu en poblaciones euroasiáticas y la población romaní estudiada. Stress: 0,037. Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), FRNC: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucrainianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, IND1: India-Cristianos, IND2: India-Hindús, IND3: India-Musulmanes.

Con el fin de reflejar la dispersión relativa de los dos grupos poblacionales (Sur de Asia y Europa) frente a la población romaní analizada, se han elaborado las elipses equiprobables del 95% para el MDS realizado sobre la matriz de distancias F_{ST} (Figura 57). Como en el caso de los SNPs analizados, la variabilidad inter poblacional es mucho mayor en Asia, hasta el punto de incluir en su interior la elipse correspondiente a Europa. La población romaní aparece en un extremo de la elipse asiática, pero alejada de la elipse europea.

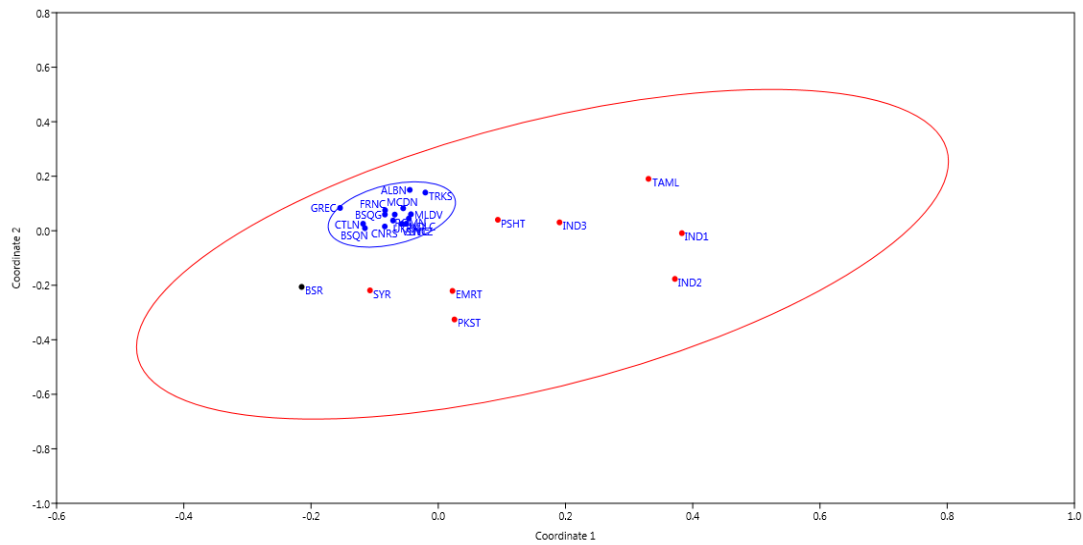


Figura 57: Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz F_{ST} de Reynolds para 8 inserciones Alu en poblaciones euroasiáticas y la población romaní estudiada, mostrando las elipses equiprobables del 95% para cada grupo poblacional. Stress: 0,113. Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNK: Valencianos (España), CNRS: Canarios (España), FRNC: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucranianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, IND1: India-Cristianos, IND2: India-Hindús, IND3: India-Musulmanes.

Quando se utiliza la distancia de Harpending y Jenkins (Figura 58), las posiciones relativas de las poblaciones son similares. Así, la población romaní del País Vasco aparece también próxima a las poblaciones de Siria y Emiratos Árabes pero, en este caso, más distanciada de la de Pakistán.

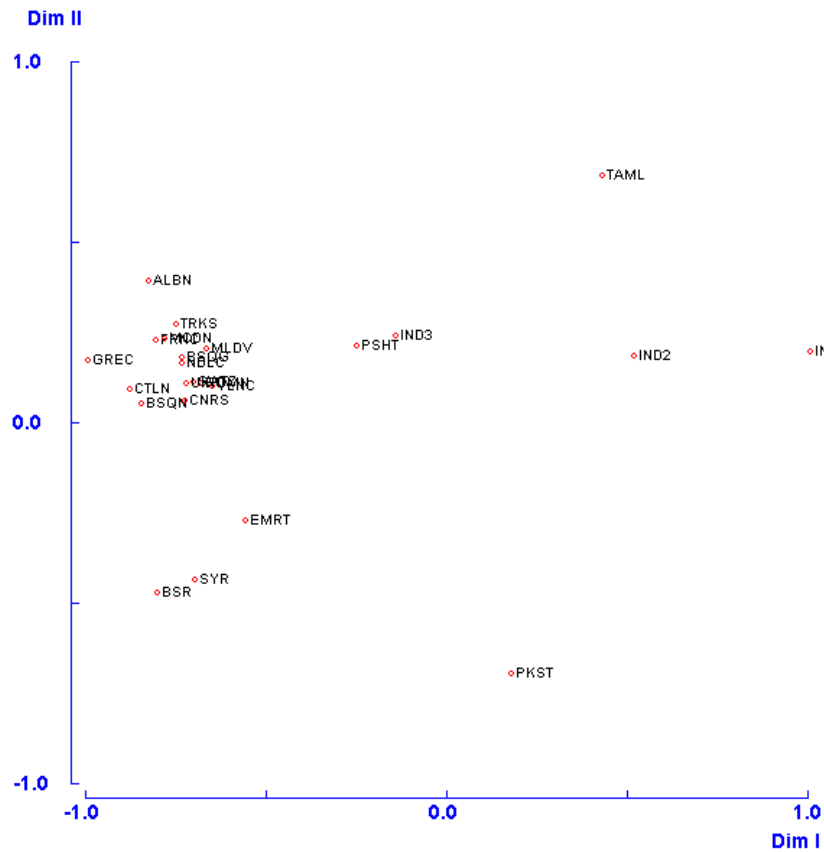


Figura 58: Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz R de Harpending y Jenkins para 8 inserciones Alu en poblaciones euroasiáticas y la población romaní estudiada. Stress: 0,048. Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), FRNC: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucranianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, IND1: India-Cristianos, IND2: India-Hindús, IND3: India-Musulmanes.

Siguiendo la metodología usada para los SNPs, se ha valorado por separado el parentesco genético de la población romaní con las poblaciones europeas y del Sur de Asia. En lo que respecta a las europeas (Figura 59) se observa a los romaníes situados en el extremo negativo del eje de abscisas (Dimensión I) y alejados del resto de poblaciones. Salvo las poblaciones de Albania y Turquía, que se sitúan relativamente alejadas hacia el extremo positivo de este mismo eje, el resto de poblaciones europeas aparecen agrupadas en una nube de puntos bastante compacta.

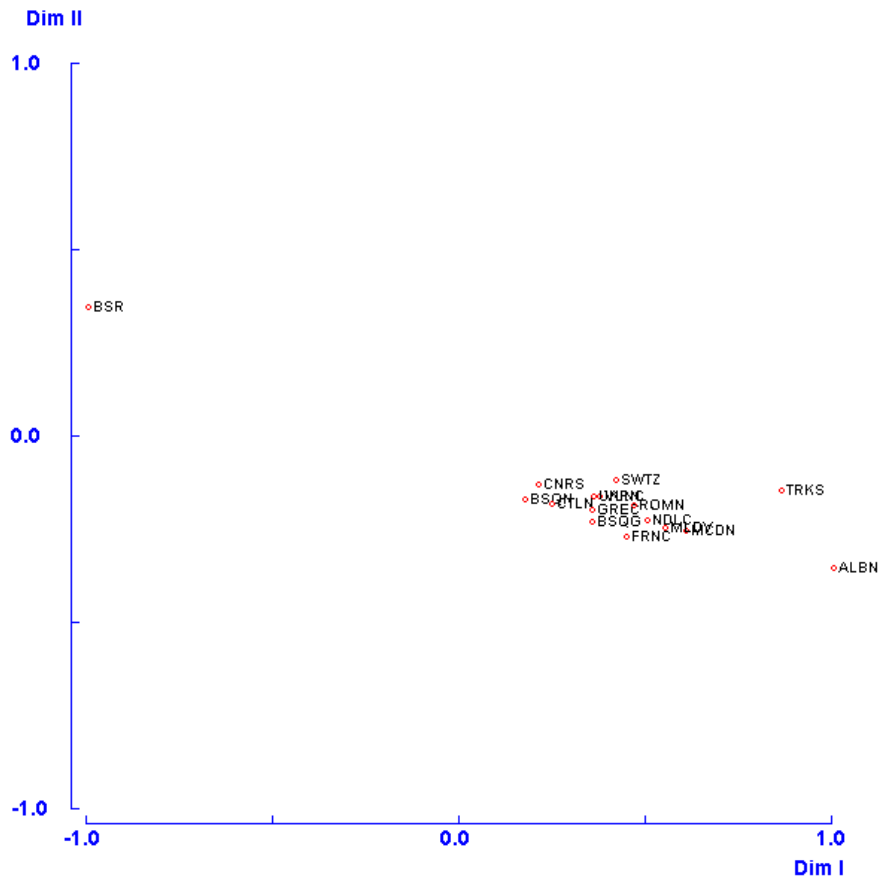


Figura 59: *Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz F_{ST} de Reynolds para 8 inserciones Alu, en poblaciones europeas y la población romaní estudiada. Stress: 0,096. Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), FRNC: Bretones (Francia), SWZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucrainianos, TRKS: Turcos.*

La Figura 60 muestra los resultados obtenidos al analizar la población romaní junto a las poblaciones asiáticas. A lo largo del eje de abscisas (Dimensión I) se observa una disposición que recuerda a la posición geográfica de las poblaciones: desde el extremo positivo, con la población Tamil situada al sur de la India, a continuación las poblaciones generales de la India, los pastunes de Afganistán y las poblaciones de Próximo Oriente, para terminar, en el extremo negativo, con la población romaní. Tan sólo la población de Pakistán aparece un tanto desubicada, más próxima a Siria que a las poblaciones de la India.

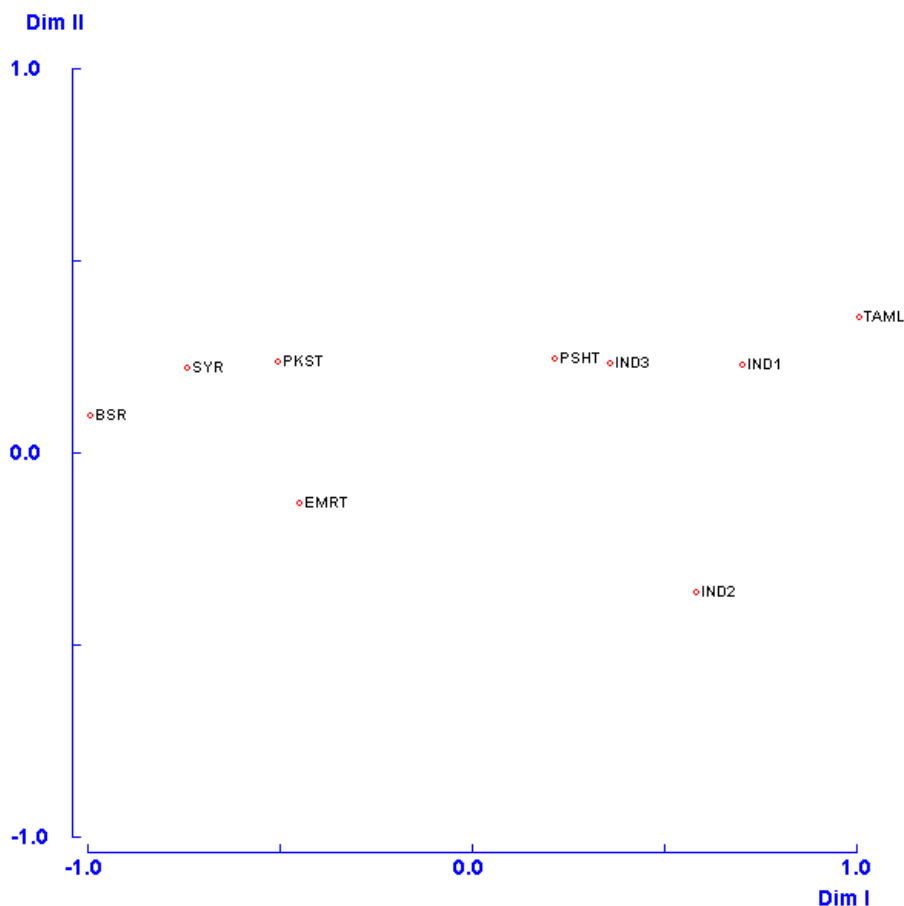


Figura 60: Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz F_{ST} de Reynolds para 8 inserciones Alu, en poblaciones asiáticas y la población romaní estudiada. Stress:0,023. Leyenda: BSR: Romaníes del País Vasco (España), EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, NSTN: Nusa, IND1: India-Cristianos, IND2: India-Hindús, IND3: India-Musulmanes.

Análisis factorial de correspondencias (AFC)

Con el fin de contrastar la robustez de los resultados obtenidos mediante los MDS, se han realizado una serie de Análisis factoriales de correspondencias (AFC). En primer lugar, se ejecutó con el conjunto de las 24 poblaciones analizadas, tanto las asiáticas como las europeas, incluyendo a la población romaní del País Vasco (Figura 61). Por un lado, en los valores negativos del eje de abscisas (eje 1) se encuentran las poblaciones asiáticas, con los valores más extremos en las poblaciones de indo-hindús, indo-cristianos y tamiles. Siguiendo este mismo eje se localizan las poblaciones de indo-musulmanes, pastunes, Emiratos Árabes y Pakistán. Por otro lado, las poblaciones europeas se encuentran en los valores positivos del eje de abscisas, con una disposición que sugiere a grandes rasgos un eje oeste/este, de acuerdo con el eje de ordenadas. Los romaníes aparecen relativamente alejados en el extremo positivo del eje 2, pero más próximos a europeos que a asiáticos. Finalmente, resulta sorprendente la posición de la

población siria, que aparece junto a algunas poblaciones del sudeste europeo, en el cuadrante inferior derecho.

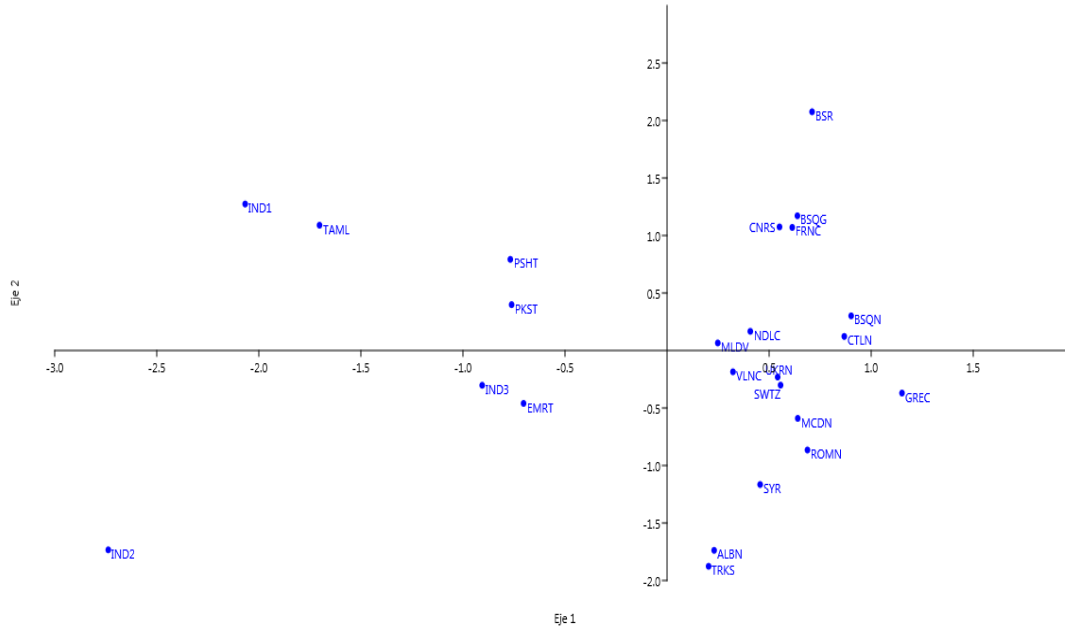


Figura 61: Análisis Factorial de Correspondencias (AFC, ejes 1 y 2) realizado a partir de las frecuencias de 8 inserciones Alu, en poblaciones euro-asiáticas y en la población romaní estudiada. Varianza explicada: 65,10 %.
Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarias (España), FRNC: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucranianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, IND1: India-Cristianos, IND2: India-Hindús, IND3: India-Musulmanes.

Si en lugar de representar los ejes 1 y 2 del AFC, se representan los ejes 1 y 3, la disposición de las poblaciones es más similar a la obtenida en los análisis MDS reflejados en la Figura 56, tal y como muestra la Figura 62. Los ejes 1 y 3 explican una proporción muy similar de la varianza (13,84% frente a 11,79 %), por lo que eventualmente pueden ser intercambiados cuando el algoritmo de análisis es diferente, como ocurre con el MDS y el AFC. En este caso, el eje 2 explica sobre todo la heterogeneidad de las poblaciones europeas (Figura 61), mientras que el 3 explica la heterogeneidad entre las poblaciones de Próximo Oriente, pakistaní y romaní frente al resto (Figura 62).

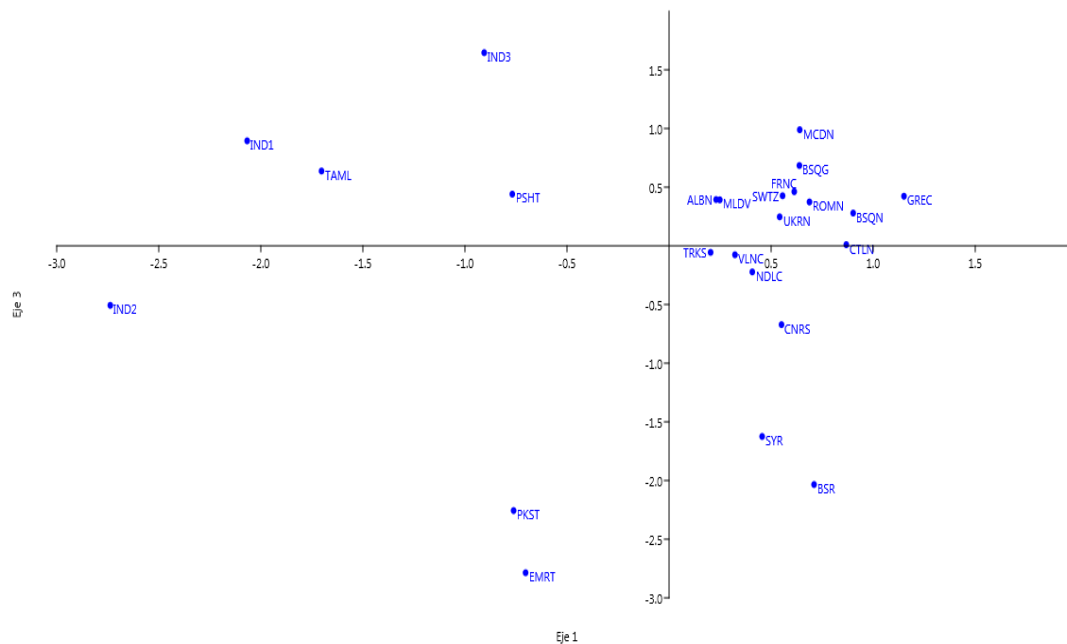
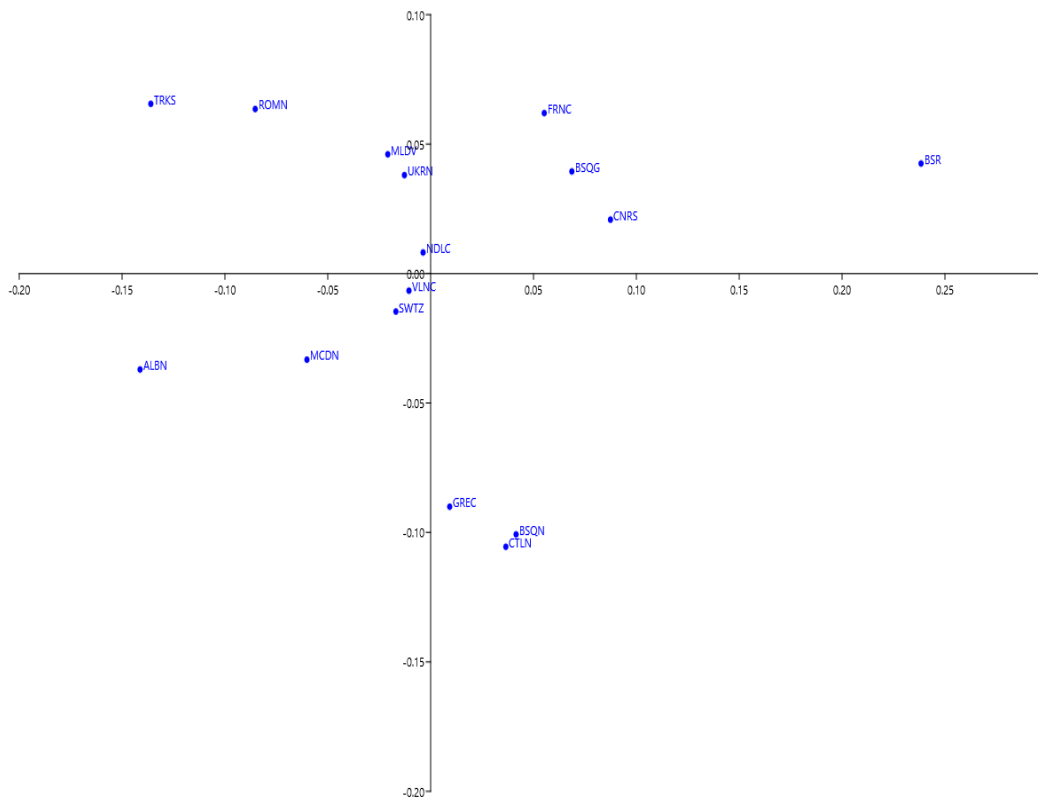


Figura 62: Análisis Factorial de Correspondencias (AFC, ejes 1 y 2) realizado a partir de las frecuencias de 8 inserciones Alu, en poblaciones euro-asiáticas y en la población romaní estudiada. Varianza explicada: 63,05 %.
Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarias (España), FRNC: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucranianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, IND1: India-Cristianos, IND2: India-Hindús, IND3: India-Musulmanes.

A continuación, se ha analizado la población romaní del País Vasco respecto a cada grupo continental por separado. El objetivo ha sido profundizar en su relación con las poblaciones de cada grupo y observar sin interferencias la distribución intragrupal de ambas poblaciones continentales. En la Figura 63 se muestra la posición de los romaníes respecto a las poblaciones europeas; como era de esperar, dicha población se emplaza en un extremo de la figura, alejada del resto de poblaciones, siendo las poblaciones francesa, canaria y guipuzcoana las más cercanas a ella. Por el contrario, las más alejadas de los romaníes del País Vasco son las muestras representativas de Albania, Turquía, Rumanía y Macedonia, que dibujan un eje que, como en la Figura 61, recuerda vagamente la disposición geográfica oeste/este de las mismas. El eje de ordenadas separa a las poblaciones griega, catalana y navarra del resto.



*Figura 63: Análisis Factorial de Correspondencias (AFC) realizado a partir de frecuencias de 8 inserciones *Alu*, en poblaciones europeas y en la población romaní estudiada. Varianza explicada: 62,24 %. Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarias (España), FRNC: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucrainianos, TRKS: Turcos.*

En la Figura 64 se muestra el resultado obtenido al comparar la población romaní del País Vasco con las poblaciones asiáticas. En los valores negativos del eje de abscisas se encuentran las poblaciones de la India y Afganistán y en los positivos se sitúan la población pakistaní y las muestras representativas de Próximo Oriente, siendo Siria la más próxima a los romaníes, localizados en el extremo positivo del eje. Hay que señalar que en los análisis realizados sobre inserciones *Alu*, romaníes y sirios mostraban un notable parentesco genético. Cabe pensar, por tanto, que la población siria, debido a su posición geográfica, tiene un patrimonio genético en cierta medida híbrido entre el Sur de Asia y Europa. Por diferentes razones la población romaní también muestra unas características híbridas, lo que explicaría una

relativa similitud entre las frecuencias alélicas de ambas poblaciones para algunos marcadores (ver Discusión).

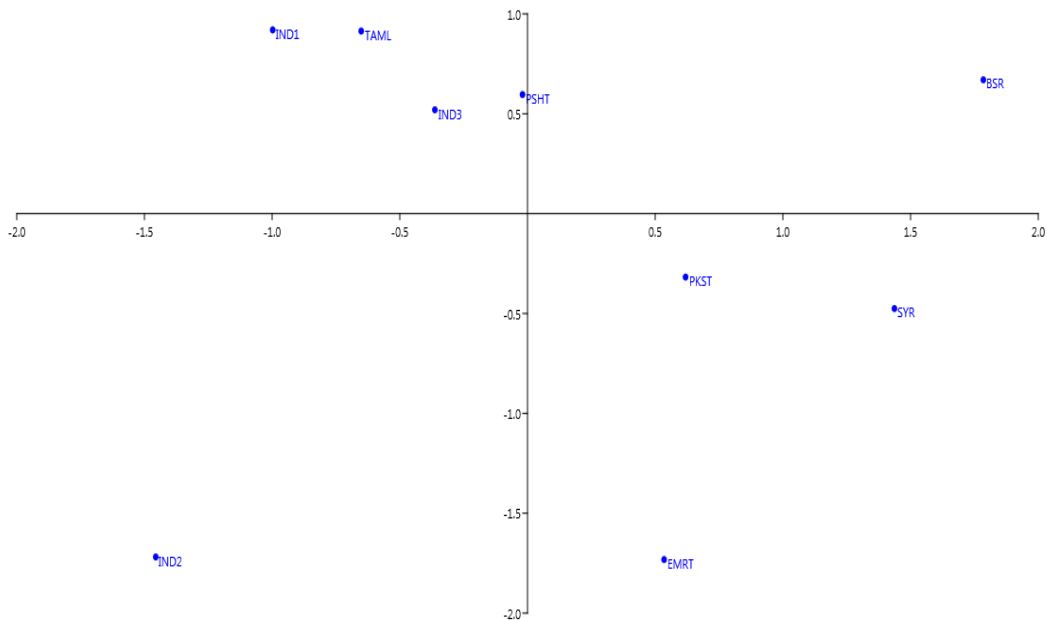


Figura 64: Análisis Factorial de Correspondencias (AFC) realizado a partir de frecuencias de 8 inserciones *Alu*, en poblaciones asiáticas y en la población romaní estudiada. Varianza explicada: 66,40 %. Leyenda: BSR Romaníes del País Vasco (España), EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, IND1: India-Cristianos, IND2: India-Hindús, IND3: India-Musulmanes.

Método del centroide

Se ha realizado, al igual que con los SNPs, un análisis del centroide con las frecuencias de las inserciones *Alu* con el objetivo de observar el flujo génico que ha experimentado la población romaní estudiada respecto al resto de poblaciones manejadas en esta memoria, así como su heterocigosidad y la varianza que nuestra población aporta al conjunto. Se han incluido las 24 poblaciones de Europa y el Sur de Asia disponibles para las 8 inserciones (Figura 65). Lo más relevante de este análisis es la posición de las poblaciones de la India y la pakistaní. Así, excepto los musulmanes de la India, el resto de poblaciones de esta región muestran valores elevados tanto de varianza como de heterocigosidad, lo cual revela un alto grado de flujo génico con otras poblaciones no incluidas en el análisis. Por su posición en el Sur de Asia, y próximas

geográficamente a otras poblaciones del centro y del Este del subcontinente Indio, es comprensible que puedan estar en contacto genético con ellas.

Los pastunes muestran un alto grado de flujo génico con las poblaciones de su entorno, en tanto que las poblaciones de Emiratos Árabes y Siria, paradójicamente, muestran un menor flujo génico del esperado. Las poblaciones europeas aportan poca heterogeneidad, con una gran variabilidad en los niveles de heterocigosidad. Los romaníes muestran unos valores medios de variabilidad y heterogeneidad y una relación entre ambas de acuerdo a lo esperado, ya que se sitúan prácticamente en la línea promedio. Esto es lo que cabría esperar de una población que ha experimentado flujo génico tanto desde poblaciones europeas como asiáticas (ver Discusión).

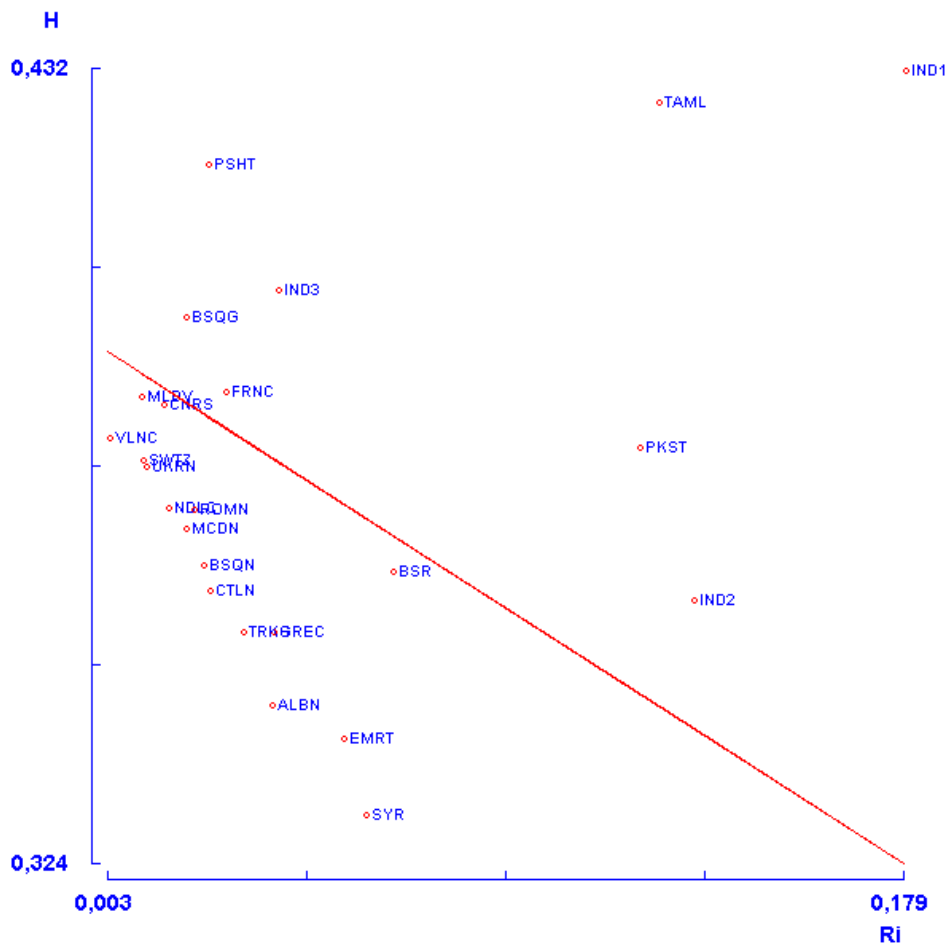


Figura 65: Método del centroide sobre 8 inserciones Alu, para un grupo de poblaciones euroasiáticas y la población romaní de estudio. Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), FRNC: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucrucianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, IND1: India-Cristianos, IND2: India-Hindús, IND3: India-Musulmanes.

No se han realizado otros análisis (clinas, deriva, mestizaje), ya que el grupo de poblaciones disponibles no se ha considerado el idóneo. En efecto, aun siendo estas inserciones *Alu* las más frecuentemente estudiadas y las más comunes, no existen datos para las poblaciones del noroeste de la India, y las muestras representativas del entorno se limitan a Afganistán, Pakistán (pastunes), el Sur de la India (tamiles) o a población general de la india de diferentes confesiones religiosas. Por ello, se ha considerado adecuado reducir el número de inserciones con el fin de ampliar el número de poblaciones representadas, particularmente del Sur de Asia.

V.2.2. Análisis de 7 inserciones *Alu*

Con el fin de mejorar la representación de las poblaciones ubicadas en la región de origen de los romaníes, se eliminó la inserción *Alu* A25. De este modo pudo incluirse un grupo de 9 poblaciones del norte de la India (Tripathi et al., 2008). Por otra parte, se han eliminado 3 de las poblaciones de la India utilizadas en el apartado anterior (indo-cristianos, indo-hindús e indo-musulmanes), por ser demasiado genéricas. Las frecuencias se encuentran recopiladas en la Tabla 4 del Anexo. En la Figura 66 se muestra el resultado del MDS realizado sobre la matriz de distancias F_{ST} de este grupo de poblaciones. En esencia, la posición de las poblaciones no es muy distinta a las del análisis realizado con las 8 inserciones *Alu*. Las nuevas poblaciones indias se sitúan dentro del área delimitada por las poblaciones pastún, tamil y pakistaní. Así, las poblaciones del Sur de Asia ocupan la mitad izquierda del gráfico, las poblaciones europeas el cuadrante superior derecho, y las poblaciones de Próximo Oriente y los romaníes (BSR) el cuadrante inferior derecho.

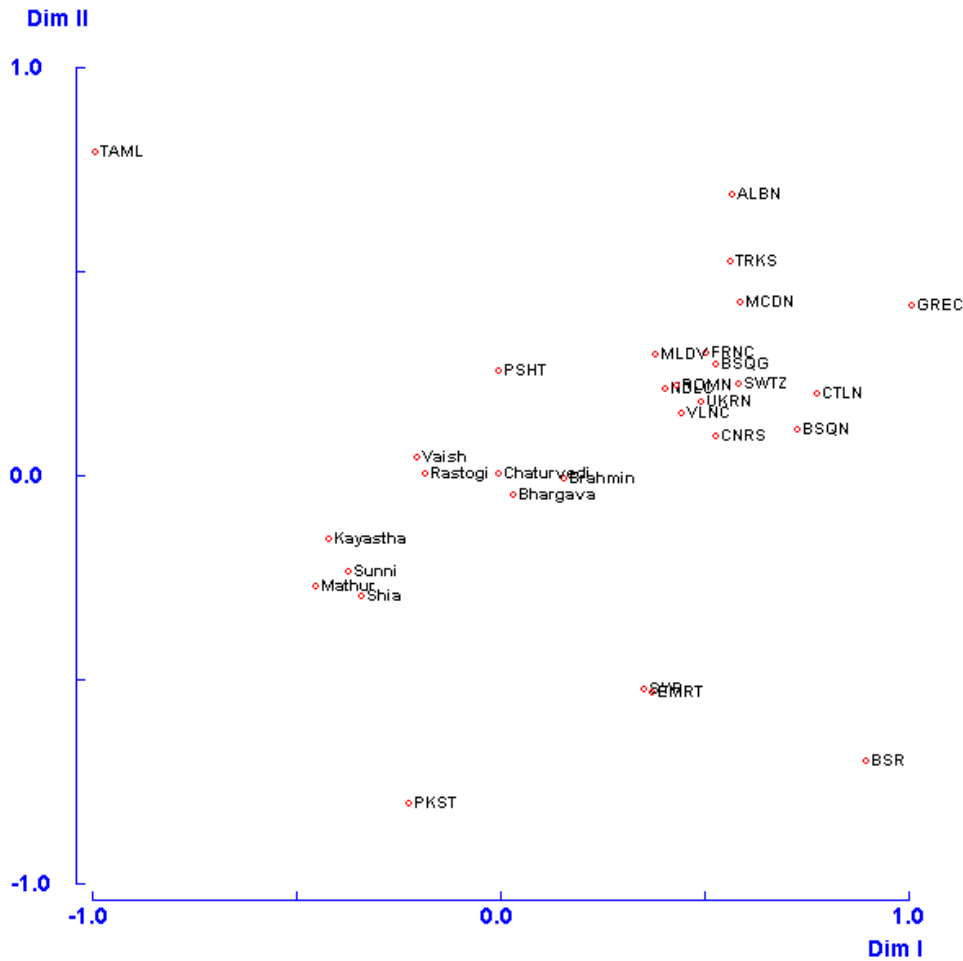


Figura 66: Análisis de Escalamiento Multidimensional (MDS) obtenido a partir de la matriz F_{ST} de Reynolds de 7 inserciones Alu, en poblaciones euroasiáticas y en la población romaní estudiada. Stress: 0,074. Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), FRNC: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucranianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, Uttar Pradesh: Kayasthas, Mathurs, Rastogies, Vaish, Brahmins, Bhargavas, Chaturvedis, Sunni: Sunitas, Shia: Chiitas.

En un AFC realizado sobre las mismas poblaciones, el resultado ha sido, a grandes rasgos, muy similar (Figura 66). Se observa, no obstante, una mayor dispersión del grupo de poblaciones del Norte de la India, y también aparece desplazada la población de Emiratos Árabes que, en lugar de encontrarse próxima a Siria, se localiza cerca de los pastunes (Figura 67). Como en el MDS, la población romaní del País Vasco se encuentra alejada de las poblaciones europeas e indias. Puesto que no aparece en una posición intermedia entre ambos grupos poblacionales, sino en una posición marginal, cabe pensar en una importante acción de la deriva sobre el patrimonio genético romaní (ver Discusión).

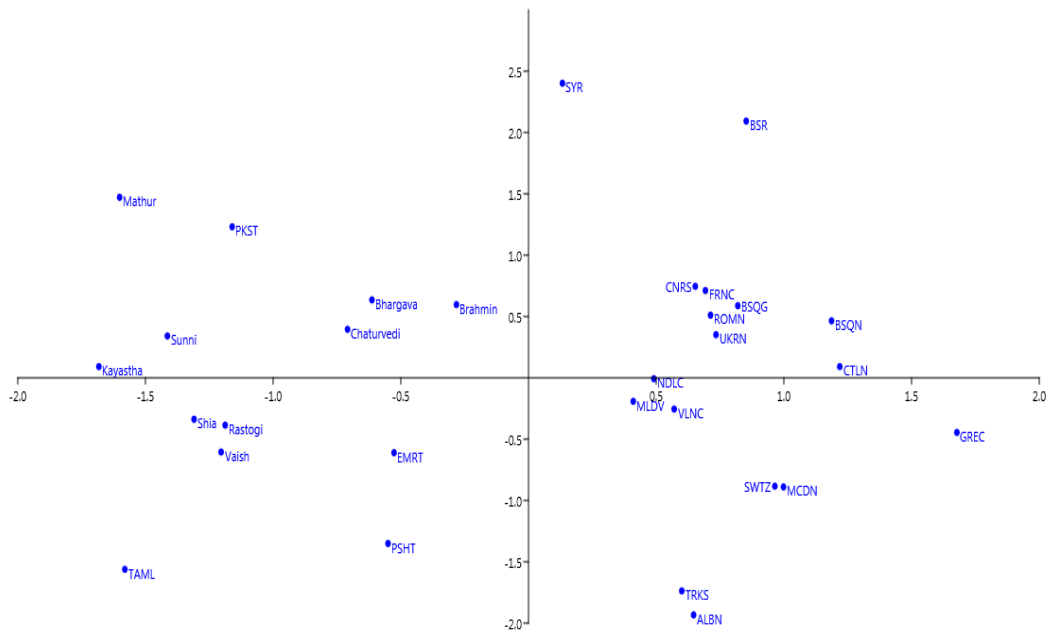


Figura 67: Análisis Factorial de Correspondencias (AFC) realizado a partir de las frecuencias de 7 inserciones Alu, en poblaciones euro-asiáticas y en la población romaní estudiada. Varianza explicada: 72,46 %. Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), FRNC: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucrainianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, Uttar Pradesh: Kayasthas, Mathurs, Rastogies, Vaish, Brahmins, Bhargavas, Chaturvedis, Sunni: Sunitas, Shia: Chiitas.

Al poder contar con un número suficientemente representativo de poblaciones de la India para el grupo de 7 inserciones *Alu*, se ha abordado el análisis de clinas. En 5 de las 7 inserciones se ha observado una clina significativa (Tabla 13, Figura 68). En dicha figura se observa que las clinas son predominantemente latitudinales con una orientación noroeste/sur-sudeste, de forma similar a lo observado con los marcadores SNP.

Tabla 13: Características de las clinas observadas en las frecuencias de 5 inserciones *Alu* en 30 poblaciones del sudoeste de Eurasia.

SNP	Ángulo	Intersección	Pendiente	r	df	p	Test de Bonferroni
D1	353	0,183011	0,207357	0,661944	28	6,78E-05	p<0,01
B65	329	0,379202	0,196751	0,593061	28	0,000553	p<0,01
ACE	121	0,349689	0,225243	0,745582	28	2,27E-06	p<0,01
APO	334	0,812445	0,173044	0,715189	28	8,94E-06	p<0,01
PV92	171	0,171113	0,257159	0,827765	28	1,67E-08	p<0,01

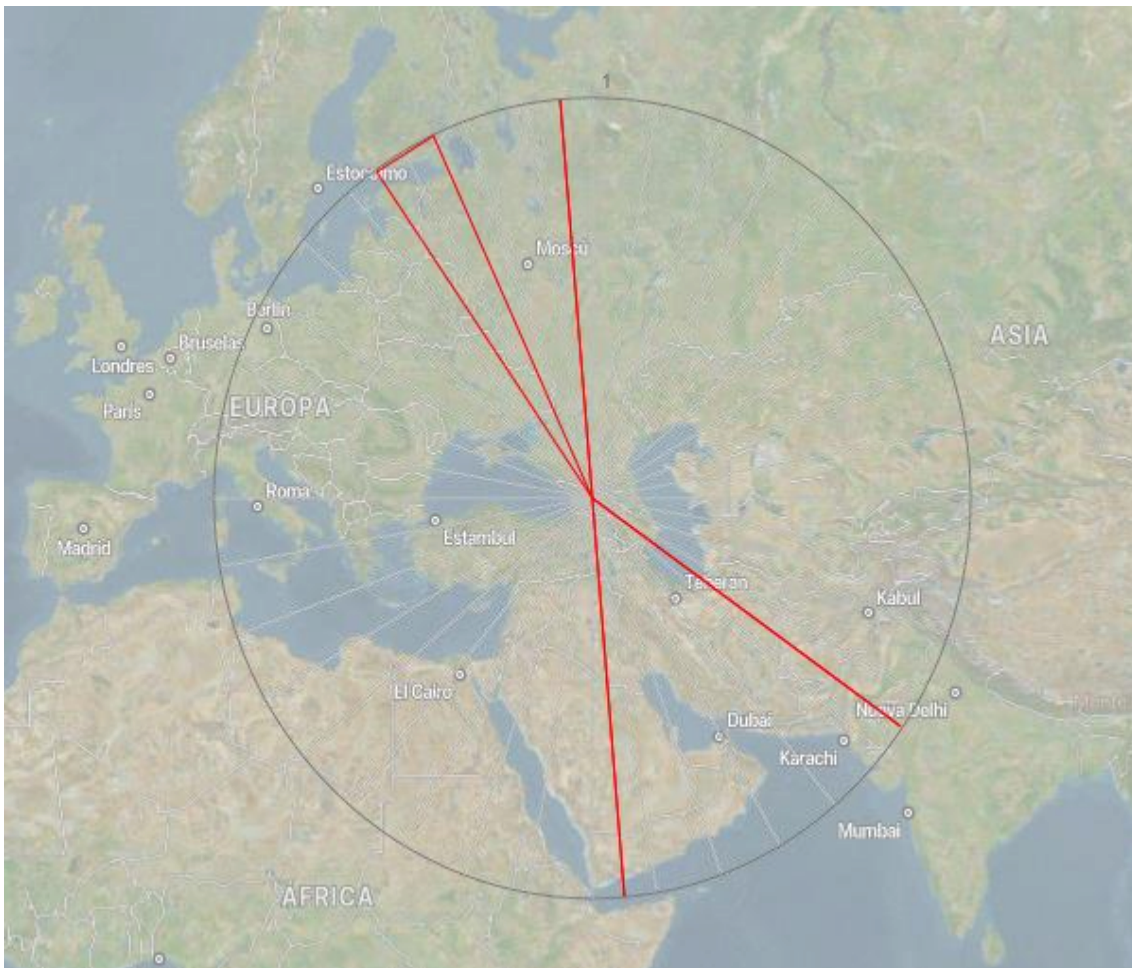


Figura 68: Orientación y frecuencia de las clinas observadas en las frecuencias de 7 inserciones *Alu* en 30 poblaciones del sudoeste de Eurasia.

Por otra parte, se observa un solapamiento en los rangos de frecuencias de asiáticos y europeos en las 5 inserciones. Los romaníes del País Vasco se sitúan en el rango de solapamiento en D1, en el rango de frecuencias asiático en B65 y APO, por debajo del rango de frecuencias de ambas regiones en ACE y en medio de ambos rangos en PV92. Así, la acción de la deriva genética sólo es perceptible –aparentemente- en una inserción de las 5 que muestran clinas significativas, concretamente en ACE (Figura 69).

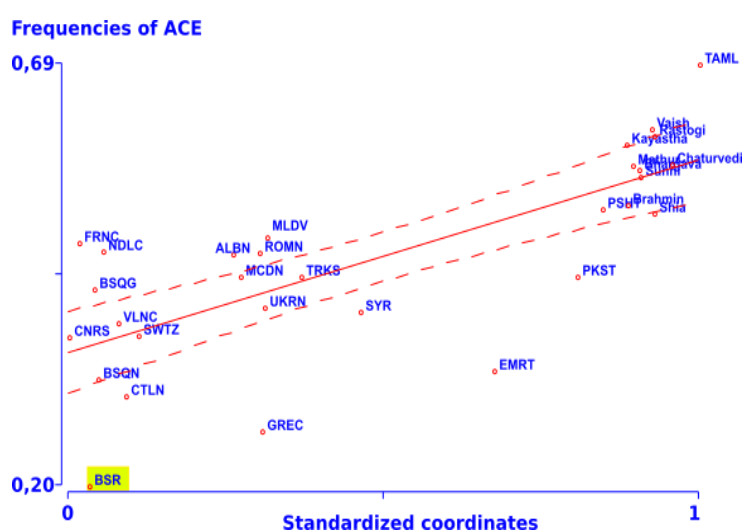


Figura 69: Rectas de regresión de ACE. Se muestra también el intervalo de confianza del 95%, así como la posición de las poblaciones consideradas. La población romaní del País Vasco (BSR) aparece señalada con un recuadro amarillo.

Se ha abordado el análisis de la acción potencial de la deriva con este grupo de 7 clinas. Como se hizo con los SNPs, se han considerado dos situaciones posibles para el cálculo del valor esperado de F_{ST} , a saber, con o sin existencia de flujo génico. En ambos casos, se han obtenido valores significativos de $-\log(P)$ en 3 de las 7 inserciones, concretamente en ACE, algo que ya se deducía del análisis de las clinas, pero también de los resultados para TPA25 y FXIIB, las dos inserciones que no mostraban clina (Tabla 14). Por tanto, cabe suponer que los valores significativos de $-\log(P)$ se deben a una acción perceptible de la deriva sobre estas 3 inserciones *Alu*. Por otra parte, aun siendo un escaso número de marcadores, la distribución de los valores de $-\log(P)$, una vez ordenados, se ajusta razonablemente bien a una curva exponencial negativa (Figura 70).

Tabla 14: Valores obtenidos del cociente de F_{ST} observado y esperado con su correspondiente P asociada y el valor de $-\log(P)$. Los valores significativos se marcan con un asterisco.

SNP	Fst obs/ Fst min	p(F)	$-\log(p)$	Fst obs/ Fst med	P(F)	$-\log(P)$
D1	0,7271	4,83E-01	0,3158	1,0004	3,68E-01	0,4345
B65	0,8093	4,45E-01	0,3515	1,0901	3,36E-01	0,4734
ACE	3,1107	4,46E-02	1,3510*	4,1375	1,60E-02	1,7969*
APO	0,8743	4,17E-01	0,3797	0,9856	3,73E-01	0,4280
TPA25	4,8102	8,15E-03	2,0891*	6,3907	1,68E-03	2,7754*
PV92	0,7904	4,54E-01	0,3433	1,0022	3,67E-01	0,4352
FXIIB	9,9357	4,84E-05	4,3150*	13,3308	1,62E-06	5,7894*

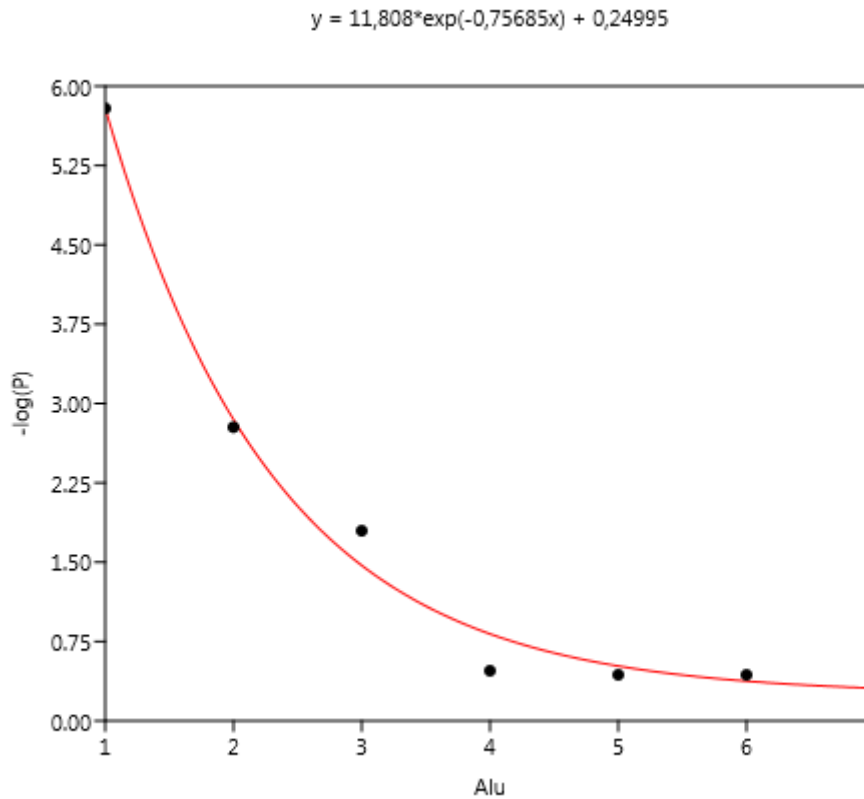


Figura 70: Regresión exponencial negativa sobre los valores ordenados de $-\log(P)$ para 7 inserciones *Alu*. Los valores representados son los del modelo con flujo génico.

En la Tabla 15 se muestran las proporciones estimadas de mestizaje obtenidas a partir de 7 inserciones *Alu*, considerando como poblaciones parentales, por un lado, el norte de la India (representada como el valor medio de las frecuencias de las muestras representativas de este territorio), y por otro la Península Ibérica (calculada también como el valor medio de las

frecuencias de las muestras españolas). Al igual que con los SNPs, se han desechado aquellas inserciones *Alu* que mostraban una perceptible acción de la deriva. Se observan de nuevo unos valores similares para la proporción de patrimonio genético original de la India y para la proporción de patrimonio genético importado de Europa. En este caso, a diferencia de los SNPs, el componente original presenta un valor (58,4%) ligeramente superior al componente importado (41,6%).

Tabla 15: Estimación de la proporción del patrimonio genético de los romaníes del País Vasco proveniente de dos poblaciones parentales, error estándar e intervalo de confianza del 95%.

Población parental	Coefficientes	Error estándar	Límite inferior	Límite superior
Norte de la India	0,584	0,109	0,116	1,052
Península Ibérica	0,416	0,095	0,006	0,825

Para finalizar con los análisis de las inserciones *Alu*, se ha realizado un dendrograma (Figura 71). Dado que es posible reconstruir el estado ancestral para todas las inserciones *Alu* y, a partir de ahí, reconstruir la población ancestral teórica (con frecuencias 0 para todas ellas), puede valorarse la posición relativa de todas las poblaciones en relación a la población ancestral que, en el caso de un dendrograma, será la que represente la raíz. Se ha realizado un análisis *Neighbour Joining* sobre la matriz de distancias F_{ST} calculada a partir de las 7 inserciones *Alu*. Con el fin de contrastar la posición de las poblaciones europeas y asiáticas en relación a la población ancestral, se han añadido una muestra africana (!Kung) (Stoneking et al., 1997) y una del sudeste asiático (Han) (Xiao et al., 2002). Las frecuencias se encuentran recopiladas en la Tabla 5 del Anexo.

Como cabría esperar, la población !Kung es la más próxima a la población ancestral y la muestra Han aparece, precisamente, en el otro extremo del gráfico; en medio se disponen todas las poblaciones europeas y del Sur de Asia. Las más próximas a los Han son, como cabía esperar, las poblaciones de la India. Las poblaciones europeas se sitúan en otra rama que se une a la del Sur de Asia, en una posición equidistante a las de Próximo Oriente (Emiratos Árabes y Siria). Como en anteriores análisis, la población de Pakistán se encuentra más cercana a las de Próximo Oriente que a las del Sur de la India. Es posible que la muestra, 43 individuos de origen pakistaní que residen y trabajan en Dubái (Stoneking et al, 1997), no sea muy representativa de la población general de Pakistán (ver Discusión). Finalmente, la población romaní del País Vasco (BSR) también aparece en este pequeño grupo, en una posición sobre la que puede haber

influido tanto la deriva experimentada en algunos marcadores como el flujo génico recibido (Ver Discusión).

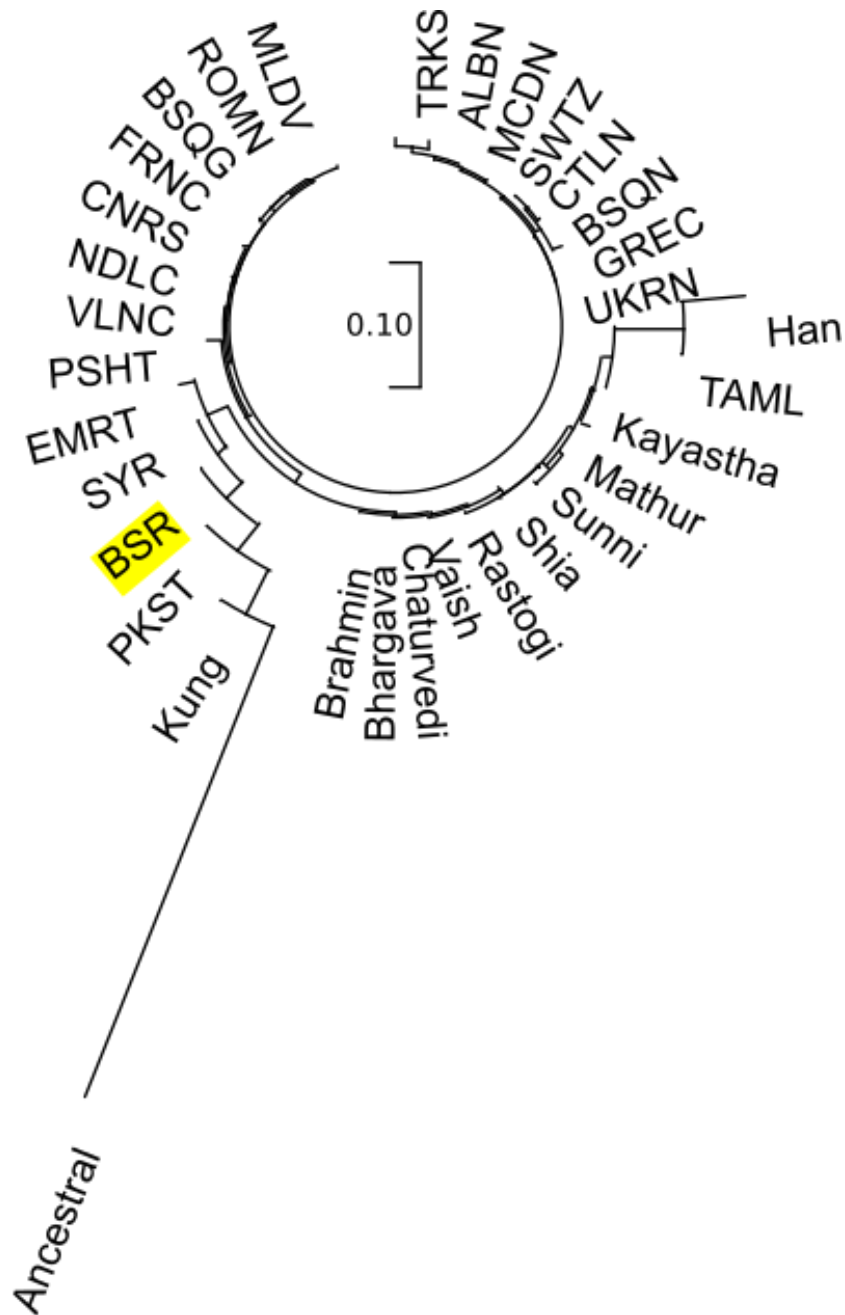


Figura 71: Dendrograma realizado a partir de las frecuencias de 7 inserciones Alu para poblaciones Ancestrales africanas, europeas y asiáticas. Leyenda: BSR: Romaníes del País Vasco (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), BRTN: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucranianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, Uttar Pradesh (Tripathi et al., 2008): Kayasthas, Mathurs, Rastogies, Vaish, Brahmins, Bhargavas, Chaturvedis, Sunni: Sunitas, Shia: Chiitas. ¡KUNG: Población ¡Kung de Angola , CHBJ: Han de China. ANCESTRAL: población ancestral hipotética.

V.3. Análisis de la población romaní para CODIS

En este apartado se presentan los resultados obtenidos a partir de las frecuencias de los 13 STRs pertenecientes al sistema CODIS. Se han incluido en los análisis 5 poblaciones romaníes residentes en Croacia, Eslovaquia, Hungría, Macedonia y Portugal, así como muestras de población general de estos mismos países. También se han considerado otras 6 poblaciones europeas, con el fin de obtener una mejor representación de la variabilidad de Europa. Aunque existen numerosas poblaciones europeas analizadas para el sistema CODIS, un número mayor de muestras sólo habría servido para aumentar la confusión en las representaciones gráficas, por lo que no se han considerado. Por último, se han incluido 9 poblaciones del Sur de Asia. Todas las frecuencias han sido tomadas de la bibliografía (ver Tabla 7 del Anexo).

Análisis molecular de la varianza (AMOVA) para los marcadores del sistema CODIS

Con la finalidad de determinar la existencia de diferencias significativas entre las poblaciones europeas y las poblaciones asiáticas (India, Pakistán e Irán) se ha realizado un análisis molecular de la varianza (AMOVA) (Tabla 16) mediante el programa Arlequín; se especificaron 1000 permutaciones. Teniendo en cuenta el valor F_{CT} , se observan diferencias significativas para 9 de los 13 STRs, lo que indica que hay una heterogeneidad genética significativa entre ambos grupos. De los distintos marcadores considerados, 4 presentan un alto nivel de significación ($P < 0,001$), 2 un nivel medio ($P < 0,01$) y el resto un nivel más bajo ($P < 0,05$). A su vez los índices F_{SC} y F_{ST} muestran variaciones significativas para todos los marcadores, revelando la existencia de heterogeneidad tanto en las poblaciones dentro de cada grupo como entre los individuos dentro de las mismas. En el caso de F_{SC} , salvo dos STRs, el resto muestra un grado de variación altamente significativa, y para el índice F_{ST} todos los STRs analizados tienen un valor muy alto de significación estadística.

Tabla 16: Índices de fijación (F_{ST} , F_{SC} y F_{CT}) generados mediante el análisis AMOVA para frecuencias de 13 STRs del CODIS, en poblaciones europeas y del Sur de Asia. (F_{CT} : varianza genética entre grupos. F_{SC} : varianza genética entre poblaciones dentro de grupos. F_{ST} : varianza genética entre individuos dentro de una población) (* $P < 0,05$; ** $P < 0,01$; *** $P < 0,001$).

STR	Índices de fijación		
	F_{ST}	F_{SC}	F_{CT}
CSFP1PO	0,01468 ***	0,01562 ***	0
TPOX	0,01509 ***	0,00673 ***	0,00842 **
D16S539	0,01666 ***	0,01053 ***	0,00620 ***
FGA	0,00724***	0,00412***	0,00314 ***
vWA	0,00559***	0,00514***	0,00045
D7S820	0,01623 ***	0,01320 ***	0,00308 *
D8S1179	0,01378***	0,00341**	0,01041***
D13S317	0,00858***	0,00927***	0
D21S11	0,00672***	0,00412**	0,00260**
TH01	0,02451***	0,01369***	0,01097***
D3S1358	0,00726***	0,00602***	0,00125*
D18S51	0,01159 ***	0,00937***	0,00225*
D5S818	0,01196***	0,01015***	0,00183

Análisis de escalamiento multidimensional (MDS)

Se han realizado diversos análisis MDS con el fin de visualizar la heterogeneidad genética de las poblaciones europeas y asiáticas, así como la posición relativa de las poblaciones romaníes. En el primer MDS se han incluido las 25 poblaciones de la base de datos que se han recopilado para el presente estudio. Los resultados se muestran en la Figura 72, donde se observa que las poblaciones europeas se disponen en el cuadrante inferior derecho. Dichas poblaciones forman una nube de puntos relativamente compacta, de la que se separan ligeramente las poblaciones polaca, vasco-francesa y sarda. Esta última aparece, de hecho, más próxima a las poblaciones asiáticas que a las europeas. Las poblaciones del Sur de Asia se encuentran en el cuadrante superior derecho, con la población iraní más cerca de las europeas, de acuerdo a su situación geográfica. En el eje de ordenadas (Dimensión II) las poblaciones romaníes se sitúan con valores intermedios entre Asia y Europa, pero desplazadas hacia los valores negativos del eje de abscisas (Dimensión I). Es notable la posición de los romaníes croatas, que muestran una notable heterogeneidad genética respecto a todo el conjunto de poblaciones europeas, asiáticas e incluso otros grupos romaníes.

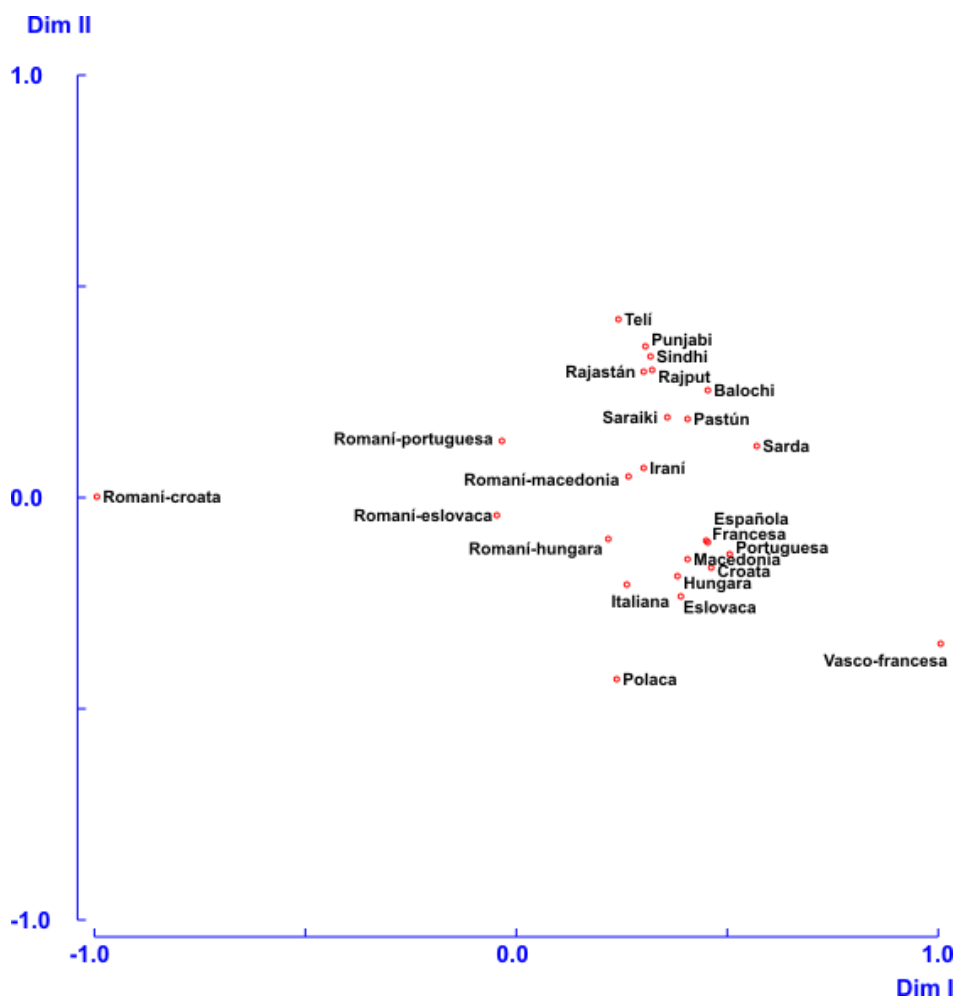


Figura 72: Análisis de Escalamiento Multidimensional a partir de la matriz F_{ST} de Reynolds sobre STRs del CODIS, para 25 poblaciones de Europa y el Sur de Asia, incluyendo 5 poblaciones romaníes. Stress: 0,065.

A continuación, con el fin de valorar la distancia genética entre las poblaciones romaníes y aquellas con las que comparten territorio en la actualidad, se realizó un análisis incluyendo sólo a las poblaciones europeas. En la Figura 73 se observa en el extremo positivo del eje de abscisas (Dimensión I) a la población vasco-francesa y en el extremo contrario a la población romaní croata. De igual manera, en el extremo inferior de la figura se localiza la población romaní macedonia alejada de la nube de puntos que conforman el resto de poblaciones. En definitiva, tanto la población romaní croata como la romaní macedonia muestran una mayor heterogeneidad con las poblaciones europeas que el resto de poblaciones romaníes incluidas en el análisis.

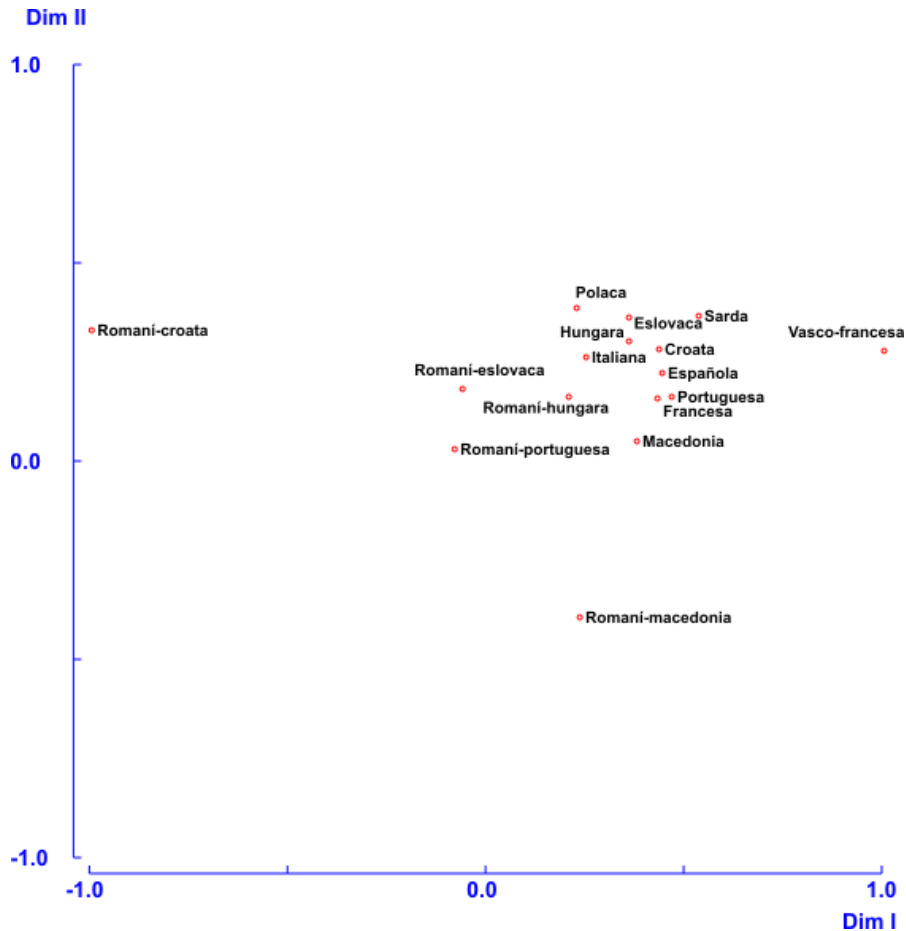


Figura 73: Análisis de Escalamiento Multidimensional a partir de la matriz F_{ST} de Reynolds sobre STRs del CODIS, para poblaciones europeas, incluyendo romaníes. Stress: 0,044.

En un análisis posterior se han incluido las poblaciones romaníes junto con las del Sur de Asia (Figura 74). Los resultados muestran que los grupos genéticamente más próximos a los romaníes europeos son los iraníes, los balochi, que de hecho se distribuyen principalmente entre Irán, Afganistán y Pakistán, y los pastunes, que viven entre Afganistán y Pakistán. Hay por tanto una coherencia geográfica en cuanto al parentesco genético de las poblaciones del Sur de Asia con las romaníes. Entre estas últimas destacan los romaníes croatas, de nuevo los más alejados del resto, y también los romaníes macedonios que, junto con los anteriores, son los que más se diferencian de las poblaciones europeas y de las asiáticas. Este resultado es sin embargo contradictorio con el análisis conjunto de todas las poblaciones, en el que los romaníes de Macedonia son de entre los grupos de esta etnia, los más próximos a las poblaciones del Sur de Asia.

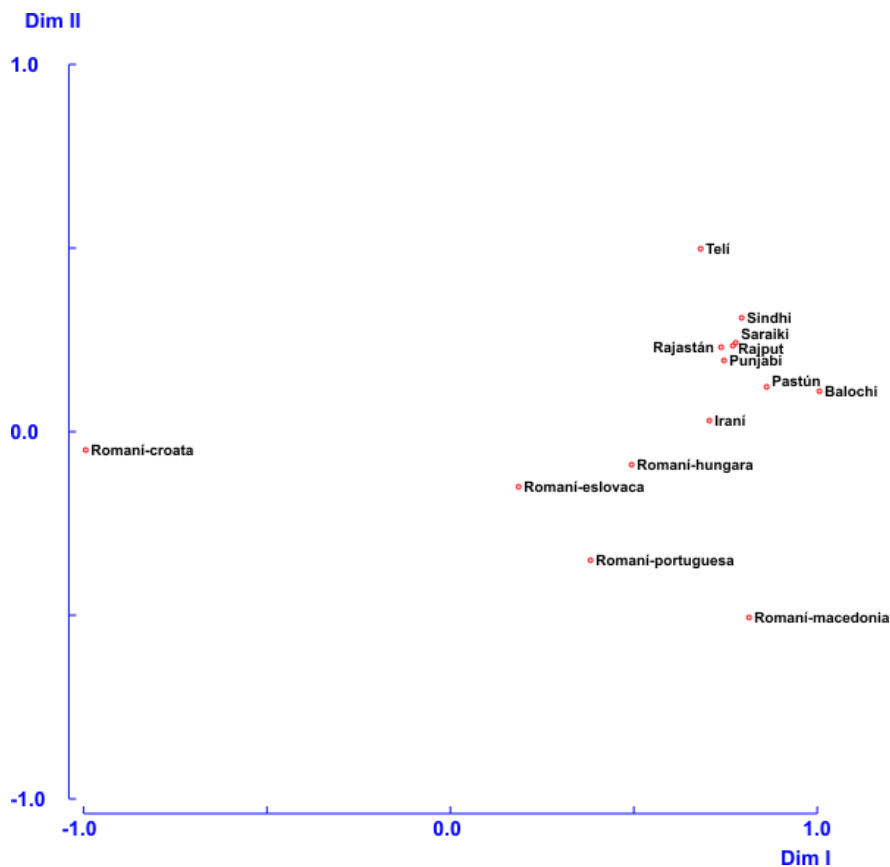


Figura 74: Análisis de Escalamiento Multidimensional a partir de la matriz F_{ST} de Reynolds sobre STRs del CODIS, para poblaciones romanas y del Sur de Asia. Stress: 0,052.

Análisis factorial de correspondencias (AFC)

Con el fin de contrastar la robustez de los resultados obtenidos mediante los MDS, se ha realizado un Análisis factorial de correspondencias (Figura 75). Como en los análisis anteriores, los dos grupos continentales muestran una gran heterogeneidad, de forma que cada uno de ellos aparece en un extremo del eje 1. Los romaníes están en el extremo positivo de este eje, al igual que las poblaciones del Sur de Asia, pero se sitúan separados de estas por el eje 2. El eje 1 mostraría la heterogeneidad Europa/Asia, en tanto que el 2 discriminaría entre los romaníes y las poblaciones asiáticas. Por otra parte, como en los resultados de los MDS, los romaníes croatas muestran un alto grado de diferenciación respecto al resto de poblaciones, en tanto que los de Macedonia son los más similares a las poblaciones del Sur de Asia.

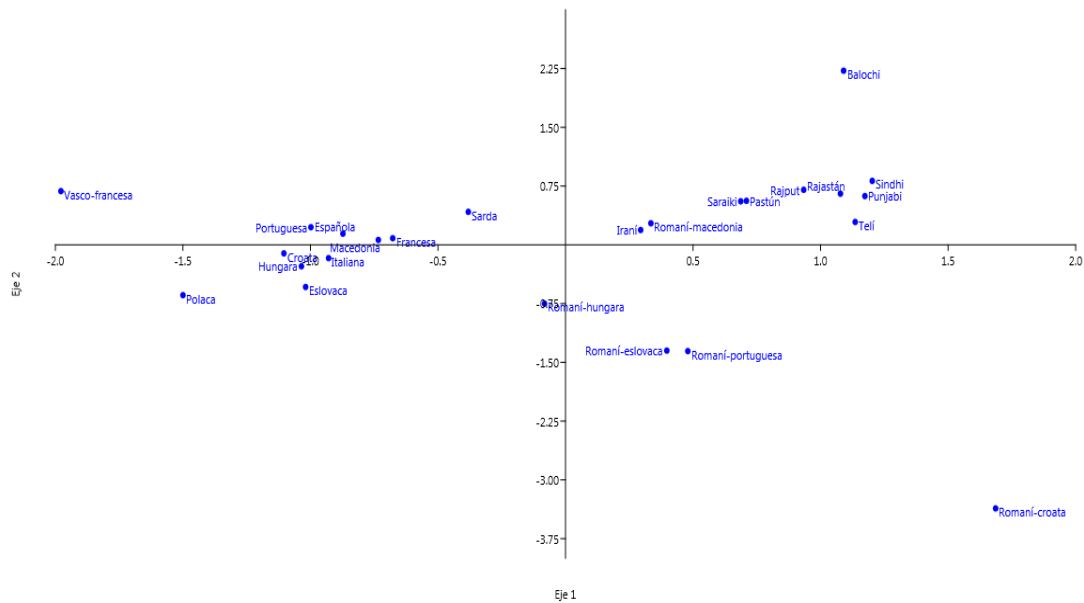


Figura 75: Análisis Factorial de Correspondencias sobre las frecuencias de los STRs del CODIS, en 23 poblaciones de Europa y el Sur de Asia. Varianza explicada: 34,72%.

Método del centroide

Con el objetivo de valorar el flujo génico experimentado por las diferentes poblaciones objeto de estudio, así como su heterocigosidad y la varianza que aporta cada una de ellas al conjunto, se realizó un análisis del centroide (Figura 76). En general, la mayor parte de las poblaciones europeas y del Sur de Asia se encuentran próximas al valor medio esperado de la relación entre varianza y heterocigosidad. Las asiáticas, no obstante, muestran una mayor influencia del flujo génico, particularmente las poblaciones pastún y Rajastán. En todo caso, debido a su posición geográfica, es algo que cabría esperar. En cuanto a los romaníes, son nuevamente los de Croacia los que muestran una posición más llamativa, reflejo probable de un notable grado de aislamiento. De modo similar, los de Portugal también aparecen claramente por debajo de la recta de regresión. Sin embargo, los romaníes macedonios, que han mostrado una posición inestable en los diferentes MDSs, no muestran, como cabía esperar, los efectos del aislamiento sino más bien de un flujo génico por encima del esperado (Ver Discusión).

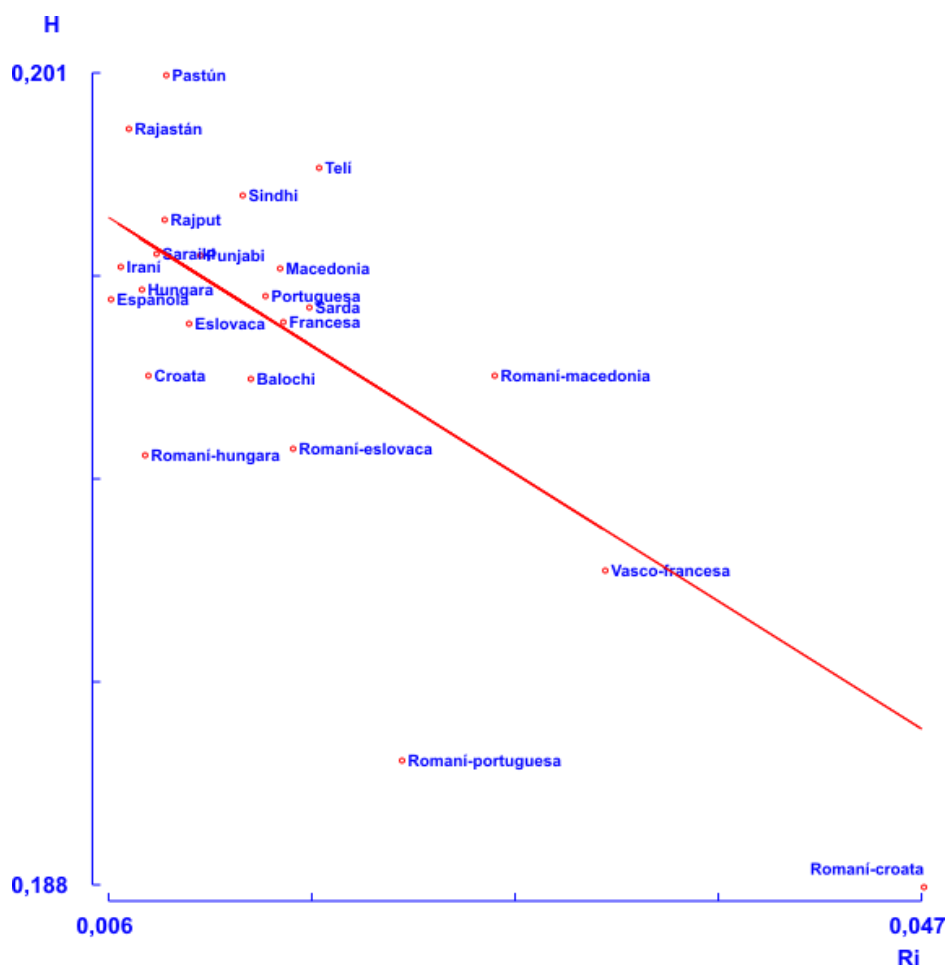


Figura 76: Método del centroide sobre STRs pertenecientes al sistema CODIS, para un grupo de poblaciones romaníes europeas, no romaníes y sur-asiáticas.

Análisis de clinas para el sistema CODIS

Se ha realizado un análisis de detección de clinas mediante GenoCline para los *loci* STR del sistema CODIS, en las poblaciones de Europa y del Sur de Asia (Tabla 17, Figura 77). Se han encontrado 43 clinas significativas, de las cuales 11 han superado la corrección de Bonferroni y 4 alcanzan una significación de $P < 0,01$. Es posible que el gran número de clinas significativas iniciales se haya producido por causa del azar, debido al elevado número de alelos analizados por el programa, de ahí la gran reducción obtenida tras la corrección de Bonferroni que ajusta la significación en función del número de replicas; en todo caso, la mayor parte de las clinas son esencialmente latitudinales (concretamente nor-noroeste sur-sudeste). Estos resultados coinciden con los observados tanto en los SNPs como en las inserciones *Alu* analizadas anteriormente. Las similitudes observadas en las 3 familias de marcadores coinciden con la tendencia latitudinal de las frecuencias europeas explicadas por Haak et al. (2015) (Ver Discusión).

Tabla 17: Características de las clinas observadas en las frecuencias de STRs del sistema CODIS, en 25 poblaciones del sudoeste de Eurasia.

Alelo	Ángulo	Intersección	Pendiente	r	df	P	Test de Bonferroni
CSFP1PO*10	12	0,2092	0,11	0,5343	23	5,93E-03	
CSFP1PO*15	21	-0,0008	0,0045	0,4114	23	4,11E-02	
TPOX*8	358	0,5534	-0,2145	0,7474	23	1,76E-05	p<0,01
TPOX*11	178	0,2564	0,1212	0,5554	23	3,95E-03	
TPOX*13	26	-0,0007	0,0038	0,4824	23	1,46E-02	
D16S539*10	23	0,0481	0,0643	0,4291	23	3,23E-02	
D16S539*12	348	0,3169	-0,1196	0,636	23	6,33E-04	
D16S539*13	218	0,1927	-0,0944	0,5982	23	1,59E-03	
FGA*19	330	0,0872	-0,0473	0,5404	23	5,29E-03	
FGA*20	325	0,1463	-0,0491	0,5504	23	4,36E-03	
FGA*21	349	0,1799	-0,047	0,3976	23	4,91E-02	
FGA*23	174	0,1314	0,0529	0,4032	23	4,56E-02	
FGA*25	172	0,0571	0,0913	0,7008	23	9,55E-05	p<0,01
FGA*26	177	0,0239	0,022	0,4014	23	4,67E-02	
D7S820*7	100	0,0097	0,029	0,5387	23	5,46E-03	
D7S820*8	180	0,1031	0,1358	0,6353	23	6,44E-04	
D7S820*9	259	0,1521	-0,0777	0,6448	23	5,03E-04	
D7S820*10	352	0,2761	-0,0765	0,4648	23	1,92E-02	
D8S1179*10	168	0,0606	0,1203	0,7514	23	1,50E-05	p<0,01
D8S1179*12	355	0,1512	-0,0919	0,6315	23	7,11E-04	
D8S1179*13	344	0,3147	-0,1471	0,8413	23	1,38E-07	p<0,01
D8S1179*15	177	0,1064	0,1029	0,6741	23	2,20E-04	p<0,05
D8S1179*16	143	0,0254	0,0501	0,6808	23	1,80E-04	p<0,05
D13S317*10	177	0,0545	0,0422	0,4662	23	1,88E-02	
D13S317*12	199	0,3343	-0,0814	0,52	23	7,72E-03	
D21S11*27	286	0,0292	-0,0215	0,4586	23	2,11E-02	
D21S11*29	222	0,2573	-0,0769	0,6499	23	4,38E-04	p<0,05
D21S11*30,2	5	0,0092	0,0465	0,6134	23	1,11E-03	
D21S11*31,2	145	0,0904	0,0362	0,4024	23	4,61E-02	
D21S11*33,2	158	0,0283	0,0513	0,7059	23	8,07E-05	p<0,01
D21S11*34,2	45	0,0033	0,0089	0,4021	23	4,63E-02	
TH01*6	159	0,2221	0,0562	0,4754	23	1,63E-02	
TH01*7	150	0,1233	0,0795	0,6097	23	1,21E-03	
TH01*9,3	347	0,2881	-0,187	0,8423	23	1,29E-07	p<0,01
D3S1358*14	352	0,0936	-0,0628	0,4345	23	3,00E-02	
D3S1358*17	165	0,1841	0,0641	0,5636	23	3,34E-03	
D3S1358*18	341	0,1892	-0,0678	0,6145	23	1,08E-03	
D18S51*11	169	0,0051	0,0287	0,7528	23	1,41E-05	p<0,01
D18S51*12	204	0,1568	-0,0772	0,5354	23	5,82E-03	
D18S51*14	175	0,1577	0,1195	0,6142	23	1,09E-03	
D18S51*18	360	0,0773	-0,0507	0,583	23	2,23E-03	
D5S818*10	158	0,0488	0,0834	0,6744	23	2,18E-04	p<0,05
D5S818*13	212	0,1981	-0,0575	0,5221	23	7,42E-03	

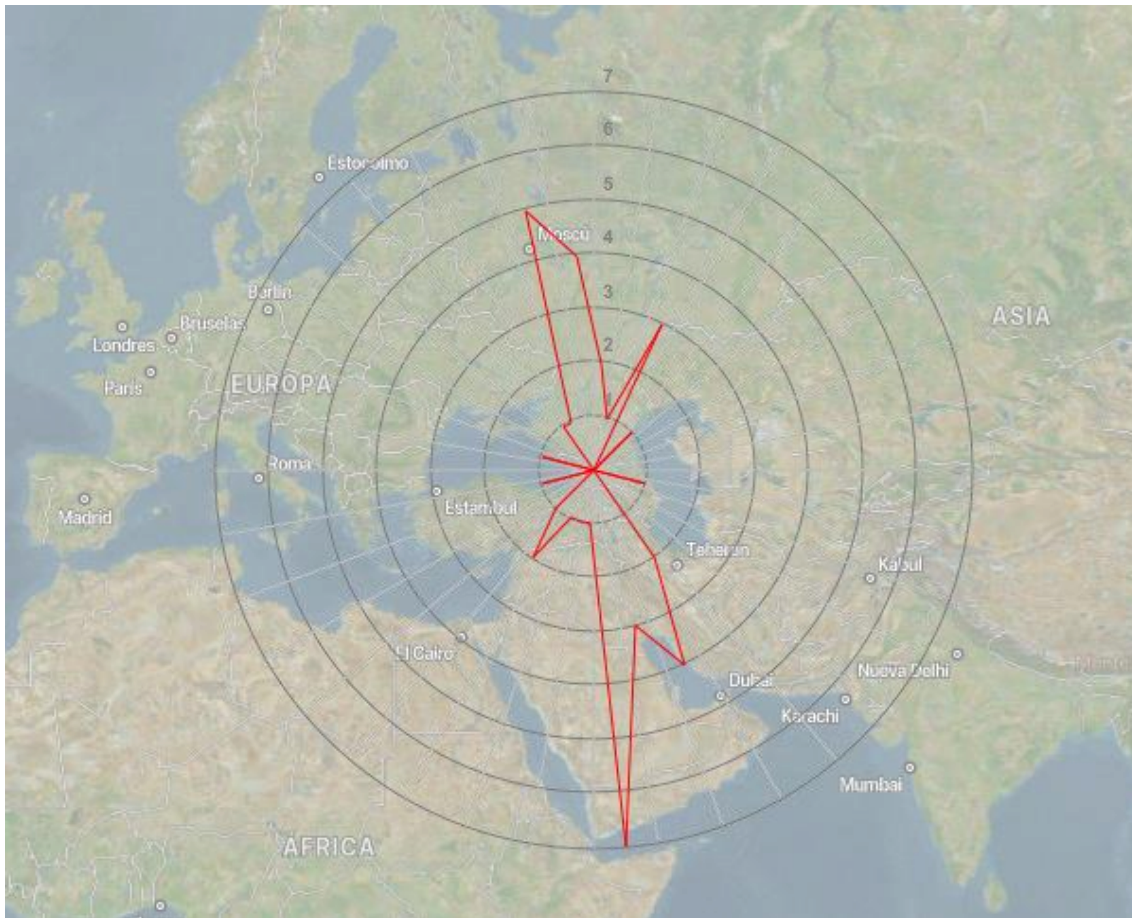


Figura 77: Orientación y frecuencia de las clinas observadas en las frecuencias de los STRs del sistema CODIS, en 25 poblaciones del sudoeste de Eurasia.

Possible acción de la deriva genética

En la Tabla 18 se muestran los valores de $-\log(P)$ para dos casos hipotéticos (con y sin migración), obtenidos a partir de las frecuencias alélicas del sistema CODIS en 5 poblaciones de romaníes europeos. Los valores significativos están marcados con un asterisco. Cuando se especifica >323 significa que el valor de $-\log(P)$ no ha podido calcularse, por ser superior a este valor y estar por encima de las capacidades del procesador. Se ha anotado un guion cuando no se ha podido calcular el cociente de F_{ST} , por existir una división por 0 en el denominador. Los resultados muestran la existencia de 53 alelos con evidencias de acción de la deriva en los romaníes de Croacia, 44 en los de Macedonia, 36 en los de Eslovaquia, 53 en los de Portugal y 31 en los de Hungría (Tabla 18).

Tabla 18: Valores obtenidos del valor de $-\log(P)$ bajo las dos premisas analizadas, con o sin migración. Los valores significativos se marcan con un asterisco.

Alelo	Romaníes Croacia		Romaníes Macedonia		Romaníes Eslovaquia		Romaníes Portugal		Romaníes Hungría	
	Sin migración	Con migración	Sin migración	Con migración	Sin migración	Con migración	Sin migración	Con migración	Sin migración	Con migración
CSF1PO										
9	6,641*	8,400*	0,246	0,421	12,889*	17,653*	0,714	1,182	2,214*	3,104*
10	0,393	0,527	108,385*	144,350*	0,701	0,946	0,377	0,432	0,331	0,444
11	>323*	>323*	0,334	0,435	0,678	0,886	18,373*	24,552*	2,209*	2,909*
12	>323*	>323*	0,427	0,524	27,607*	36,910*	12,258*	16,177*	178,033*	237,479*
13	1,207	1,655*	0,684	0,727	0,309	0,434	0,48	0,786	0,316	0,449
14	1,148	1,284	1,094	1,707*	0,743	0,816	2,644*	3,783*	1,963*	2,343*
15	0,434	0,868	0	0	0,218	0,435	0,434	0,868	0	0
TPOX										
7	0,866	0,867	0,866	0,867	1,788*	2,115*	4,849*	6,065*	0,866	0,867
8	0,347	0,466	0,913	1,227	0,326	0,436	0,34	0,457	0,355	0,475
9	4,311*	5,768*	5,765*	8,098*	0,436	0,58	1,476*	2,041*	0,445	0,608
10	2,487*	3,049*	1,115	1,438*	7,556*	10,033*	3,691*	4,805*	>323*	>323*
11	0,335	0,439	0,975	1,275	0,368	0,468	0,468	0,607	0,343	0,433
12	33,669*	45,730*	9,877*	13,295*	2,580*	3,231*	23,731*	31,162*	4,263*	5,506*
13	0,867	0,868	0,867	0,868	0,565	0,565	1,043	1,139	0,867	0,868
D16S539										
8	0,871	0,914	1,154	1,333*	0,679	0,713	0,698	0,779	0,932	1,034
9	5,006*	6,371*	>323*	>323*	0,563	0,731	50,687*	67,744*	8,786*	11,487*
10	0,836	0,991	0,388	0,469	1,021	1,241	0,449	0,542	0,58	0,693
11	20,771*	27,840*	130,765*	174,786*	0,359	0,481	47,617*	63,354*	10,745*	14,261*
12	0,334	0,465	1,084	1,496*	0,311	0,435	0,641	0,881	0,314	0,434
13	0,315	0,451	0,699	0,981	0,615	0,876	0,644	0,935	0,301	0,438
14	1,675*	2,399*	2,325*	3,216*	1,861*	2,212*	12,236*	16,830*	0,984	1,428*
FGA										
18	1,533*	3,063*	0	0	0,236	0,472	0,434	0,867	0,304	0,607
19	2,774*	3,850*	0,511	0,743	0,991	1,394*	0,319	0,441	1,436*	2,034*
20	0,826	1,153	0,75	1,047	4,083*	5,601*	2,385*	3,344*	0,307	0,438
21	0,54	0,739	0,373	0,514	0,325	0,451	4,759*	6,503*	0,725	1,001
22	0,321	0,471	0,617	0,876	0,326	0,468	0,379	0,543	0,299	0,435
23	3,400*	4,397*	7,998*	10,532*	0,41	0,501	1,456*	1,881*	0,398	0,496
24	0,591	0,73	0,355	0,443	0,486	0,618	0,404	0,501	0,464	0,57
25	9,806*	12,774*	0,961	1,207	>323*	>323*	14,674*	19,151*	0,368	0,458
26	1,323*	1,604*	46,485*	62,490*	0,615	0,786	10,554*	13,826*	1,325*	1,781*
VWA										
14	9,288*	12,601*	0,64	0,822	1,575*	2,202*	>323*	>323*	7,762*	10,407*
15	0,61	0,84	1,132	1,635*	0,326	0,484	0,334	0,501	0,298	0,442
16	1,679*	2,151*	0,486	0,633	2,306*	2,992*	7,740*	10,253*	0,537	0,7
17	9,855*	13,251*	3,378*	4,453*	1,945*	2,561*	3,359*	4,426*	2,205*	2,950*

Resultados

18	0,325	0,443	1,573*	2,161*	3,335*	4,482*	229,460*	304,845*	0,359	0,473
19	7,955*	10,184*	0,572	0,752	0,465	0,582	0,408	0,52	3,592*	4,572*
D7S820										
7	>323*	>323*	1,903*	2,464*	4,818*	6,138*	3,929*	4,963*	0,414	0,481
8	1,312*	1,647*	0,618	0,773	1,04	1,322*	0,341	0,434	0,569	0,716
9	0,564	0,823	0,323	0,464	1,374*	1,924*	1,156	1,591*	0,513	0,712
10	0,624	0,867	0,319	0,441	0,354	0,486	1,868*	2,570*	0,546	0,758
11	17,940*	23,021*	3,593*	4,822*	>323*	>323*	16,963*	22,677*	15,636*	20,988*
12	37,137*	48,754*	0,583	0,76	16,132*	21,402*	>323*	>323*	1,039	1,368*
13	1,695*	2,053*	0,684	0,727	1,214	1,558*	6,069*	7,719*	1,741	2,147*
14	0,867	0,868	-	-	0,468	0,601	-	-	24,128*	31,026*
D8S1179										
8	28,593*	37,183*	0,999	1,414*	0,7	1,006	12,212*	15,665*	12,212*	15,665*
9	0,434	0,868	0,419	0,837	0,245	0,489	0,434	0,867	0,283	0,564
10	2,168*	2,646*	0,933	1,129	0,545	0,654	2,057*	2,554*	0,378	0,462
11	3,140*	3,955*	1,251	1,633*	3,025*	3,854*	4,204*	5,549*	0,447	0,575
12	0,963	1,353*	3,170*	4,091*	0,624	0,877	1,520*	2,058*	0,401	0,572
13	0,296	0,432	0,342	0,487	0,351	0,509	0,556	0,8	0,413	0,594
14	1,763*	2,399*	0,476	0,659	0,369	0,502	0,325	0,448	0,381	0,528
15	0,443	0,539	0,344	0,445	0,428	0,527	0,408	0,504	0,477	0,568
16	0,818	0,955	0,363	0,449	0,956	1,116	0,753	0,89	0,379	0,432
17	0,867	0,868	-	-	13,098*	16,843*	3,347*	4,120*	0,368	0,43
D13S317										
8	0,702	0,86	0,347	0,443	0,403	0,518	0,423	0,449	0,345	0,436
9	187,821*	248,256*	15,792*	21,228*	>323*	>323*	0,398	0,462	0,929	1,251
10	1,196	1,522*	1,198	1,459*	0,48	0,626	3,678*	4,753*	0,474	0,594
11	1,187	1,663*	1,078	1,486*	0,844	1,161	0,448	0,622	0,489	0,677
12	0,371	0,486	9,729*	13,049*	1,1	1,446*	0,451	0,619	0,828	1,096
13	0,369	0,507	0,825	1,111	0,401	0,557	1,17	1,761*	0,495	0,681
14	0,739	1,1	44,936*	60,312*	0,988	1,389*	0,479	0,562	0,751	1,081
D21S11										
27	0,664	1,134	0,872	1,306*	0,469	0,805	0,677	1,152	0,254	0,461
28	5,784*	7,914*	>323*	>323*	0,627	0,87	24,167*	32,541*	0,402	0,561
29	4,199*	5,605*	0,338	0,466	3,364*	4,538*	6,065*	8,211*	7,450*	9,884*
30	0,959	1,285	2,445*	3,206*	0,333	0,436	3,120*	4,231*	0,535	0,715
30,2	0,371	0,599	0,522	0,838	0,253	0,426	0,513	0,801	0,308	0,502
31	241,079*	>323*	0,513	0,659	1,533*	2,092*	70,446*	94,682*	11,293*	15,296*
31,2	69,003*	92,233*	109,457*	147,211*	0,61	0,759	>323*	>323*	0,908	1,182
32	0,434	0,867	0,434	0,868	0,203	0,405	0,203	0,406	0,245	0,489
32,2	1,874*	2,322*	2,384*	3,137*	0,379	0,482	1,470*	1,822*	0,579	0,701
33	0,434	0,868	0	0	0,218	0,435	0,434	0,868	0	0
33,2	0,344	0,435	0,929	1,152	0,366	0,428	0,498	0,608	1,652*	2,088*
34	0	0	0	0	0	0	0	0	0	0
34,2	0,731	0,936	3,336*	4,064*	16,039*	22,115*	2,712*	3,304*	2,574*	3,296*
35	0	0	0	0	0	0	0,434	0,868	0	0
TH01										
6	0,765	0,984	0,999	1,305*	0,345	0,438	0,482	0,615	0,617	0,791

Procesos microevolutivos en la población romaní residente en el País Vasco

7	1,064	1,353*	0,743	0,955	0,823	1,064	42,889*	56,886*	1,956*	2,568*
8	9,278*	12,864*	0,91	1,353*	>323*	>323*	1,406*	2,004*	5,123*	7,141*
9	0,774	1,001	0,389	0,493	0,448	0,588	0,636	0,821	0,551	0,715
9,3	0,394	0,579	0,301	0,433	0,303	0,451	0,439	0,64	0,312	0,457
D3S1358										
14	1,849*	2,708*	5,343*	7,359*	0,403	0,613	0,887	1,288	0,601	0,848
15	55,949*	74,335*	1,036	1,371*	0,374	0,482	10,009*	13,293*	0,331	0,437
16	2,106*	2,766*	0,559	0,748	0,49	0,635	0,401	0,526	0,888	1,161
17	0,333	0,435	1,236	1,630*	0,365	0,475	1,038	1,367*	-	-
18	1,318*	1,836*	4,898*	6,659*	0,415	0,599	0,542	0,75	0,662	0,923
19	2,390*	2,892*	1,988*	2,553*	0,41	0,475	3,006*	3,686*	140,414*	188,780*
D18S51										
11	0,533	0,645	1,241	1,786*	3,006*	3,686*	1,963*	2,343*	12,564*	16,294*
12	0,408	0,624	0,403	0,641	0,328	0,488	0,293	0,468	0,352	0,546
13	22,782*	30,218*	19,232*	25,589*	2,632*	3,387*	43,581*	57,680*	31,143*	42,053*
14	0,918	1,16	0,358	0,433	0,484	0,601	0,352	0,433	0,387	0,485
15	3,392*	4,398*	2,204*	2,975*	4,307*	5,761*	15,796*	20,821*	0,339	0,434
16	>323*	>323*	4,075*	5,500*	0,462	0,631	21,281*	28,576*	0,783	1,002
17	>323*	>323*	1,242	1,608*	4,145*	5,718*	7,408*	9,986*	0,453	0,615
18	0,25	0,452	0,238	0,422	0,234	0,423	0,237	0,418	0,236	0,424
19	0,357	0,528	1,208	1,714*	1,531*	2,193*	0,344	0,504	0,294	0,444
20	0,947	1,511*	0,725	1,105	0,49	0,79	0,278	0,436	0,254	0,431
D5S818										
9	0,709	1,198	0,255	0,443	0,243	0,419	0,618	1,004	0,411	0,691
10	0,43	0,548	0,471	0,547	1,264	1,614*	0,365	0,445	0,351	0,442
11	15,139*	20,063*	0,577	0,749	30,518*	40,815*	298,450*	>323*	6,448*	8,559*
12	188,971*	251,532*	0,693	0,931	7,203*	9,533*	13,740*	18,350*	>323*	>323*
13	3,734*	5,083*	0,341	0,474	0,809	1,085	15,706*	21,203*	2,117*	2,847*
14	58,058*	78,775*	0,42	0,569	0,472	0,604	14,975*	20,319*	0,289	0,441

Con el fin de valorar la posible existencia de un sesgo sistemático en las estimaciones de acción de la deriva entre las diferentes poblaciones, los alelos se han clasificado en dos categorías, codificando como 1 aquellos que tenían un valor significativo para la prueba de $-\log(P)$ y como 2 aquellos que en los que no había significación. Se realizó una correlación de Spearman entre las 5 poblaciones, cuyos resultados se muestran en la Tabla 19. Sólo en 3 casos la correlación resultó significativa. Este resultado se considera aceptable, ya que no se observa una repetición sistemática de alelos con evidencias de deriva en todas las poblaciones y se considera que, por ser poblaciones con un origen común, una parte de su patrimonio genético (y por tanto de su historia microevolutiva) debería también ser común. Hay que hacer notar que los 3 coeficientes de correlación implican a la población portuguesa, que es precisamente la más distante de todo el resto de romaníes europeos.

Tabla 19: Matriz de coeficientes de correlación de Spearman entre los alelos de los STRs del sistema CODIS que han mostrado (o no) evidencias de acción de la deriva genética, en 5 poblaciones romaníes europeas.

	RomCro	RomMac	RomEsl	RomPor	RomHun
RomCro	-				
RomMac	0,1770	-			
RomEsl	0,2648	0,0811	-		
RomPor	0,4476*	0,2557	0,3462*	-	
RomHun	0,3199	0,1248	0,2724	0,3616*	-

La distribución de los valores de $-\log(P)$ entre las 5 poblaciones es desigual, pero su comparación es difícil, ya que hay varios alelos en los que no ha podido calcularse el valor exacto por haberse alcanzado el límite de decimales del procesador. Son, como ya se ha señalado anteriormente, los casos anotados como >323 . En su lugar se han comparado las distribuciones de los valores de los cocientes entre F_{ST} observado y F_{ST} esperado en el supuesto de que haya existido migración (Figura 78). La población romaní de Croacia tiene, junto con la de Portugal, el mayor número de alelos con valores significativos. Además, como se aprecia en la Figura 78, en general la curva del cociente F_{ST} de ambas está por encima del resto, en particular la de Croacia que tiene valores más altos. Este hecho concuerda con la posición de dicha población en los análisis MDS y AFC (Figuras 72 y 75), donde aparecía como la más heterogénea. Además, el análisis del centroide revelaba un alto grado de aislamiento de esta población (Figura 76), por lo que previsiblemente es la población romaní con una mayor incidencia de la deriva (Ver Discusión).

La población romaní de Portugal presenta los valores más altos dentro del rango medio, pero sus valores máximos no son de los más elevados. En todo caso, ha experimentado también un notable efecto de la deriva. La población romaní de Eslovaquia muestra valores bajos en general, aunque el grupo de alelos con valores más altos se encuentra entre los máximos, junto a los croatas, como ya se ha señalado en el párrafo anterior. Puede decirse por tanto que la deriva ha influido de forma moderada en su patrimonio genético y de manera distinta a la población portuguesa. Por la distribución de los valores, podría predecirse algún proceso intenso de cuello de botella, aunque anterior en el tiempo al experimentado por la población portuguesa, en la que este proceso sería más reciente. El aislamiento de la población eslovaca, por el contrario, habría sido más sostenido en el tiempo.

Finalmente, las poblaciones romaníes de Hungría y Macedonia presentan los valores más bajos para todo el rango de alelos, aunque el número de alelos afectados ha sido mayor. En conjunto puede inferirse que la deriva ha sido más intensa en los romaníes croatas y portugueses, aunque con características diferentes en cada caso, con una intensidad media en

los macedonios, mientras que la menor cantidad de deriva se ha producido en los eslovacos y húngaros. Estos valores permiten entender la posición de las poblaciones en los diferentes MDSs y AFCs (Figuras 72 y 75), particularmente la de las poblaciones croata y portuguesa (Ver Discusión).

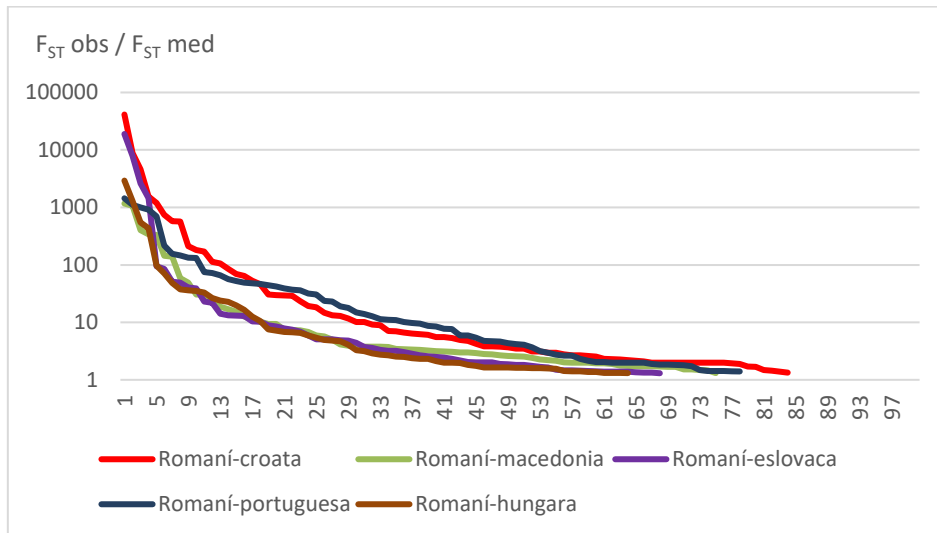


Figura 78: Distribución de valores ordenados del cociente de F_{ST} con migración en 5 poblaciones de romaníes. El eje de ordenadas se muestra en escala logarítmica. Se representan sólo los valores significativos.

Estimación del grado de mestizaje

Finalmente, se ha estimado el grado de mestizaje para las cinco poblaciones romaníes. Para ello, se ha considerado como la población parental de Asia la resultante de calcular la media de las frecuencias alélicas de las poblaciones del noroeste de la India, es decir, Punjab, Rajastán, sindhi y saraiki. En cada caso la población parental europea ha sido aquella con la que cada muestra romaní comparte actualmente su residencia. Como en los apartados anteriores, para mejorar la eficacia de los resultados, se han eliminado de los cálculos aquellos alelos con valores significativos de $-\log(P)$, es decir, que hayan mostrado claros indicios de acción de la deriva. Los resultados se muestran en la Tabla 20. La población romaní con un mayor componente original indio es la de Portugal, con un 63,5%, seguida de las de Croacia (55,4%) y Eslovaquia (50,3%) y, finalmente, con un componente menor, se encuentran Hungría (40,3%) y Macedonia (29,9%). En líneas generales, aquellas poblaciones romaníes con un mayor efecto de la deriva (según los valores observados de $-\log(P)$), muestran un componente indio mayoritario (Portugal, Eslovaquia y Croacia), en tanto que Hungría, con un menor efecto de la deriva, tiene un mayor grado de mestizaje con poblaciones europeas. No obstante, Macedonia, que ha mostrado una notable incidencia de la deriva, posee un componente indio minoritario. En cuanto a los

romaníes del País Vasco, que han mostrado unos porcentajes de mestizaje con poblaciones europeas del 52,7% a partir de los análisis de SNPs (Tabla 10), y del 41,6% a partir de las inserciones *Alu* (Tabla 15), parecen ser más similares al grupo de romaníes de Portugal, Croacia y Eslovaquia, que a los de Hungría y Macedonia (ver Discusión).

Tabla 20: Estimación del grado de mestizaje (\pm error estándar) e intervalo de confianza del 95% para las poblaciones romaníes de estudio frente a su población ancestral y su respectiva población huésped europea.

Población romaní	Componentes	Coefficientes	Error estándar	Límite inferior	Límite superior
Croacia	NW India	0,554	0,131	0,291	0,817
	Croacia	0,446	0,116	0,212	0,680
Portugal	NW India	0,635	0,094	0,446	0,825
	Portugal	0,365	0,090	0,183	0,546
Eslovaquia	NW India	0,503	0,064	0,375	0,630
	Eslovaquia	0,497	0,060	0,377	0,617
Macedonia	NW India	0,299	0,079	0,140	0,458
	Macedonia	0,701	0,079	0,541	0,860
Hungría	NW India	0,403	0,061	0,282	0,525
	Hungría	0,597	0,058	0,482	0,712

La Figura 79 muestra la comparación entre el grado de mestizaje estimado frente al porcentaje de la población romaní en el país de residencia correspondiente. Ambos valores se recogen en la Tabla 6 del Anexo. Las dos variables han mostrado una correlación significativa, como se muestra en la Tabla 21.

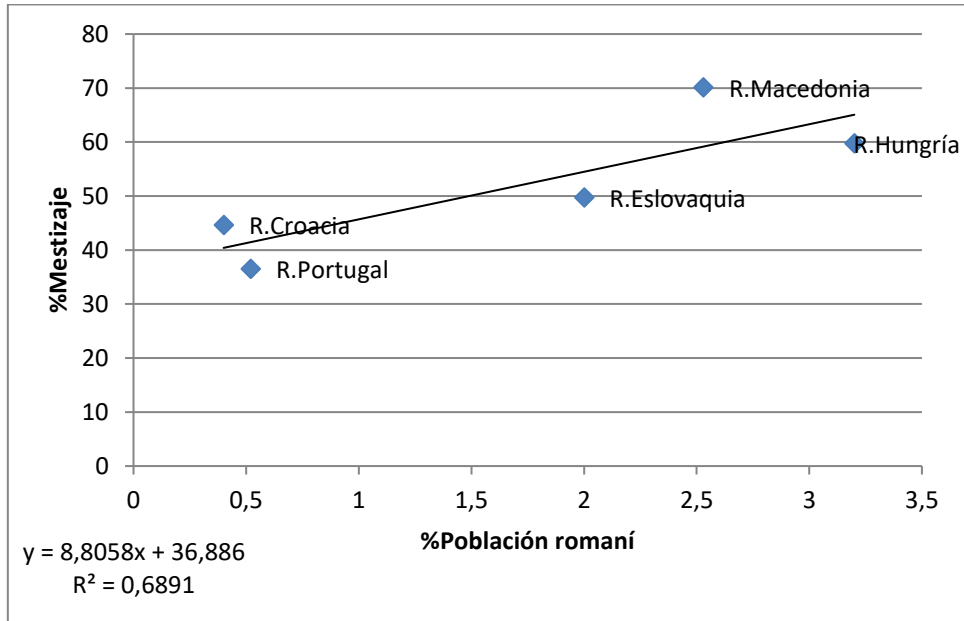


Figura 79: Recta de regresión obtenida al comparar entre el porcentaje que representa la población romaní en la población huésped frente al porcentaje de mestizaje para los marcadores del sistema CODIS.

Tabla 21: Matriz de coeficientes de correlación de Pearson y Spearman entre el porcentaje que representa la población romaní en la población huésped frente al porcentaje de mestizaje para los marcadores del sistema CODIS.

r Pearson		
	% Población total	%Mestizaje CODIS
% Población total	-	0,043151
%Mestizaje CODIS	0,82522	-

Spearman		
	% Población total	%Mestizaje CODIS
% Población total	-	0,033333
%Mestizaje CODIS	0,88571	-

V.4. Análisis de las poblaciones romaníes europeas para marcadores de linaje

V.4.1. Análisis de la población romaní para haplogrupos del Cromosoma Y

Análisis molecular de la varianza (AMOVA)

Se han recopilado las frecuencias de haplogrupos del cromosoma Y en 7 poblaciones romaníes (Martínez-Cruz et al., 2016), 14 poblaciones europeas (Semino et al., 2000; Martínez-Cruz et al., 2016) y 20 poblaciones del Sudoeste de Asia (Kivisild et al., 2003; Flores et al., 2005; Regueiro et al., 2006; Haber et al., 2012; Khurana et al., 2014). Como en casos anteriores, con el fin de comprobar la existencia de heterogeneidad genética entre las muestras europeas y asiáticas, se ha realizado un AMOVA. Como se observa en la Tabla 22, se han encontrado diferencias significativas elevadas ($P < 0,001$), tanto a nivel de individuos dentro de una población (F_{ST}), como entre poblaciones dentro de un grupo (F_{SC}), y entre los grupos continentales (F_{CT}). Estos resultados indican una gran heterogeneidad en todos los niveles de análisis.

Tabla 22: Índices de fijación (F_{ST} , F_{SC} y F_{CT}) obtenidos mediante el análisis AMOVA a partir de frecuencias de haplogrupos del cromosoma Y en dos grupos de poblaciones euroasiáticas (F_{CT} : varianza genética entre grupos. F_{SC} : varianza genética entre poblaciones dentro de grupos. F_{ST} : varianza genética entre individuos dentro de una población) (*) $P < 0,001$.**

	Índices de Fijación		
	F_{ST}	F_{SC}	F_{CT}
Haplogrupos Cromosoma Y	0,21464***	0,16315***	0,06154***

Análisis de escalamiento multidimensional (MDS)

Con el fin de visualizar las diferencias observadas de un modo más preciso, se ha realizado un análisis MDS sobre la matriz de distancias F_{ST} de las 40 poblaciones europeas y asiáticas recopiladas. Sus frecuencias se incluyen en la Tabla 8 del Anexo y los resultados del análisis se muestran en la Figura 80. La Dimensión I distribuye a las poblaciones fundamentalmente entre Europa y Asia. Así, en los valores positivos se encuentran las poblaciones koya (sur de la India), vasava, gamit, nana chaudhari y mota chaudhari (Gujarati, oeste de la India), todas ellas tribus, es decir, no pertenecientes a castas; las poblaciones romaníes de Grecia, Rumanía, Eslovaquia y Bulgaria se sitúan junto a ellas. Próximas al centroide se emplaza el resto de poblaciones del Sur de Asia, Próximo Oriente y las poblaciones romaníes española, ucraniana y húngara. Finalmente, en el extremo negativo se sitúan las poblaciones

Europeas junto a las poblaciones asiáticas de Punjab y pavagadhi chaudhari. La Dimensión II caracteriza fundamentalmente a las poblaciones de Próximo Oriente y a la población romaní española, que muestran los valores positivos más altos.

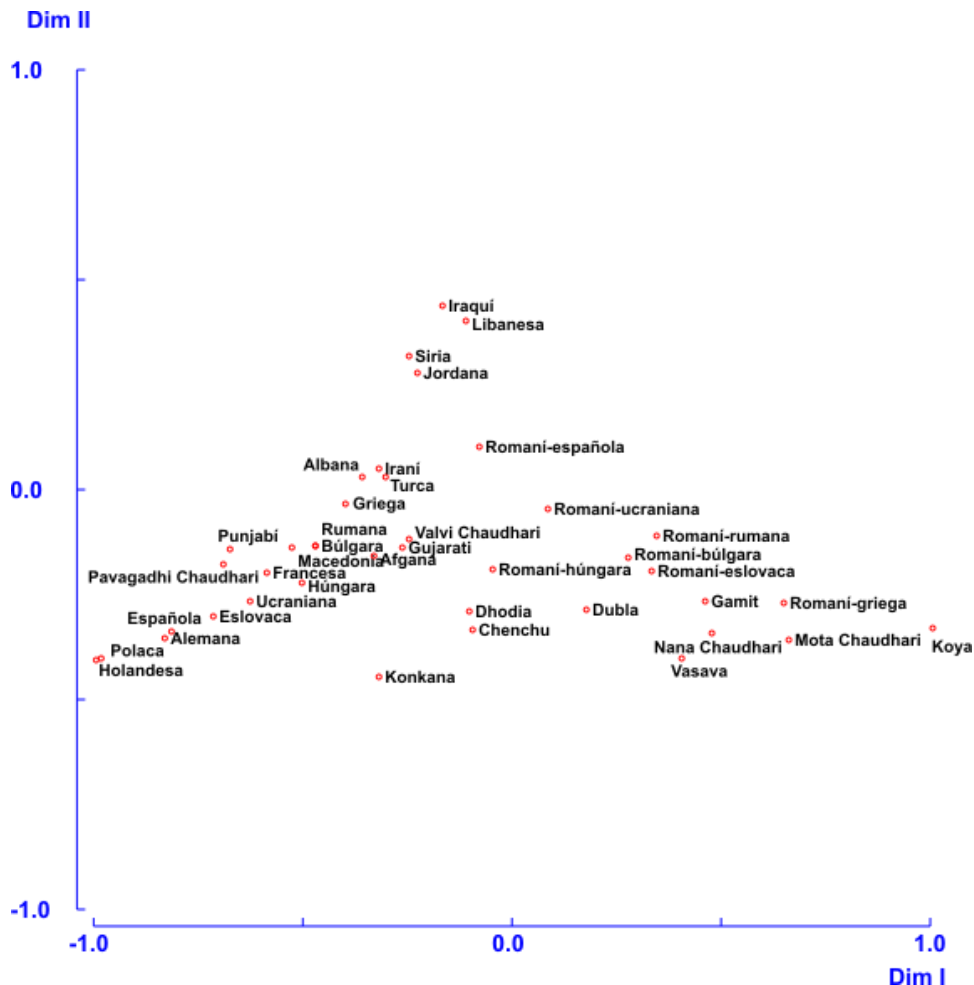


Figura 80: Análisis de Escalamiento Multidimensional a partir de la matriz F_{ST} de Reynolds sobre haplogrupos del cromosoma Y en 40 poblaciones de Europa y el Sur de Asia, incluyendo 7 poblaciones romaníes. Stress: 0,092.

Posteriormente, se han realizado sendos MDS frente a los diferentes grupos continentales. En el análisis ejecutado sobre las poblaciones europeas, incluyendo las romaníes, se observa una tendencia muy clara de variación, prácticamente en un eje (Figura 81). Así, en el lado positivo del eje de abscisas (Dimensión I) se sitúan las poblaciones no romaníes y en el extremo negativo las romaníes. Las poblaciones europeas se ordenan desde el centro y el oeste hacia el sudeste, con las poblaciones albanesa, griega y rumana situadas más próximas al centroide. Entre los grupos romaníes, los húngaros, ucranianos y españoles son los más próximos al centroide, quedando los griegos, rumanos y eslovacos como los más alejados.

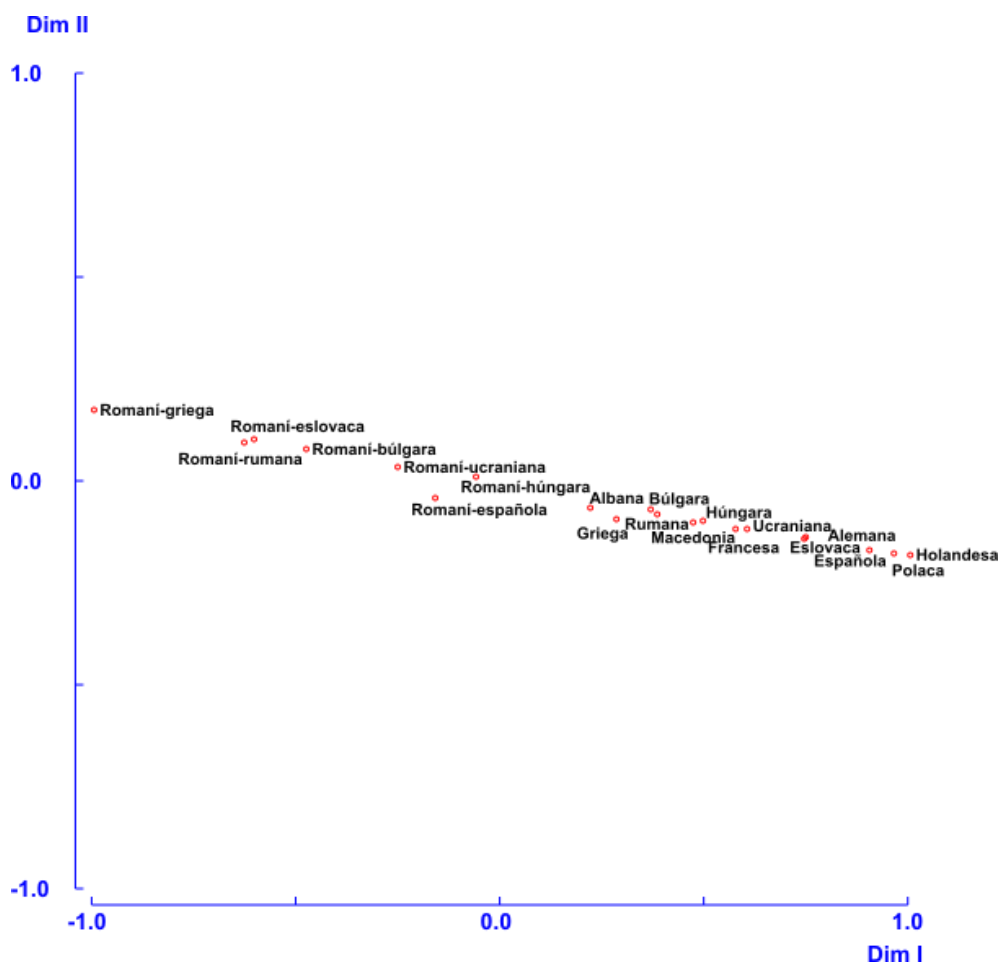


Figura 81: Análisis de Escalamiento Multidimensional a partir de la matriz F_{st} de Reynolds sobre haplogrupos del cromosoma Y en 20 poblaciones de Europa, incluyendo romaníes. Stress: 0,070.

Cuando se analizan las poblaciones de Asia junto a los romaníes, se observan dos tendencias de variación muy claras (Figura 82). En el primer eje (Dimensión I) se diferencian por un lado un grupo de poblaciones tribales de la India junto a los romaníes en los valores positivos y, por otro, el resto de poblaciones en los negativos. Así, en los valores positivos se encuentran las poblaciones tribales de koya (del este de la India), vasava, gamit, nana chaudhari y mota chaudhari (Gujarati, oeste), junto a los romaníes de Grecia, Rumanía, Eslovaquia y Bulgaria. Próximos al centroide se sitúan los romaníes de Hungría y Ucrania junto a las poblaciones chenchu (del Este de la India), dubla y dhodia (Gujarati, Oeste), también catalogadas como tribus. En los valores negativos se encuentran la población romaní española, las del Próximo Oriente y el resto de poblaciones del Sur de Asia, incluyendo también algunas tribus de la India. El segundo eje (Dimensión II) diferencia a las poblaciones del Próximo Oriente del resto. En definitiva, las poblaciones más próximas a la mayor parte de los romaníes son algunas

poblaciones tribales de la India, tanto del Este como del Oeste del subcontinente. La excepción se encontraría en la población romaní española más próxima a las del Próximo Oriente.

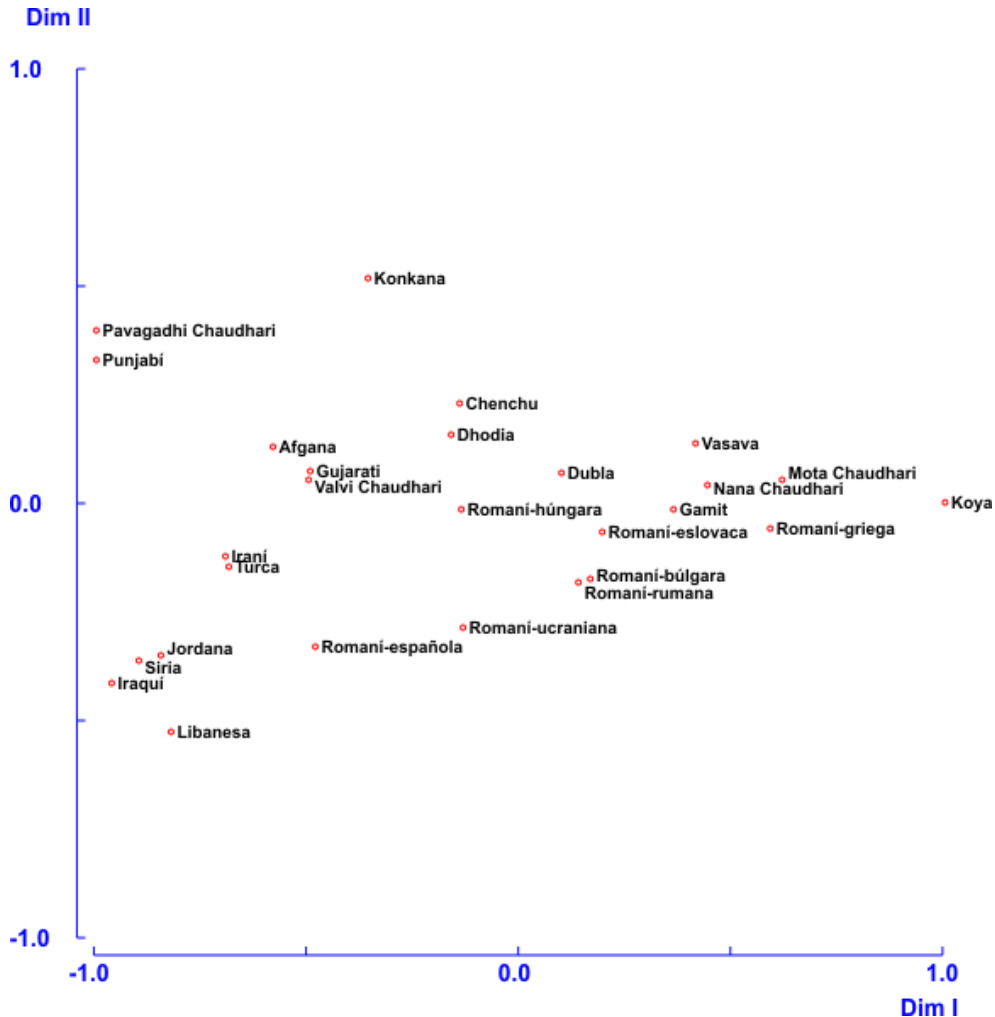


Figura 82: *Análisis de Escalamiento Multidimensional a partir de la matriz F_{ST} de Reynolds sobre haplogrupos del cromosoma Y en 27 poblaciones del sudoeste de Asia y romaníes europeos. Stress: 0,068.*

Análisis factorial de correspondencias (AFC)

Con el fin de contrastar la robustez de los resultados obtenidos mediante análisis MDS, se realizó un AFC, como en apartados anteriores. El resultado, que se muestra en la Figura 83, es básicamente similar. En el extremo negativo del eje 1 se encuentra la población koya. Siguiendo este eje hacia los valores positivos se encuentran las poblaciones mota chaudhari, nana chaudhari, vasava y gamit; a continuación, en los valores negativos del eje 2 se emplazan las poblaciones romaníes de Grecia, Eslovaquia, Rumanía y Bulgaria y, solapadas con ellas, pero

en valores positivos de dicho eje, las de dubla, dhodia y konkana, tribus del oeste de la India, y chenchu, una tribu del sur. Próximas al valor 0 del eje 1 se sitúan el resto de poblaciones romaníes, poblaciones del Sur de Asia y una parte de las del Próximo Oriente. Finalmente, en los valores positivos se localizan las poblaciones europeas y también varias del Próximo Oriente. Precisamente, el eje 2 diferencia fundamentalmente a las poblaciones del Próximo Oriente (libanesa, iraquí, siria, jordana y en menor medida iraní y turca) de todas las demás (ver Discusión).

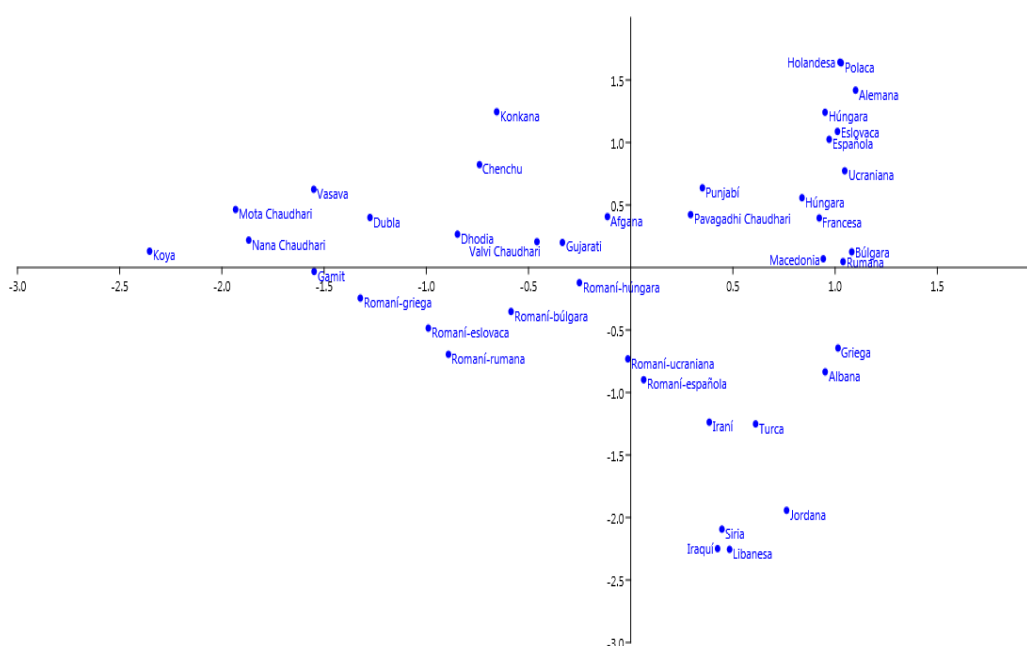


Figura 83: Análisis Factorial de Correspondencias realizado sobre las frecuencias de haplogrupos del cromosoma Y en 40 poblaciones de Europa y el Sur de Asia. Varianza explicada: 50,98%.

Método del centroide

Se ha realizado un análisis del centroide de las frecuencias de haplogrupos del cromosoma Y con el objetivo de observar el flujo génico que han experimentado las poblaciones romaníes europeas respecto al resto de poblaciones analizadas, así como su heterocigosidad y la varianza que aportan al conjunto. Este análisis no muestra un patrón geográfico claro (Figura 84). Destacan, en todo caso, las poblaciones afgana y koya; la primera es la que más claramente se sitúa por encima de la línea de los valores esperados. Dado que esta población tiene valores relativamente altos tanto de la varianza que aporta al conjunto (R_i) como de heterocigosidad (H), esto podría indicar un cierto flujo génico desde otras poblaciones ajenas al área de estudio, presumiblemente desde el centro de Asia. Los koya, por el contrario, reflejan un notable grado de aislamiento; la R_i es elevada, pero poseen un bajo valor de H ; esto explica, al menos en parte, su posición en los diferentes análisis multivariantes. Así, tanto en los MDS como en el AFC, este grupo estaba situado en un extremo del eje de abscisas, es decir, aquél que discrimina entre las poblaciones de Europa y del Sur de Asia, indicando unos valores extremos para diversos haplogrupos. De hecho, los Koya se caracterizan por tener las frecuencias más altas de los haplogrupos F y H en todo el sudoeste de Eurasia, con valores 0 para el resto de haplogrupos, excepto para el R (ver Discusión).

En lo que respecta a las poblaciones romaníes europeas estas se encuentran en su mayoría por debajo de la recta de regresión, lo que indica un grado variable de aislamiento. Este iría desde la población romaní húngara, próxima a la recta de regresión (lo que indica un grado bajo), seguido de la romaní búlgara, con un grado un poco más alto, hasta la romaní griega que presentaría el mayor grado de aislamiento. La excepción se encuentra en los romaníes ucranianos que se localizan por encima de la recta de regresión, con un mayor flujo génico de lo esperado, presumiblemente desde poblaciones europeas (ver Discusión).

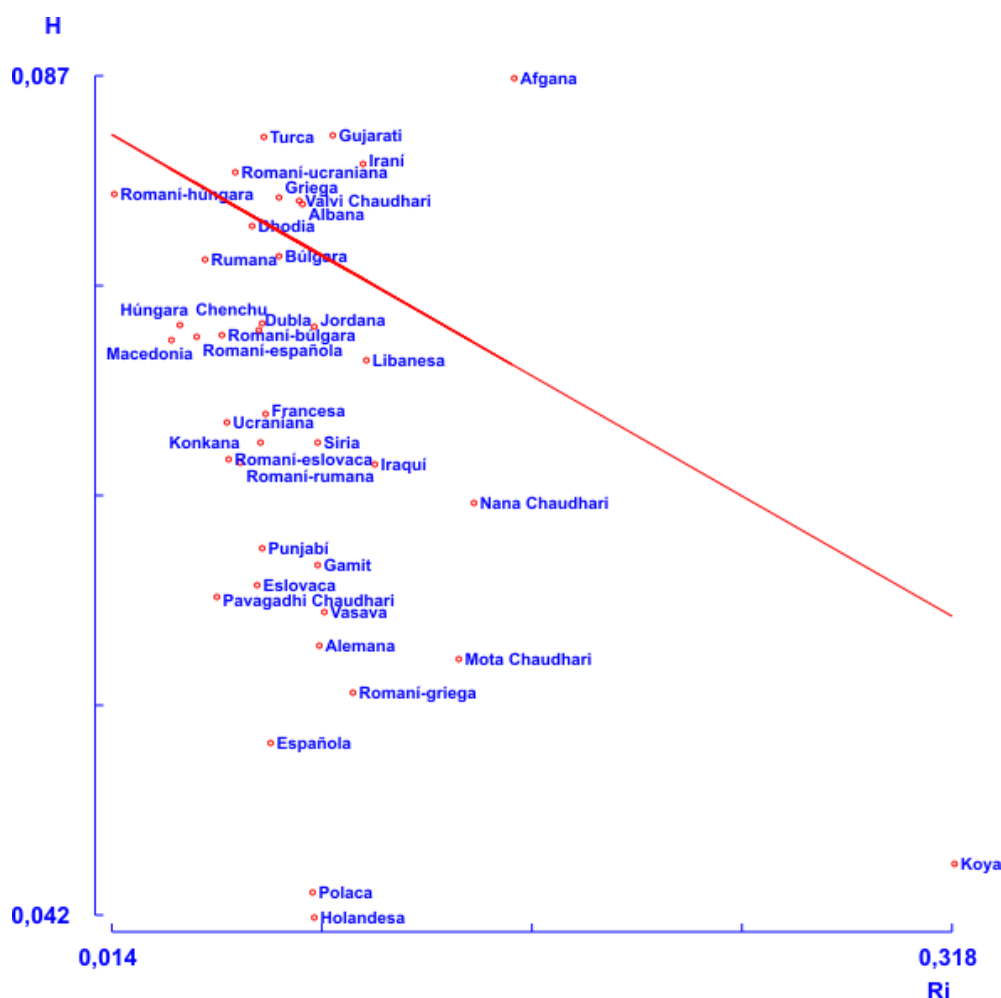


Figura 84: Método del centroide sobre frecuencias de haplogrupos del cromosoma Y en 40 poblaciones del Sudoeste de Eurasia.

Análisis de clinas para haplogrupos del cromosoma Y

Se ha abordado asimismo el análisis de clinas, como se hizo en los análisis de las familias de marcadores precedentes y, como en ellos, la mayor parte de las clinas se han dispuesto en sentido latitudinal (Figura 85). Sin embargo, las clinas que siguen siendo significativas tras la corrección de Bonferroni tienen una distribución irregular (Tabla 23). De hecho, 4 de estos haplogrupos se caracterizan por unas distribuciones marcadamente diferentes entre Europa y el Sur de Asia, con frecuencias de 0 en Europa y mayores de 0 en Asia (por ejemplo C y L), o frecuencias de 0 en Asia y por encima de 0 en Europa (como E e I). Por otra parte, aunque la clina observada para el haplogrupo H no ha resistido la corrección de Bonferroni, su distribución de frecuencias es muy interesante. En la Figura 86 se observa que este haplogrupo se encuentra

con frecuencias elevadas en el Sur de Asia y próximas a 0 en Europa, excepto en los romaníes, en los que los valores también son elevados. De hecho, este es el haplotipo más frecuente en todos los romaníes, excepto en los españoles. El hecho de que los romaníes muestren unas diferencias tan notables del resto de europeos afecta necesariamente al análisis de la clina. Así, una vez eliminados los romaníes del análisis, la clina resulta altamente significativa ($R: 0,745$; g.l.: 31; $P = 6,75 \cdot 10^{-7}$; $P_{\text{Bonferroni}} = 1,28 \cdot 10^{-5}$).

Tabla 23: Características de las clinas observadas en las frecuencias de haplogrupos del cromosoma Y en 40 poblaciones del Sudoeste de Eurasia.

Haplogrupo	Ángulo	Intersección	Pendiente	r	df	p	Test de Bonferroni
C	111	-0,0247	0,1078	0,6297	39	1,03E-05	p<0,01
E	242	0,1576	-0,1437	0,4785	39	0,0016	p<0,05
F	172	-0,0107	0,0641	0,4230	39	0,0059	
H	172	0,0447	0,3562	0,4388	39	0,0041	
I	358	0,2384	-0,3144	0,8214	39	4,7E-11	p<0,01
K	180	0,0001	0,0259	0,3492	39	0,0252	
L	67	-0,0261	0,0868	0,5965	39	3,86E-05	p<0,01
M	27	-0,0006	0,0013	0,3342	39	0,03273	
N	17	-0,0048	0,0209	0,4931	39	0,0011	p<0,05
O	27	-0,0046	0,0097	0,3569	39	0,0220	
Q	86	-0,0094	0,0369	0,4388	39	0,0041	
T	214	0,0087	-0,0109	0,3345	39	0,0325	

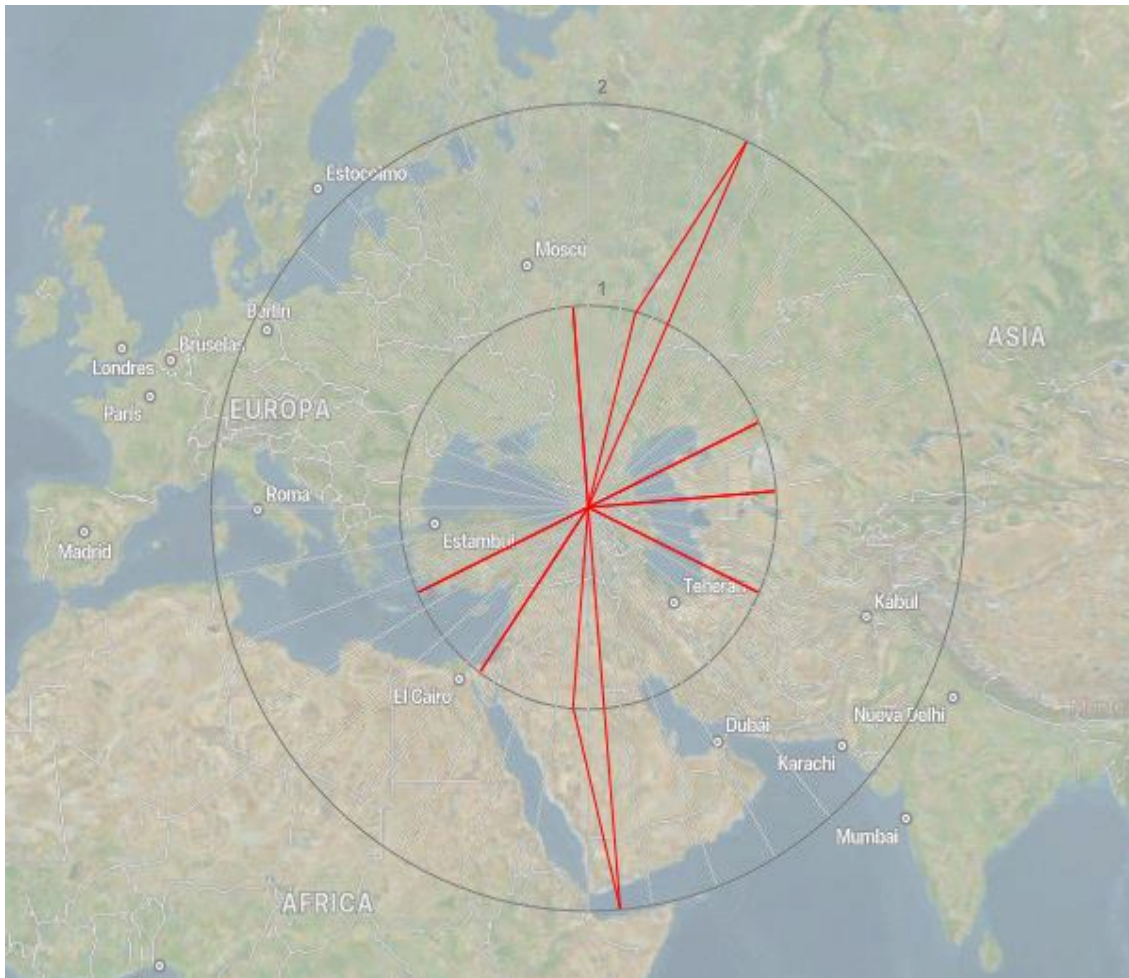


Figura 85: Orientación y frecuencia de las clinas observadas en las frecuencias de haplogrupos del cromosoma Y en 40 poblaciones del Sudoeste de Eurasia.

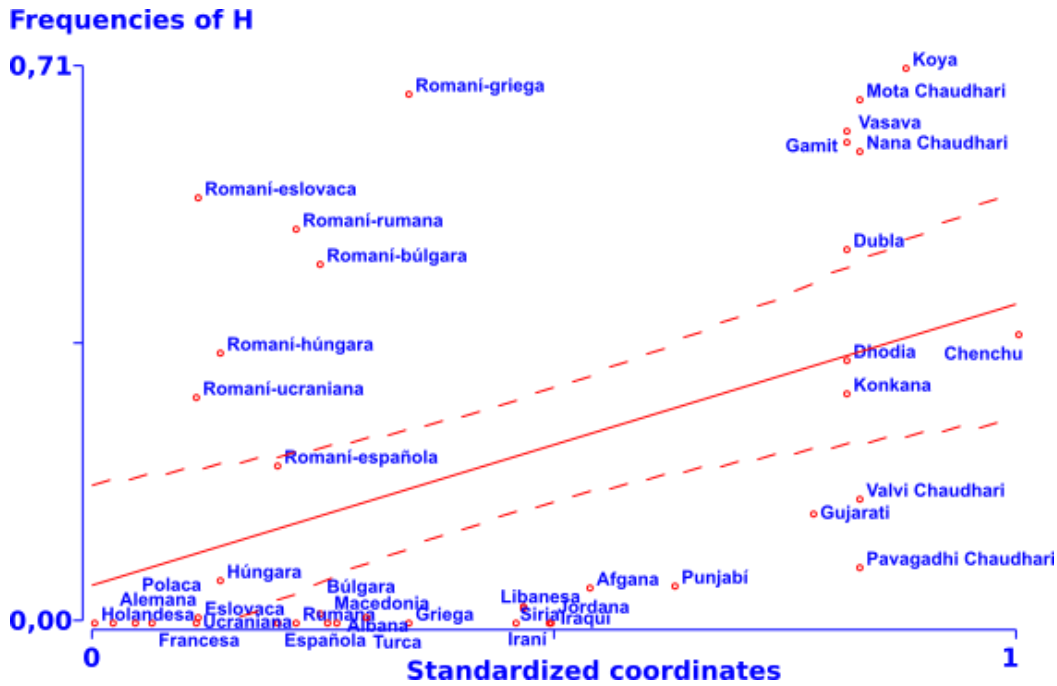


Figura 86: Clina observada para el haplogrupo H en el Sudoeste de Eurasia, en la que se muestra la distribución de frecuencias en las 40 poblaciones analizadas.

Possible acción de la deriva genética

En lo que respecta al análisis de la posible acción de la deriva genética se ha realizado mediante el estadístico $-\log(P)$ (Peña et al., 1997). Para las estimaciones del flujo génico y la deriva se han utilizado como población de referencia del Sur de Asia las medias de las frecuencias alélicas de las poblaciones tribales del estado de Gujarati (dhodia, dubla, konkana, vasava, gamit, valvi chaudhari, nana chaudhari y mota chaudhari), excepto la población pavagadhi chaudhari; tampoco se ha utilizado la muestra de población general de Gujarati. Se observan evidencias de deriva en los haplogrupos J, L y R (Tablas 24 y 25). Tanto en el supuesto sin migración como en el supuesto con migración, J y R parecen haber experimentado variaciones al azar en varias poblaciones romaníes. Así, el haplogrupo J muestra valores significativos en todas ellas, excepto en la romaní griega, y el R ha sido significativo en las de Bulgaria, Grecia, Rumanía y Ucrania. Finalmente, el haplogrupo L ha mostrado evidencias de acción de la deriva en los romaní griegos. Cabe pensar que el caso de los haplogrupos J y R, por ser un fenómeno generalizado, se trate del reflejo un proceso de cuello de botella temprano o de un efecto fundador, quizá en las primeras fases de la diáspora de esta etnia (ver Discusión).

Los valores más elevados de $-\log(P)$ se han obtenido en las poblaciones romaníes de España, Bulgaria y Ucrania. Estos resultados son paradójicos debido a lo observado previamente

en los análisis MDS y AFC. En las Figuras 80 y 83 dichos análisis situaron a estas poblaciones más desviadas que el resto respecto de la zona intermedia del gráfico (entre las poblaciones europeas y asiáticas), donde tendrían que haberse emplazado en el caso de no existir influencia alguna de la deriva. No obstante, es difícil precisar cuáles han sido las poblaciones con una mayor acción de la deriva, ya que el número de haplogrupos es relativamente escaso.

Tabla 24: Valores de $-\log(P)$ para el caso hipotético sin migración, obtenidos a partir de las frecuencias de haplogrupos del cromosoma Y en 7 poblaciones de romaníes europeos. Los valores significativos se marcan con un asterisco. Cuando se especifica >323 , el valor de $-\log(P)$ no ha podido calcularse, por ser superior a este valor y estar por encima de las capacidades del procesador. Se ha anotado un guion cuando no se ha podido calcular el cociente de F_{ST} , por existir una división con 0 en el denominador.

Haplogrupo	Romaníes Bulgaria	Romaníes Grecia	Romaníes Hungría	Romaníes Rumanía	Romaníes Eslovaquia	Romaníes España	Romaníes Ucrania
C	0,434	0,434	0,393	0,434	0,434	0,434	0,434
E	0,434	0,458	0,468	0,724	0,528	0,848	0,621
F	0,434	0,434	0,434	0,434	0,434	0,434	0,434
G	0,47	0,853	0,403	0,862	0,853	0,862	0,861
H	0,269	0,393	0,258	0,28	0,308	0,25	0,232
I	0,386	0,397	0,414	0,443	0,494	0,722	0,4
J	74,470*	0,424	8,016*	3,977*	2,629*	>323*	26,259*
K	0,434	0,434	0,434	0,434	0,434	0,434	0,434
L	0,434	0,944	0,434	0,434	0,434	0,434	0,434
N	0,865	-	0,434	0,866	0,859	0,866	0,875
Q	0,434	0,434	0,434	0,434	0,434	0,434	0,434
R	5,772*	5,689*	0,468	2,427*	0,712	0,519	1,137
T	0,862	0,859	-	0,866	0,866	0,861	-

Tabla 25: Valores de $-\log(P)$ para el caso hipotético con migración, obtenidos a partir de las frecuencias de haplogrupos del cromosoma Y en 7 poblaciones de romaníes europeos. Los valores significativos se marcan con un asterisco. Se ha anotado un guion cuando no se ha podido calcular el cociente de F_{ST} , por existir una división con 0 en el denominador.

Haplogrupo	Romaníes Bulgaria	Romaníes Grecia	Romaníes Hungría	Romaníes Rumanía	Romaníes Eslovaquia	Romaníes España	Romaníes Ucrania
C	0,854	0,854	0,773	0,854	0,854	0,854	0,854
E	0,452	0,486	0,474	0,747	0,532	0,858	0,631
F	0,863	0,863	0,863	0,863	0,863	0,863	0,863
G	0,471	0,861	0,407	0,865	0,861	0,865	0,865
H	0,478	0,712	0,429	0,508	0,552	0,453	0,42
I	0,414	0,408	0,436	0,47	0,514	0,735	0,423
J	98,327*	0,519	10,443*	4,945*	4,015*	>323*	36,137*
K	0,865	0,865	0,865	0,865	0,865	0,865	0,865
L	0,864	1,506*	0,864	0,864	0,864	0,864	0,864
N	0,867	-	0,434	0,867	0,863	0,867	0,879
Q	0,863	0,863	0,863	0,863	0,863	0,863	0,863
R	7,479*	7,350*	0,595	3,112*	0,927	0,693	1,450*
T	0,865	0,864	-	0,867	0,867	0,865	-

Estimación del grado de mestizaje

Para el cálculo del grado de mestizaje se han eliminado, como en apartados anteriores, los haplogrupos que han presentado evidencias notables de la acción de la deriva. Los resultados se muestran en la Tabla 26. En promedio, el cromosoma Y muestra un mayor grado de aislamiento, hasta el punto de que algunas poblaciones no parecen haber tenido flujo génico desde otras poblaciones, asociado éste a los linajes masculinos. Es por ejemplo el caso de los romaníes de Grecia y Eslovaquia, en los que no se aprecia ningún grado de mestizaje. Los rumanos, con un 90,2 % parecen mantener también en gran medida el patrimonio genético original en lo que se refiere al cromosoma Y. Los romaníes de España (71,8 %), Bulgaria (69,8 %) y Hungría (65,3 %) mantienen un componente asiático mayoritario en la fracción masculina de su patrimonio genético. Por último, los romaníes de Ucrania han experimentado un mayor grado de mestizaje, conservando el 44,1 % de sus linajes masculinos originales.

Tabla 26: Estimación del grado de mestizaje (\pm error estándar) e intervalo de confianza del 95% en 7 poblaciones romaníes europeas, obtenidas a partir de frecuencias de haplogrupos del cromosoma Y.

Población romaní	Componentes	Coefficientes	Error estándar	Límite inferior	Límite superior
Bulgaria	NW India	0,698	0,136	0,390	1,005
	Bulgaria	0,302	0,163	-0,066	0,671
Grecia	NW India	1,000	0,189	0,573	1,427
	Grecia	0,000	0,279	-0,631	0,631
Hungría	NW India	0,653	0,081	0,472	0,833
	Hungría	0,347	0,082	0,166	0,529
Rumanía	NW India	0,902	0,106	0,662	1,143
	Rumanía	0,098	0,151	-0,243	0,439
Eslovaquia	NW India	1,000	0,110	0,750	1,250
	Eslovaquia	0,000	0,239	-0,540	0,540
España	NW India	0,718	0,186	0,304	1,133
	España	0,282	0,140	-0,029	0,593
Ucrania	NW India	0,441	0,107	0,198	0,684
	Ucrania	0,559	0,171	0,172	0,947

La Figura 87 muestra una comparación entre el porcentaje de mestizaje con la población huésped frente al porcentaje de población romaní en el país correspondiente; ambos valores se muestran en la Tabla 10 del Anexo. En la figura no se observa, a simple vista, una relación entre ambas variables. Las correlaciones de Pearson y Spearman no han mostrado valores

significativos en ninguno de los casos (Tabla 27). Estos resultados contrastan con los obtenidos en los STR del sistema CODIS que sí mostraban correlación significativa (Tabla 21). Esta discrepancia podría deberse, en primer lugar, a que hemos analizado únicamente el linaje masculino y, en segundo lugar, a que las poblaciones romaníes europeas analizadas en ambos casos no son las mismas, con la excepción de la húngara y la eslovaca. El porcentaje de mestizaje de la población romaní húngara con la población general, mediante marcadores STR, fue del 59,7% (Tabla 21), muy superior al 34,7% obtenido con los haplogrupos del cromosoma Y. Un caso más extremo es el eslovaco, con un 49,7% en los análisis basados en marcadores STR frente al 0% mediante el uso de haplogrupos del cromosoma Y.

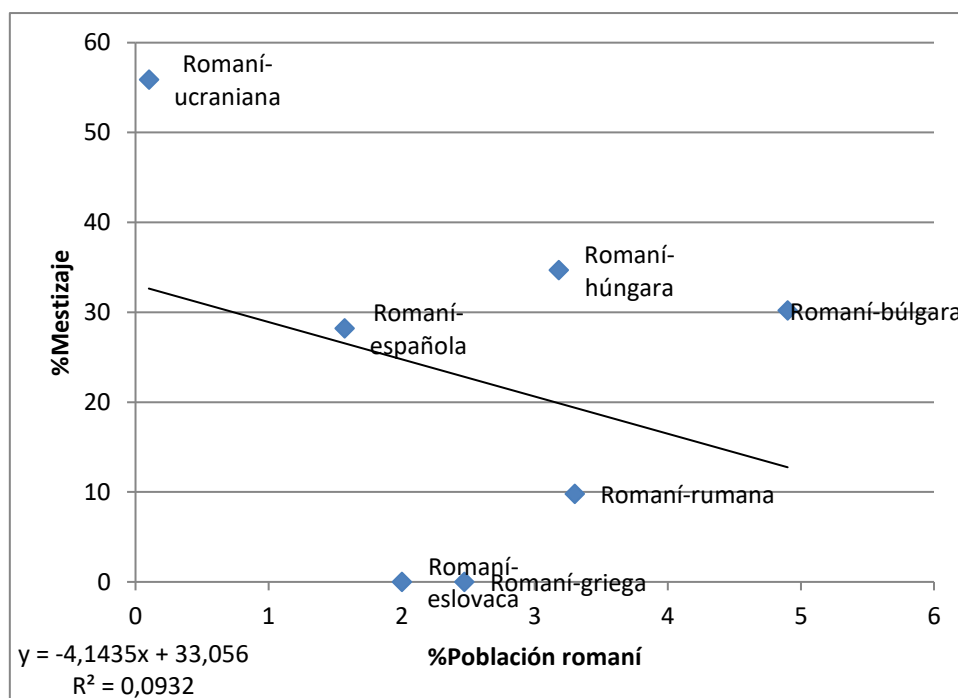


Figura 87: Recta de regresión al comparar entre el porcentaje que representa la población romaní en la población huésped frente al porcentaje de mestizaje para los haplogrupos del cromosoma Y.

Tabla 27: Matriz de coeficientes de correlación de Pearson y Spearman entre el porcentaje que representa la población romaní en la población huésped frente al porcentaje de mestizaje para los haplogrupos del cromosoma Y.

r Pearson			Spearman		
	% Población total	%Mestizaje Crom Y		% Población total	%Mestizaje Crom Y
% Población total	-	0,50545	% Población total	-	0,79286
%Mestizaje Crom Y	-0,30535	-	%Mestizaje Crom Y	-0,12613	-

V.4.2. Análisis de la población romaní para haplogrupos de ADN mitocondrial

A partir de los trabajos de Martínez-Cruz et al. (2016); Metspalu et al. (2004) y Ottoni et al. (2011), se obtuvieron las frecuencias de los haplogrupos de ADNmt de 7 poblaciones romaníes europeas y sus correspondientes poblaciones huésped (Bulgaria, Hungría, Eslovaquia, España, Grecia, Rumanía y Ucrania), así como las frecuencias de una serie de poblaciones europeas y del Sur de Asia, sumando un total de 43 poblaciones (ver Tabla 9 del anexo).

Análisis molecular de la varianza (AMOVA)

En la Tabla 28 se muestran los resultados del AMOVA para 28 haplogrupos de ADNmt en los 2 grupos continentales establecidos, correspondientes a Europa y el Sur de Asia. Se observan diferencias significativas elevadas ($P < 0,001$) tanto a nivel de individuos dentro de una población (F_{ST}), de las poblaciones dentro de un grupo (F_{SC}), como entre grupos continentales (F_{CT}). Estos resultados indican una gran heterogeneidad en todos los niveles de análisis.

Tabla 28: Índices de fijación (F_{ST} , F_{SC} y F_{CT}) generados mediante el análisis AMOVA para frecuencias de 28 haplogrupos de ADNmt en poblaciones euroasiáticas. (F_{CT} : varianza genética entre grupos. F_{SC} : varianza genética entre poblaciones dentro de grupos. F_{ST} : varianza genética entre individuos dentro de una población) (* $P < 0,05$; ** $P < 0,01$; * $P < 0,001$).**

Índices de Fijación			
	F_{ST}	F_{SC}	F_{CT}
Haplogrupos ADNmt	0,19858***	0,10462***	0,10494***

Análisis de escalamiento multidimensional (MDS)

El análisis MDS realizado a partir de la matriz de distancias de Reynolds revela la existencia de dos poblaciones claramente diferenciadas, las romaníes de Ucrania y la población rajput (Figura 88). Las primeras aparecen claramente separadas de europeas y asiáticas, por lo que cabría esperar una elevada acción de la deriva genética. La población rajput, por su parte, muestra una escasa variabilidad genética para el ADNmt, con un haplogrupo claramente mayoritario, el M, con una frecuencia muy por encima del resto de poblaciones (0,714) y tan sólo con 5 haplogrupos presentes. El resto de poblaciones se distribuye en una nube sin solución de continuidad, pero polarizadas entre un extremo europeo (en la parte superior izquierda del gráfico) y un extremo asiático (en la parte inferior derecha del gráfico). Las poblaciones del Próximo y Medio Oriente tienden a situarse en una posición intermedia (ver Discusión).

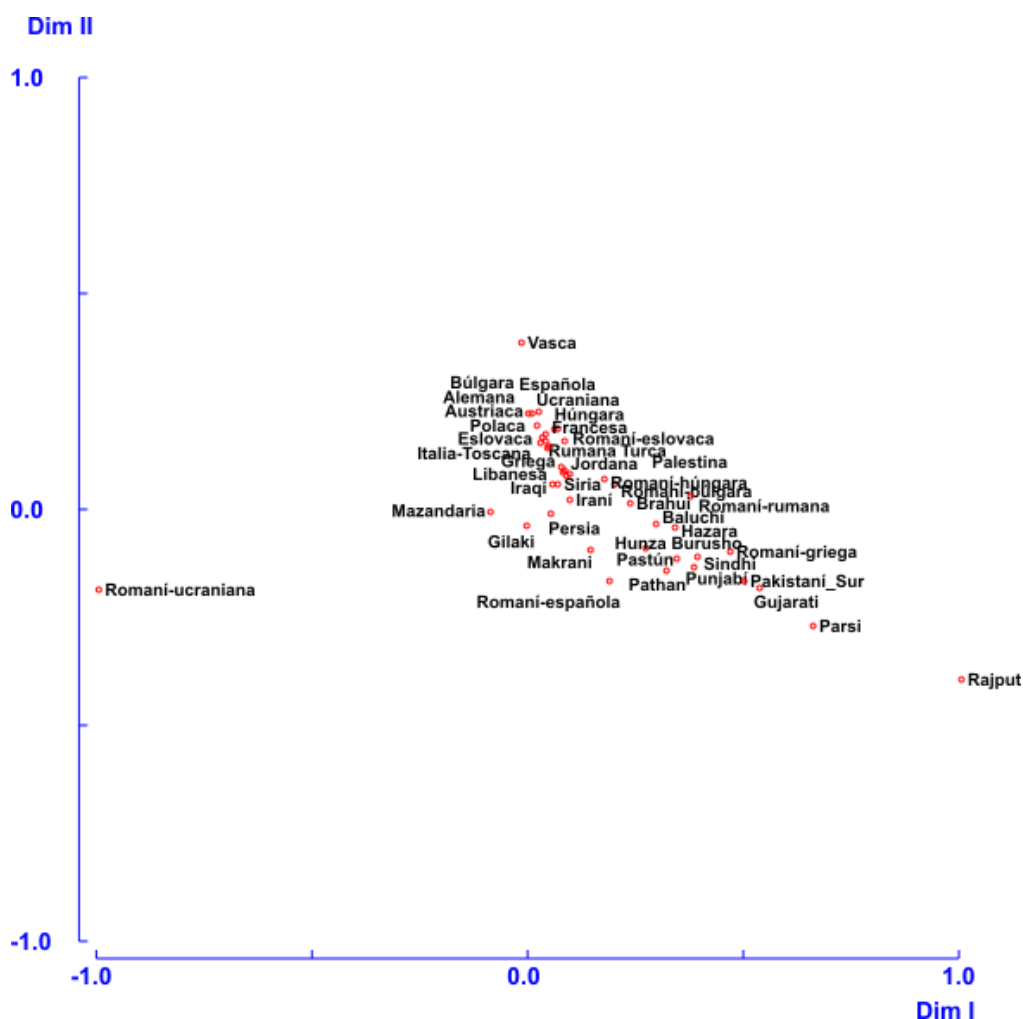


Figura 88: Análisis de Escalamiento Multidimensional a partir de la matriz F_{ST} de Reynolds sobre haplogrupos de ADN mitocondrial en 43 poblaciones de Europa y el Sur y Oeste de Asia, incluyendo 7 poblaciones romaníes. Stress: 0,102.

Considerando exclusivamente las 20 poblaciones europeas, incluyendo a las romaníes, se observa una disposición muy característica de éstas en el análisis MDS (Figura 89). Como era de esperar, la población romaní de Ucrania aparece alejada del resto, sugiriendo de nuevo una importante acción de la deriva. Las otras 6 poblaciones romaníes se sitúan ordenadas en relación a las europeas, de modo que las más alejadas son las de Grecia, España y Rumanía y las más próximas son las de Bulgaria, Hungría y Eslovaquia, lo que posiblemente refleja los diferentes grados de mestizaje de cada población (ver Discusión).

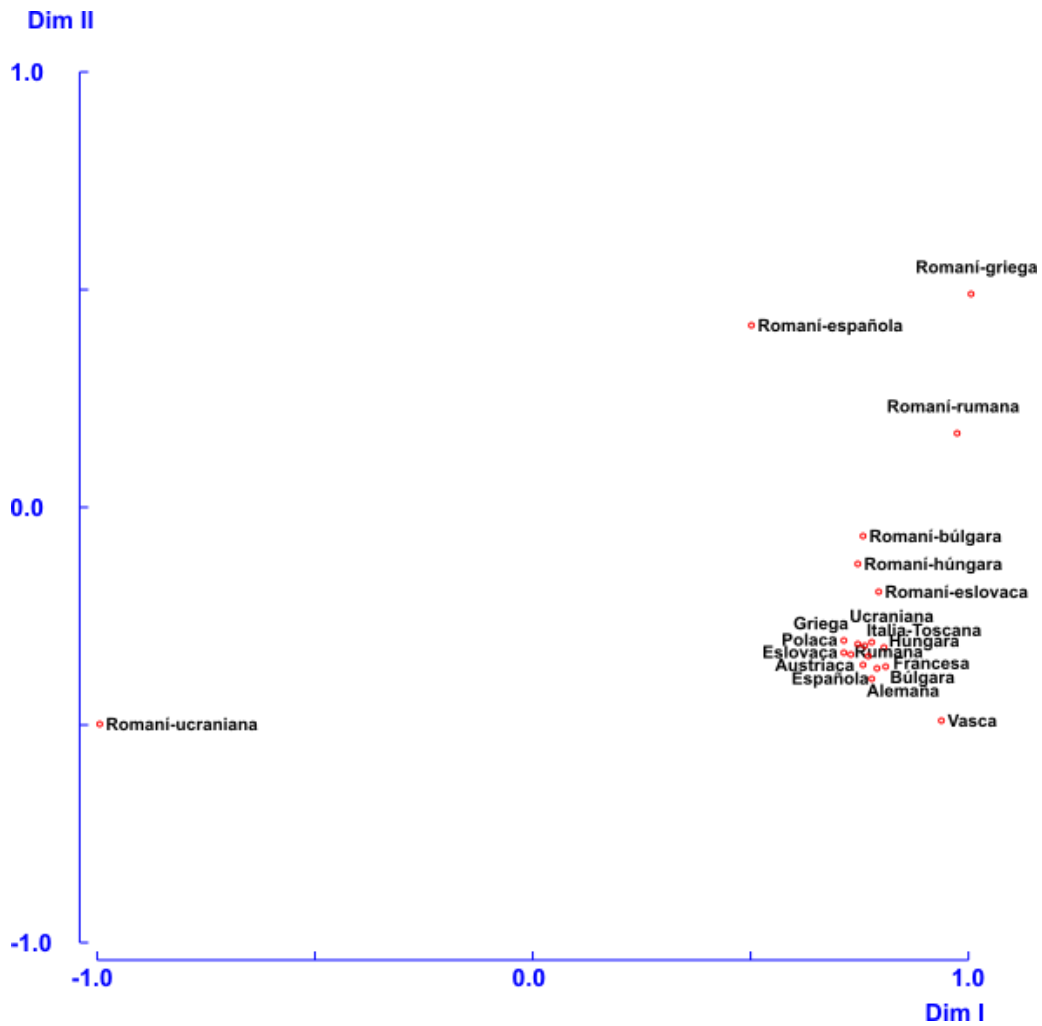


Figura 89: Análisis de Escalamiento Multidimensional a partir de la matriz F_{ST} de Reynolds sobre haplogrupos de ADN mitocondrial en 20 poblaciones de Europa, incluyendo 7 poblaciones romaníes. Stress: 0,070.

En el análisis MDS realizado sobre 13 poblaciones del Sur de Asia y 7 poblaciones romaníes europeas, éstas se solapan con las asiáticas, con la excepción de la población romaní ucraniana, que de nuevo muestra una gran heterogeneidad, y la población de rajput, localizadas ambas en los extremos de la Dimensión I (Figura 90).

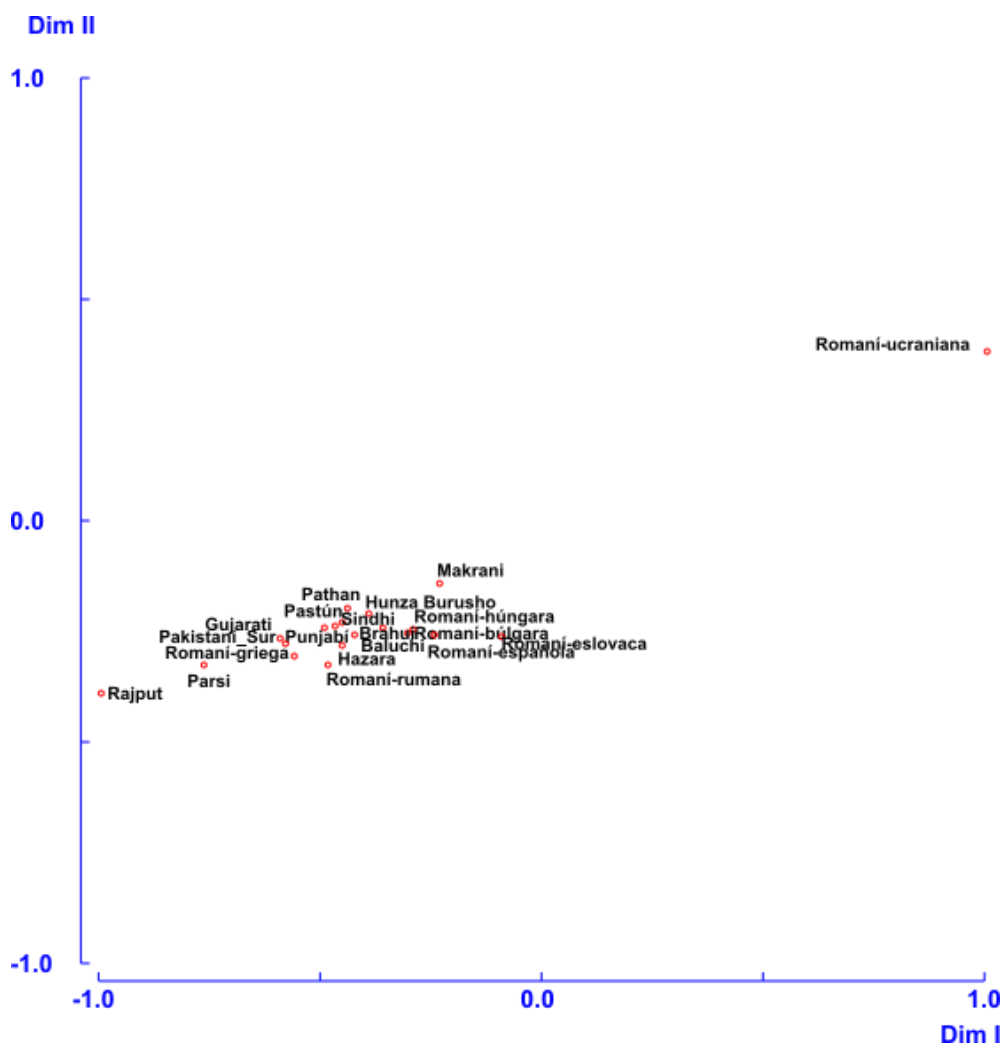


Figura 90: Análisis de Escalamiento Multidimensional a partir de la matriz F_{st} de Reynolds sobre haplogrupos de ADN mitocondrial en 13 poblaciones del Sur de Asia y 7 poblaciones romaníes. Stress: 0,064.

Análisis factorial de correspondencias (AFC)

Se ha realizado un AFC sobre las frecuencias del ADNmt de las 43 poblaciones euroasiáticas (Figura 91), con el fin de contrastar los resultados obtenidos previamente mediante el análisis MDS (Figura 88). El gráfico muestra un eje 1 condicionado por las diferentes frecuencias observadas entre las poblaciones europeas y las asiáticas. En el extremo positivo se localizan las poblaciones europeas, y también en la mitad positiva, pero más próximas al centroide, las del Próximo y Medio Oriente (palestina, siria, iraquí, jordana, libanesa, turca y varias muestras iraníes). En esta misma mitad del eje 1 se encuentran también varias poblaciones romaníes, como las de Ucrania, Eslovaquia, Hungría, España y Bulgaria. En la mitad negativa del

eje 1 se disponen las muestras de la India, Pakistán y Afganistán, junto a las poblaciones romaníes de Rumanía y Grecia. Rajput, como se observó en los análisis MDS, aparece en un extremo del gráfico. El eje 2 separa a los romaníes de Ucrania y Eslovaquia del resto de poblaciones.

Estos resultados son básicamente similares a los obtenidos con el análisis MDS, aunque en este caso la población romaní eslovaca muestra una mayor heterogeneidad respecto a otras poblaciones no romaníes. Por otra parte, algunas poblaciones romaníes tienden a encontrarse más próximas a las europeas que en el análisis MDS, donde se solapaban con las poblaciones de procedencia asiática.

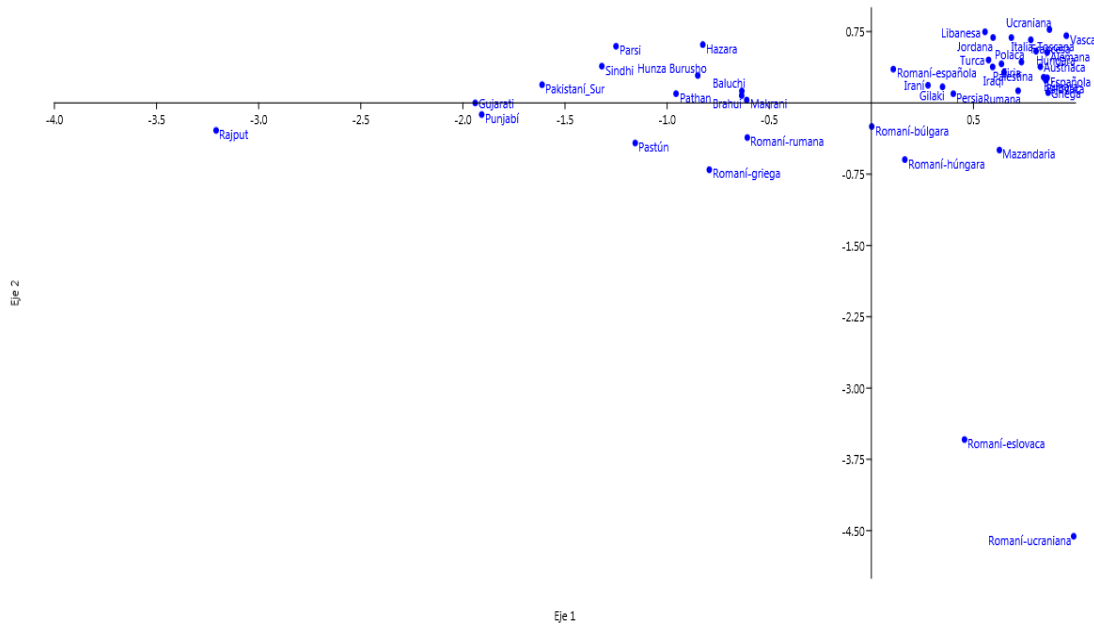


Figura 91: Análisis Factorial de Correspondencias realizado sobre las frecuencias de haplogrupos de ADNmt en 43 poblaciones de Europa y el Sur de Asia. Varianza explicada: 38,08 %.

La figura 92 muestra los resultados anteriores junto a la posición de los haplogrupos. Los haplogrupos responsables del alto grado de diferenciación de la población Rajput son el M, que ya se ha mencionado anteriormente por presentar un valor mayoritario, pero también los R10 y R, que son otros dos de los 5 haplogrupos presentes en esta población. Para los tres haplogrupos mencionados la población rajput muestra los valores máximos observados en Eurasia occidental. La posición de las poblaciones romaníes de Eslovaquia y Ucrania se encuentran condicionadas fundamentalmente por el haplogrupo I, que presenta los dos valores más altos en ambas (0,319 y 0,118, respectivamente) y, en menor medida, por el J.

Presumiblemente, estos valores deberían ser el resultado de la acción de la deriva (ver Discusión).

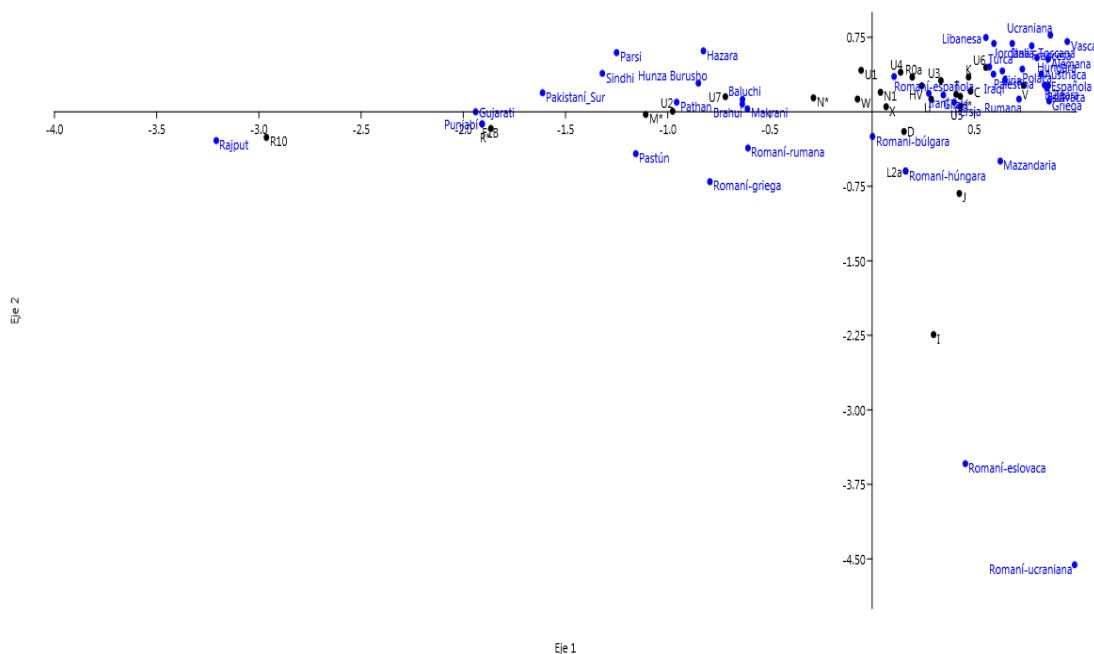


Figura 92: Análisis Factorial de Correspondencias realizado sobre las frecuencias de haplogrupos de ADNmt en 43 poblaciones de Europa y el Sur de Asia. Varianza explicada: 38,08 %. En este gráfico, que es el resultado del mismo análisis que el anterior, se muestran las posiciones de poblaciones y haplogrupos simultáneamente.

Método del centroide

El análisis centroide confirma el alto grado de aislamiento de las poblaciones rajput y romaní de Ucrania localizadas muy por debajo de la recta de regresión que se muestra en la Figura 93. Ambas aparecen en la parte inferior derecha del gráfico, y muestran un valor elevado de R_i en relación a su grado de heterocigosidad observada. Las poblaciones romaníes española y eslovaca, por el contrario, muestran unos valores que reflejan un flujo génico mayor del esperado, aunque el origen del mismo no puede deducirse de este análisis. La situación en el gráfico de las poblaciones romaníes griega y húngara, que se encuentran sobre la recta, indica cierto grado de flujo génico, a diferencia de las poblaciones romaníes búlgara, y en mayor medida, rumana, que posiblemente han experimentado un cierto grado de aislamiento (ver Discusión).

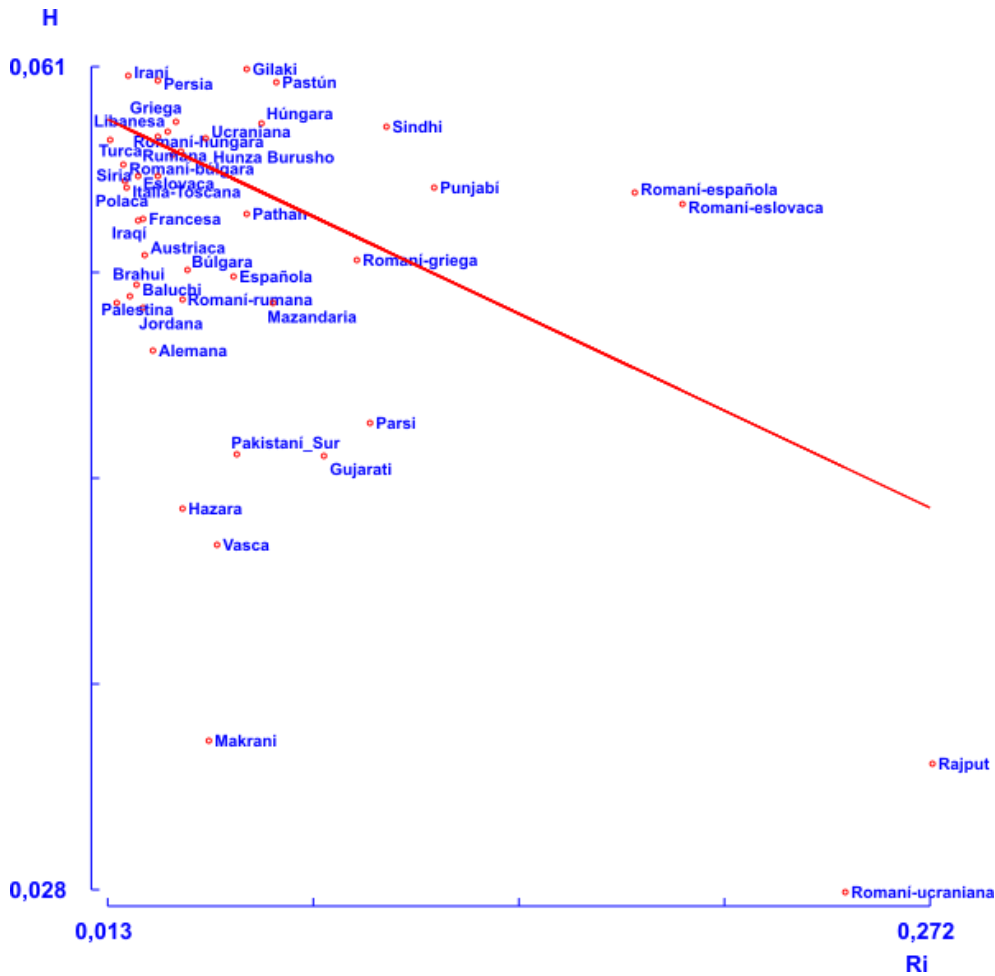


Figura 93: Análisis del centroide sobre frecuencias de haplogrupos de ADNmt en 43 poblaciones del Sudoeste de Eurasia.

Análisis de clinas de haplogrupos del ADNmt

Al igual que en los análisis previos con otras familias de marcadores la mayor parte de las clinas toman, en este análisis, una disposición latitudinal (Figura 94). En particular, tras la corrección de Bonferroni sólo 7 haplogrupos del ADNmt siguen mostrando significación (Tabla 29). De entre ellos, sólo el haplogrupo M es responsable de la diferenciación de una población surasiática (rajput); el resto muestra una mayor permanencia en poblaciones europeas. De hecho, a diferencia de casos anteriores, un número considerable de clinas no sigue esa tendencia latitudinal, llegando a mostrar un patrón longitudinal noroeste-sureste, coincidente con la línea migratoria que une el subcontinente indio y Europa.

Resultados

Tabla 29: Características de las clinas observadas en las frecuencias de haplogrupos de ADNmt en 43 poblaciones del Sudoeste de Eurasia.

Haplogrupo	Ángulo	Intersección	Pendiente	r	df	p	Test de Bonferroni
H*	310	0,4430	-0,3959	0,8195	41	1,82E-11	p<0,01
I	6	-0,0192	0,0694	0,3344	41	2,84E-02	
J	11	0,0093	0,1584	0,3480	41	2,22E-02	
K	342	0,0578	-0,0496	0,4877	41	9,08E-04	p<0,05
L2,6	206	0,0022	-0,0034	0,3884	41	1,01E-02	
M*	163	-0,0147	0,3921	0,6419	41	3,50E-06	p<0,01
R*	46	-0,0102	0,0311	0,3422	41	2,47E-02	
R0a	193	0,0165	-0,0177	0,3240	41	3,40E-02	
T	284	0,0993	-0,0586	0,3795	41	1,21E-02	
U1	166	0,0030	0,0191	0,3141	41	4,02E-02	
U2	133	-0,0144	0,0990	0,6500	41	2,39E-06	p<0,01
U3	208	0,0932	-0,1165	0,4408	41	3,09E-03	
U5	359	0,0828	-0,0925	0,6756	41	6,68E-07	p<0,01
U6	203	0,0121	-0,0167	0,6509	41	2,30E-06	p<0,01
U7	108	-0,0120	0,0700	0,6571	41	1,70E-06	p<0,01
V	291	0,0134	-0,0146	0,3665	41	1,56E-02	
W	17	-0,0056	0,0270	0,3437	41	2,40E-02	

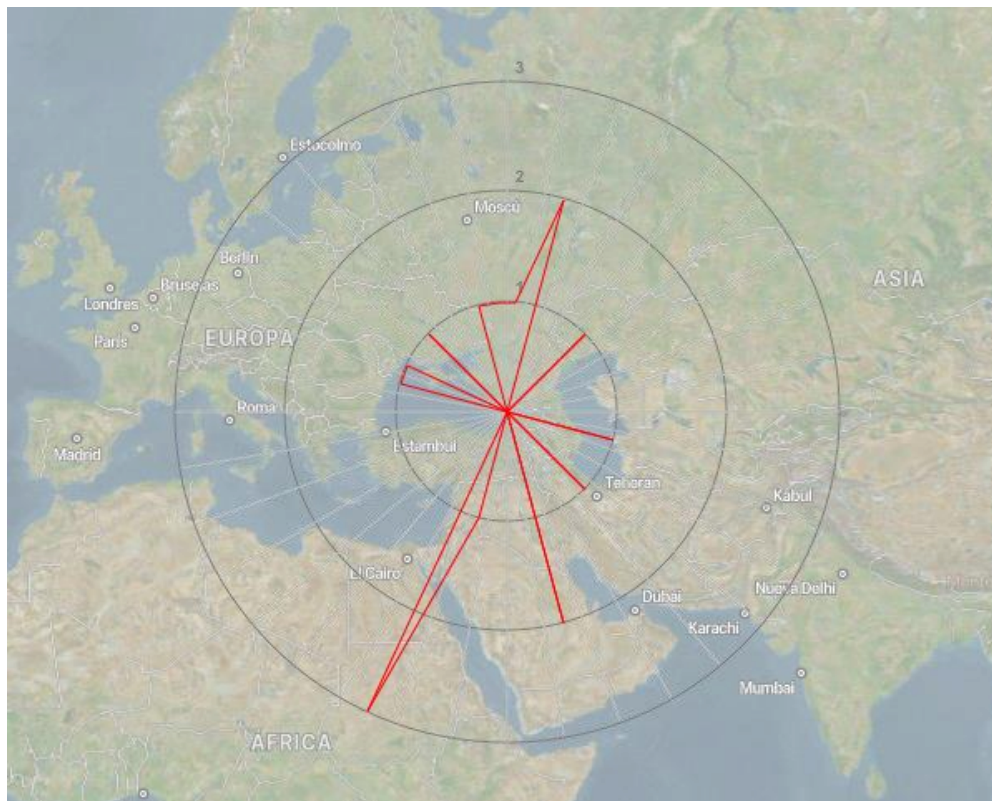


Figura 94: Orientación y frecuencia de las clinas observadas en las frecuencias de haplogrupos de ADNmt en 43 poblaciones del Sudoeste de Eurasia.

Possible acción de la deriva genética

En lo que concierne al análisis de la posible acción de la deriva genética, mediante $-\log(P)$ (Peña et al., 1997), se ha utilizado como población de referencia del Sur de Asia para las estimaciones del flujo génico y la deriva la muestra del Punjab. En conjunto, se han encontrado evidencias de deriva en 9 de los 27 haplogrupos analizados, concretamente en HV, I, J, N*, N1, T, U1, U5, y W, tanto en el supuesto sin migración (Tabla 30) como con migración (Tabla 31). Los haplogrupos con un mayor número de poblaciones afectadas han sido el I y W en 5 de ellas, y pueden haberse generado por un efecto fundador inicial durante las primeras fases de la migración, o bien por un cuello de botella en el mismo periodo.

La población con un mayor número de haplogrupos con evidencias del efecto de la deriva ha sido la rumana, con 6 haplogrupos afectados, todos ellos presentes con valores muy bajos en esta población, por lo que no ha modificado de forma notable su posición en los análisis MDS y AFC. En un segundo grupo, con 4 haplogrupos afectados, se encuentran la población romaní griega, también con haplogrupos minoritarios, la eslovaca y la ucraniana. En estas dos últimas se ha visto afectado el haplogrupo I, que presenta unas frecuencias notables. Además, en la población ucraniana se ha visto afectado el haplogrupo J, que es mayoritario en esta población, como se ha mencionado a lo largo de la exposición de resultados. Por último, como puede apreciarse en la Tabla 9 del Anexo, estas romaníes son la población con menos haplogrupos presentes, tan sólo 4. Es decir, la deriva ha promovido un haplogrupo hacia una frecuencia alta, pero además, ha provocado la pérdida de otros haplogrupos, disminuyendo considerablemente la diversidad genética de los linajes de ADN mitocondrial de esta población. Entre el resto de poblaciones, las romaníes húngaras han mostrado evidencias de deriva en 3 haplogrupos y las españolas y búlgaras en 1.

Resultados

Tabla 30: Valores de $-\log(P)$ para el caso hipotético sin migración, obtenidos a partir de las frecuencias de haplogrupos del ADN mitocondrial en 7 poblaciones de romaníes europeos. Se ha anotado un guion cuando no se ha podido calcular el cociente de F_{ST} por existir una división con 0 en el denominador.

Haplogrupo	Romaní-búlgara	Romaní-griega	Romaní-húngara	Romaní-rumana	Romaní-eslovaca	Romaní-española	Romaní-ucraniana
C	-	-	0,433	-	-	-	-
D	-	0,814	-	0,865	0,866	-	-
F1B	0,434	0,434	0,434	0,434	0,434	0,434	0,434
H*	0,365	0,434	0,427	0,382	0,370	0,496	0,645
HV	0,401	0,996	0,424	0,938	0,808	0,940	2,324*
I	0,471	1,067	3,704*	15,415*	21,197*	2,515*	6,789*
J	-	0,403	134,945*	0,678	0,480	0,689	164,320*
K	0,378	0,924	0,638	1,105	0,909	0,391	0,902
L2,6	-	-	-	-	-	0,864	-
L2a	-	-	-	-	-	-	-
M*	0,230	0,314	0,225	0,284	0,302	0,241	0,434
N*	-	0,865	-	3,642*	-	-	-
N1	0,226	-	0,434	34,443*	0,859	0,434	0,434
R*	0,504	0,434	0,650	0,434	0,493	0,369	0,434
R0a	0,862	0,862	-	-	-	-	-
R10	0,434	0,434	0,434	0,434	0,434	0,434	0,434
T	0,751	17,918*	0,467	1,223	1,662*	0,778	1,022
U	0,499	0,865	-	0,865	0,441	0,432	0,861
U1	23,174*	9,585*	0,434	100,437*	0,434	0,434	0,434
U2	0,402	0,434	0,434	0,709	0,292	0,564	0,434
U3	-	0,499	0,434	-	0,864	-	0,853
U4	0,412	0,853	0,431	0,865	0,538	0,857	0,838
U5	0,690	1,453*	0,376	0,799	0,385	0,626	0,369
U6	-	0,865	-	-	-	0,784	-
U7	0,434	0,434	0,434	0,575	0,434	0,434	0,434
V	0,855	0,853	0,433	0,861	0,857	0,784	0,861
W	0,674	1,084	7,130*	20,331*	1,020	0,434	187,180*
X	0,604	-	-	-	0,861	-	0,861

Tabla 31: Valores de $-\log(P)$ para el caso hipotético con migración, obtenidos a partir de las frecuencias de haplogrupos del ADN mitocondrial en 7 poblaciones de romaníes europeos. Se ha anotado un guion cuando no se ha podido calcular el cociente de F_{ST} , por existir una división con 0 en el denominador.

Haplogrupo	Romaní-búlgara	Romaní-griega	Romaní-húngara	Romaní-rumana	Romaní-eslovaca	Romaní-española	Romaní-ucraniana
C	-	-	0,434	-	-	-	-
D	-	0,815	-	0,866	0,867	-	-
F1B	0,865	0,865	0,865	0,865	0,865	0,865	0,865
H*	0,431	0,505	0,496	0,445	0,432	0,586	0,751
HV	0,455	1,090	0,473	1,013	0,891	1,136	2,809*
I	0,518	1,181	4,314*	18,912*	42,345*	3,052*	13,561*
J	-	0,465	177,448*	0,780	0,544	0,804	205,699*
K	0,431	0,984	0,670	1,255	0,962	0,426	0,953
L2,6	-	-	-	-	-	0,866	-
L2a	-	-	-	-	-	-	-
M*	0,418	0,570	0,420	0,514	0,562	0,423	0,809
N*	-	0,867	-	3,649*	-	-	-
N1	0,450	-	0,865	44,904*	1,393*	0,865	0,865
R*	0,898	0,854	1,105	0,854	0,916	0,727	0,854
R0a	0,865	0,865	-	-	-	-	-
R10	0,865	0,865	0,865	0,865	0,865	0,865	0,865
T	0,852	23,370*	0,516	1,501*	1,990*	0,873	1,142
U	0,501	0,867	-	0,866	0,444	0,436	0,864
U1	31,770*	13,354*	0,866	135,703*	0,866	0,866	0,866
U2	0,790	0,853	0,853	1,179	0,496	0,997	0,853
U3	-	0,502	0,435	-	0,866	-	0,861
U4	0,414	0,860	0,434	0,866	0,540	0,862	0,853
U5	0,782	1,698*	0,430	0,891	0,443	0,742	0,429
U6	-	0,867	-	-	-	0,786	-
U7	0,855	0,855	0,855	1,012	0,855	0,855	0,855
V	0,862	0,860	0,434	0,865	0,862	0,786	0,864
W	1,142	1,692*	9,061*	27,633*	1,604*	0,860	251,887*
X	0,607	-	-	-	0,864	-	0,864

Estimación del grado de mestizaje

Para el cálculo del mestizaje se han eliminado, como en apartados anteriores, los haplogrupos que han presentado evidencias notables de acción de la deriva. Los resultados obtenidos se muestran en la Tabla 32. Por un lado, con un componente europeo mayoritario, se encuentran las poblaciones romaníes de Eslovaquia (73,5%), Hungría (63,6%), Ucrania (61,5%) y Bulgaria (56,3%). En la otra vertiente están las poblaciones de Rumanía (37%), España (29,9%) y Grecia (10,7%), con componentes minoritarios en la fracción femenina de sus respectivas poblaciones huésped, siendo el caso más notable el de Grecia.

Tabla 32: Estimación del grado de mestizaje (\pm error estándar) e intervalo de confianza del 95% en 7 poblaciones romaníes europeas, obtenidas a partir de frecuencias de haplogrupos de ADN mitocondrial.

Población romaní	Componentes	Coefficientes	Error estándar	Límite inferior	Límite superior
Bulgaria	Punjab	0,437	0,088	0,257	0,618
	Búlgara	0,563	0,071	0,417	0,708
Grecia	Punjab	0,893	0,138	0,608	1,178
	Griega	0,107	0,136	-0,175	0,388
Hungría	Punjab	0,364	0,097	0,164	0,564
	Húngara	0,636	0,104	0,420	0,851
Rumanía	Punjab	0,630	0,146	0,327	0,934
	Rumana	0,370	0,144	0,071	0,669
Eslovaquia	Punjab	0,265	0,056	0,149	0,381
	Eslovaca	0,735	0,052	0,627	0,843
España	Punjab	0,701	0,222	0,239	1,163
	Española	0,299	0,177	-0,069	0,667
Ucrania	Punjab	0,385	0,433	-0,518	1,288
	Ucraniana	0,615	0,426	-0,274	1,503

En la Figura 95 se muestra una comparación entre el porcentaje de cada población romaní respecto al tamaño poblacional de sus respectivas poblaciones huésped, frente al porcentaje de mestizaje de cada grupo de etnia romaní. Los valores obtenidos se muestran en la Tabla 10 del Anexo. En primer lugar, no se observa una relación entre ambas variables, lo que se ha confirmado también mediante los coeficientes de correlación de Pearson y Spearman (Tabla 33). Esta ausencia de correlación coincide además con los resultados obtenidos con los haplogrupos del cromosoma Y (Tabla 28) y, en ambos casos, contrastan con la correlación positiva obtenida al comparar estas variables mediante STRs autosómicos (Tabla 22). Finalmente, en la mayoría de casos, los grados de mestizaje obtenidos mediante el ADNmt son muy superiores respecto a los observados por medio del cromosoma Y (Tabla 26).

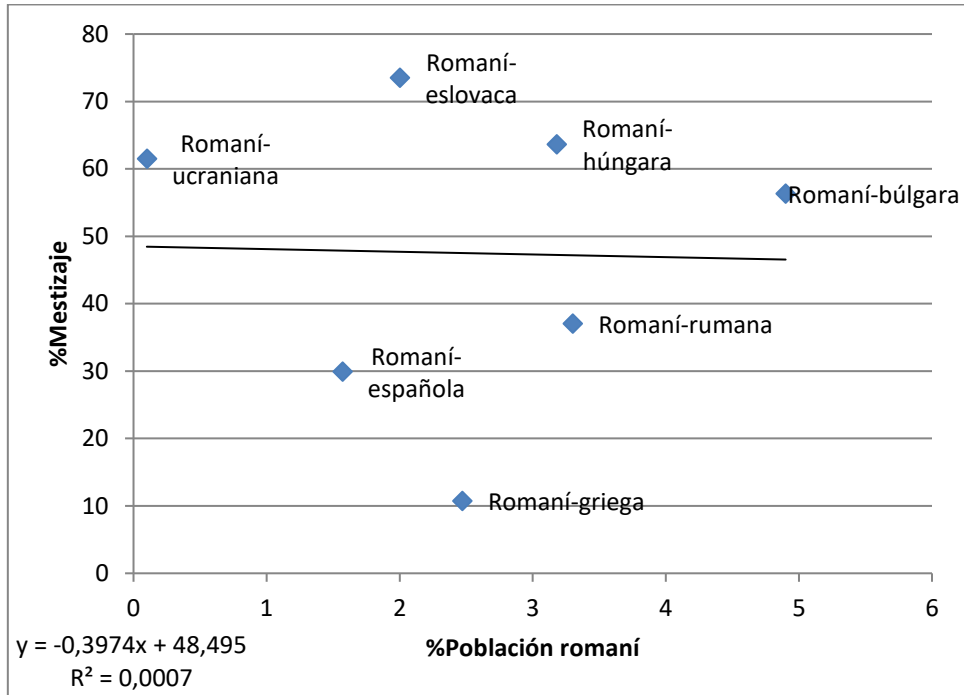


Figura 95: Recta de regresión al comparar el porcentaje que representa la población romaní en la población huésped frente al porcentaje de mestizaje para los haplogrupos de ADNmt.

Tabla 33: Matriz de coeficientes de correlación de Pearson y Spearman entre el porcentaje que representa la población romaní en la población huésped frente al porcentaje de mestizaje para los haplogrupos de ADNmt.

r Pearson			Spearman		
	% Población total	%Mestizaje ADNmt		% Población total	%Mestizaje ADNmt
% Población total	-	0,95423	% Población total	-	0,90635
%Mestizaje ADNmt	-0,026972	-	%Mestizaje ADNmt	-0,071429	-



Discusión

VI. Discusión

Como se ha señalado a lo largo de la presente memoria, agentes evolutivos como la deriva genética y el flujo génico, así como sus interacciones con otros factores complejos, como el contexto sociocultural y geopolítico, tienen un efecto fundamental en los cambios microevolutivos de las poblaciones (Arnold et al., 2001; Relethford, 2012). Estas interacciones pueden verse afectadas no sólo por el contexto actual de una población concreta, sino por los procesos históricos experimentados con anterioridad como, por ejemplo, eventos migratorios y fundadores que pueden suponer un filtro en la diversidad genética de una determinada población (Peter y Slatkin, 2015). Asimismo, otros factores demográficos que pueden verse afectados por los agentes microevolutivos mencionados serían el tamaño poblacional efectivo (N_e) o el equilibrio entre los sexos. A partir del asentamiento de un grupo de población en una región diferente de la original, su propio contexto sociocultural puede actuar como una barrera a la panmixia, limitando el flujo génico con las poblaciones hospedadoras o receptoras (Flores-Bello et al., 2021). Bajo esta situación, la posible divergencia entre una población respecto a su población ancestral, dependerá tanto del cese del intercambio genético entre ambas (y de la posible incorporación de genes de la población receptora), como del grado de deriva experimentado a lo largo de su historia, que será más notable en poblaciones con tamaños efectivos reducidos (Lynch et al., 2016; Wang et al., 2020). Por ello, a la hora de analizar la heterogeneidad genética de una población, y para poder interpretar adecuadamente los diferentes factores y procesos a los que ha sido sometida, es necesario abordar su estudio mediante un enfoque multidisciplinar. En esta investigación se han tratado de estimar los efectos de la deriva y el flujo génico sobre la diversidad genética de una muestra de población de etnia romaní residente en el País Vasco, considerando que, como agentes microevolutivos, han podido realmente actuar sobre esta población debido a sus particulares características socioculturales e históricas.

Como se expuso en la Introducción, los orígenes de la etnia romaní se han establecido en el subcontinente indio en base a análisis lingüísticos, antropológicos y genéticos (Chaix et al., 2004; Fraser, 1992; Gresham et al., 2001; Hancock, 2000; Kalaydjieva et al., 2001b; Marushiakova y Popov, 1997; Morar et al., 2004, 2005; Melegh et al., 2017). Dentro de este subcontinente su localización ha sido cuestión de debate; por un lado, estudios lingüísticos defienden que su población fundadora se encontraría en el noreste de la India (Fraser et al., 1992; Hancock et al., 2002; Iovita et al., 2004), mientras que los estudios genéticos, tanto con marcadores clásicos como con marcadores de linaje, sugieren un origen en el Noroeste del

subcontinente, debido concretamente a su similitud con poblaciones de los estados de Rajasthan y Punjab (Mastana et al., 1992; Mendizábal et al., 2011). En lo que respecta a la estructura social interna de esta etnia presenta una gran similitud con la cultura india, con un marcado sistema de castas y comportamientos endogámicos, donde los enlaces matrimoniales con poblaciones externas son muy limitados y se promueven las uniones intra-poblacionales (Melegh et al., 2017). Sin embargo, diversos estudios han encontrado pruebas de un considerable flujo génico entre las diferentes poblaciones romaníes europeas y sus respectivas poblaciones huésped (Bianco et al., 2020; Gusmão et al., 2010; Martínez-Cruz et al., 2016). Por esta “dualidad” sobre los factores implicados en el patrimonio genético de los romaníes (aislamiento y/o flujo génico) nos hemos planteado cuál habría sido el agente microevolutivo con mayor impacto en dicha población, en particular en la residente en el País Vasco. Es decir, si su alto componente endogámico habría mantenido su acervo ancestral indio sujeto únicamente a los cambios propios de la deriva genética, o si el flujo génico desde las poblaciones euro-asiáticas, durante y posterior a su diáspora, habría configurado su patrimonio genético actual.

Dado el gran número de datos manejados en cuanto a marcadores genéticos, poblaciones consideradas tanto a nivel analítico como comparativo, y los diversos análisis realizados, la discusión se ha estructurado siguiendo un orden concordante con los diferentes resultados obtenidos y descritos en el apartado correspondiente. Hay también que señalar que, en el apartado de Resultados (IV) se han discutido algunas cuestiones referidas a los análisis realizados, es decir, nos hemos extendido en algunas explicaciones de tipo “técnico” según los distintos procedimientos utilizados, así como en la explicación de las distintas figuras, con el fin de poder abordar la discusión final de forma más precisa y concreta.

VI.1 SNPs de ascendencia (AIMs)

Para abordar el estudio de los procesos de microdiferenciación experimentados por la población romaní del País Vasco se han empleado en primer lugar 48 SNPs altamente informativos de ascendencia (AIMs). Su elección se justifica tanto por su gran disponibilidad, ya que son la forma más abundante de variación genética humana, como por sus bajas tasas de mutación, lo que les dota de cierta utilidad para analizar eventos ancestrales (Fan et al., 2007; Hu et al., 2017). Los resultados obtenidos indican que los SNPs seleccionados presentan carácter polimórfico, como es propio de un marcador AIMs (Tabla 6). Cabe destacar que el SNP rs200789833, ha mostrado significación estadística en el test de equilibrio Hardy-Weinberg, indicando un exceso de heterocigotos, lo que podría estar relacionado con su localización

cromosómica. Dicho SNP se encuentra en el brazo corto del cromosoma 6 próximo al gen HLA-DPA1 (Lantermann et al., 2002) perteneciente al complejo mayor de histocompatibilidad, por lo que la región (incluyendo el SNP), podría encontrarse bajo cierto grado de selección.

El análisis de AMOVA realizado para determinar la existencia de diferencias entre los grupos continentales de Europa y Sur de Asia ha detectado que el 95,8% de los SNPs analizados mostraban diferencias significativas en el índice F_{ST} , el 85,41% mostraron significación estadística para el F_{SC} y el 62,5% para el F_{CT} . Estos resultados indican una gran heterogeneidad tanto entre los individuos dentro de una población, como entre poblaciones dentro de un grupo, y entre los grupos analizados. No obstante, a pesar de haber considerado SNPs con características AIMs, en una parte de ellos no se ha detectado heterogeneidad significativa entre los grupos poblacionales. Esto podría deberse a que los grupos asiáticos seleccionados no corresponden a todo el continente, sino a una región más delimitada, como el Sur de Asia, restando capacidad diferenciadora a los SNPs que se han seleccionado (Tabla 7).

Posteriormente, al comparar las frecuencias de la población romaní del País Vasco frente al rango de frecuencias de los dos grupos continentales (Europa y Sur de Asia) (Figura 41) se han podido observar los primeros indicios del efecto de la deriva genética sobre dicha población. Si se considera que el origen de las poblaciones romaníes, incluida la del País Vasco, se encuentra en el Sur de Asia, y que desde que abandonaron ese territorio los romaníes han convivido la mayor parte del tiempo con poblaciones europeas, el gran número de SNPs estudiados que han mostrado valores extremos parece indicar una importante incidencia de la deriva, con un 23% situados fuera de los rangos continentales. Estos resultados se confirman con los valores obtenidos de $-\log(P)$, para el cociente entre el valor de F_{ST} observado y el esperado, considerando, por un lado las frecuencias en los romaníes iguales a las frecuencias en el noroeste de la India (F_{ST} min) y las frecuencias en los romaníes intermedias entre España y el noroeste de la India por otro (F_{ST} med). En el primer caso se omite la posibilidad de flujo génico y en el segundo se considera que la población ha sufrido un 50% de flujo. Mediante este método se ha identificado la acción de la deriva genética entre el 46 y el 48% de los SNPs, lo que sugiere un efecto aún más notable de este agente microevolutivo. Además, los valores observados de $-\log(P)$ han sido muy heterogéneos, tal y como era de esperar por acción, precisamente, de la deriva (Tabla 9). Asimismo, los resultados obtenidos han mostrado una distribución exponencial negativa (Figura 54), lo que supone en términos generales una demostración empírica del efecto de la deriva en las frecuencias alélicas de un marcador dado a lo largo de las generaciones. Este proceso implica, al cabo de un número infinito de generaciones, la fijación de un alelo y la

desaparición del resto, correspondiendo a las fases intermedias una distribución exponencial negativa de las frecuencias (Fisher, 1930; Wright, 1931), tal y como se ha observado en la presente investigación. Esta situación sería extrapolable a cualquier población que sufre un alto efecto de la deriva y sin influencia de otros agentes microevolutivos como el flujo génico.

Se han realizado análisis MDS y AFC con el objetivo de conocer el resultado del posible flujo génico experimentado por la población romaní estudiada de manera conjunta al efecto de la deriva. Los resultados de análisis MDS muestran la posición intermedia de la población romaní del País Vasco respecto a los dos grupos continentales analizados, como cabría esperar debido al efecto de un cierto grado de flujo génico (Figura 42). Además, se puede observar la disposición particular de ambos grupos continentales; por un lado el europeo, con posiciones relativas próximas entre sí, y por otro el de Sur de Asia, mucho más disperso. Con el objetivo de comparar dicha variabilidad e identificar la posición de los romaníes respecto a ella, se ha realizado el mismo análisis añadiendo una elipse equiprobable del 95 % (Figura 43). Como era de esperar, la elipse europea es mucho más reducida que la que comprende a las poblaciones del Sur de Asia y, en lo que respecta a los romaníes, estos se mantienen en una posición intermedia próximos al área de solapamiento de ambas elipses. Con el fin de contrastar los resultados anteriores, obtenidos mediante una matriz de distancias F_{ST} de Reynolds, se ha realizado un análisis MDS a partir de la matriz de distancias R de Harpending y Jenkins (Figura 44). Los resultados muestran ciertos cambios en la disposición de algunas poblaciones surasiáticas, pero en lo que atañe a la población romaní esta se sigue manteniendo en una posición relativa muy similar a la observada en la Figura 42. En los distintos análisis realizados, los romaníes del País Vasco se han mantenido en valores intermedios, lo que indica un cierto grado de flujo génico, pero hay que destacar que su posición relativa no se encuentra exactamente entre ambos grupos continentales, lo que cabía esperar si el flujo génico fuera el único agente microevolutivo en juego. Este resultado es, por tanto, un claro indicio de la acción conjunta del flujo génico junto con la deriva genética experimentada por esta población.

El análisis del centroide (Figura 50) ha revelado que las poblaciones asiáticas más cercanas a la población romaní del País Vasco, brahui y balochi (según los análisis MDS, Figuras 46 y 49), muestran unas posiciones claramente diferenciadas entre sí. Por un lado, los balochi muestran una alta diversidad, debida posiblemente a un flujo génico intenso con poblaciones próximas. Hay que recordar que se trata de un grupo étnico de gran tamaño, con más de 20 millones de individuos distribuidos por varios países de la región, que incluyen Afganistán, Irán, Pakistán y la India. Su tamaño y su posición geográfica central han facilitado sin duda los

intercambios genéticos con otros grupos de esta amplia región. Los brahui también muestran una gran diversidad, pero el análisis del centroide sugiere que han experimentado un flujo génico externo en relación a los grupos poblacionales analizados. Este grupo habita principalmente en el Baluchistán, en un entorno montañoso, rodeado de poblaciones de la etnia balochi pero, a diferencia de ellos, su lengua es dravídica. Su origen es controvertido, pudiendo ser un grupo residual de los primitivos pobladores dravidianos de la región, o bien provenir del centro de la India con posterioridad al año 1000 (Sergeant, 1997; Elfenbein, 1987). En cualquier caso, a pesar de su innegable similitud genética con los balochi, debido sin duda a su proximidad geográfica, los brahui mantienen un mayor grado de diversidad genética de lo esperado. El resultado obtenido es plausible debido a su localización geográfica, puesto que se encuentran en un paso natural entre Asia central y el subcontinente indio.

Del análisis del centroide (Figura 50) también cabe destacar la posición de los romaníes del País Vasco, que revela un grado de aislamiento promedio en el grupo de poblaciones considerado. Esta población muestra una diversidad media en el contexto de las poblaciones analizadas, de modo que, aparentemente, ha experimentado un nivel medio de flujo génico. En cuanto a la proximidad genética relativa observada entre los romaníes del País Vasco con los brahui y balochis, estas poblaciones asiáticas se encuentran en la encrucijada que conecta Asia Central, el subcontinente indio y la península arábiga, por lo que se sitúan en la zona de paso natural de la población romaní durante su diáspora desde el Noroeste de la India. Por tanto, es posible que se haya producido un cierto grado de flujo génico entre todas ellas. En este sentido, Bianco et al. (2020) analizaron la secuencia genómica completa de individuos de etnia romaní pertenecientes a 6 grupos poblacionales residentes en países europeos, y encontraron evidencias de flujo génico entre ellos y poblaciones de Armenia y de la península de Anatolia, que también están situadas en la ruta más probable de la diáspora de la etnia romaní. Igualmente, Font-Porterías et al. (2019) encontraron evidencias significativas de componente poblacional del Próximo Oriente (Irán, Siria, Jordania y Palestina) en el acervo genético de varias poblaciones romaní europeas, entre ellas la española. En todo caso, siendo los brahuis y balochis las poblaciones del Sur de Asia más próximas genéticamente a las poblaciones europeas en la base de datos utilizada, y teniendo en cuenta la posición intermedia de los romaníes, debido a su patrimonio genético mixto entre las poblaciones europeas y las del Sur de Asia, parece lógico que los resultados obtenidos a partir de los análisis MDS y AFC indiquen cierta similitud en las posiciones de romaníes, brahuis y balochis.

En el análisis de clinas (Tabla 8 y Figura 51), 17 marcadores de los 25 para los que se dispuso de un número suficiente de poblaciones han mostrado clinas significativas. Estos resultados serían similares a los obtenidos en los análisis de AMOVA (Tabla 7), donde no todos los marcadores alcanzaron significación estadística, a pesar de haber sido seleccionados por sus características AIMs. En principio, la existencia de heterogeneidad significativa entre los grupos de poblaciones europeas y asiáticas debería facilitar la existencia de clinas tal y como se ha observado en esta investigación; cabe destacar que la mayor parte presenta una orientación nor-noroeste sur-sudeste. Estos resultados se podrían explicar debido a que, por un lado, el continente europeo muestra un gradiente latitudinal de las frecuencias alélicas en un gran número de marcadores, debido a las variaciones geográficas de las proporciones de origen del Neolítico y de la Estepa Póntica en su patrimonio genético (Haak et al., 2015). Por otro lado, dicha disposición se mantendría también en el subcontinente indio debido a la distribución de los componentes indoeuropeo y dravídico. Además, las poblaciones europeas se encuentran más al norte que las del Sur de Asia, por lo que todos estos factores justifican la orientación de las clinas para los marcadores utilizados.

Antes de calcular el grado de mestizaje de la población romaní del País Vasco, se estimó el efecto de la deriva sobre los diferentes SNPs. Cabe esperar que una acción significativa de la deriva alteraría las frecuencias alélicas, de forma que no serían útiles para realizar una correcta valoración del flujo génico entre las poblaciones implicadas. Así, se ha observado un valor significativo de $-\log(P)$ en 21 o 22 SNPs (dependiendo del método utilizado), lo que implica entre un 46% a un 48% de los todos los SNPs analizados (Tabla 9 y Figura 54). En la mayor parte de los casos, las frecuencias alélicas de los romaníes se han situado fuera del rango de las poblaciones parentales, por lo que estos SNPs actuarían necesariamente como una fuente de error en el cálculo del mestizaje. Una vez desechados aquellos marcadores con un efecto significativo de la deriva, la población romaní del País Vasco ha mantenido un 47,3% de su patrimonio genético ancestral. En consecuencia, el flujo génico experimentado desde el inicio de su diáspora se ha estimado en un 52,7% (Tabla 10).

En resumen, los resultados indican que, pese a mantener cierto aislamiento, la población romaní analizada ha experimentado flujo génico desde su población huésped. A su vez, cierto grado de aislamiento y un tamaño efectivo poblacional reducido han propiciado cierta actuación de la deriva. Estos resultados son similares a los publicados por Bianco et al. (2020), donde observan que, en promedio, más del 50% del patrimonio genético de varias poblaciones romaníes europeas proviene de las poblaciones europeas huésped. Por otro lado, estos autores

también han encontrado evidencias de aislamiento durante las primeras fases de la diáspora y una gran reducción del tamaño efectivo de la población en dichas fases, lo que ha propiciado la acción de la deriva genética.

VI.2. Inserciones *Alu*

Con objeto de complementar los resultados obtenidos acerca de los procesos de microdiferenciación experimentados por la población romaní residente en el País Vasco a partir de un conjunto de SNPs, se ha realizado una segunda serie de análisis empleando 8 inserciones *Alu*. Estas inserciones son altamente estables, de modo que la posibilidad de mutación hacia una nueva copia de la inserción o una delección es mínima y, como consecuencia, si dos individuos presentan una misma inserción estarán necesariamente emparentados (Batzer et al., 1994; Batzer et al., 2002; Salem et al., 2005), como ya se ha señalado en apartados anteriores. A su vez, estos marcadores tienen la inusual característica de que se conoce su estado ancestral, que se corresponde con la no inserción, lo que permite añadir una hipotética población ancestral con frecuencias de inserción 0 en los análisis estadísticos (Novick et al., 1998; Salem et al., 2005). Además son marcadores selectivamente neutros (Comas et al., 2000), por lo que las frecuencias de sus polimorfismos dependerán exclusivamente del flujo génico y la deriva, las dos presiones microevolutivas analizadas en la presente memoria.

En relación a las frecuencias observadas (Tabla 11), el hecho más destacable es que la inserción ACE se encuentra en la muestra romaní fuera del rango registrado en Europa y el Sur de Asia, lo que puede considerarse un claro indicio del efecto de la deriva sobre dicha inserción en esta población (Figura 55). La comparación mediante un AMOVA de las frecuencias entre los dos grupos continentales que se vienen manejando en este trabajo ha revelado diferencias significativas en los 3 niveles de organización considerados, indicando una alta heterogeneidad (Tabla 12) y revelan la idoneidad de este grupo de inserciones *Alu* para el análisis comparativo del patrimonio genético de los romaníes con el de europeos y asiáticos.

Posteriormente, al igual que con los SNPs, se han realizado varios análisis MDS y AFC. En los MDS se ha observado que la población romaní del País Vasco se encuentra distanciada tanto del grupo europeo como del asiático del Sur (Figuras 56, 58-60), y relativamente próxima a las poblaciones de Próximo Oriente. Estos resultados indican, en primer lugar, un claro efecto de la deriva genética. Por otra parte, podemos afirmar que su posición, cercana a las poblaciones de Próximo Oriente, podría ser circunstancial, debido a que se observa en el análisis MDS, pero no en el AFC, donde los romaníes se encuentran más próximos a algunas poblaciones europeas (Figura 61). Esto nos hace pensar que las poblaciones de Próximo Oriente, debido a su posición

geográfica, presentan unas características intermedias entre el Sur de Asia y Europa, mientras que los romaníes portan un patrimonio genético híbrido debido al flujo génico con poblaciones europeas a lo largo de su tránsito migratorio. Puesto que el origen del mestizaje en las poblaciones de Próximo Oriente y en los romaníes es diferente, su cercanía en el análisis MDS es, probablemente, un tanto forzada. En todo caso, no puede descartarse la posibilidad de un cierto parentesco genético entre ambos grupos poblacionales, debido a los resultados encontrados por Bianco et al. (2020), donde se describen evidencias de flujo génico desde Armenia y la Península de Anatolia hacia la población romaní europea, así como en el trabajo de Font-Porterías et al. (2019) donde se constata esta misma evidencia en la población romaní española.

Al igual que con los análisis de SNPs, los grupos continentales han mostrado diferente grado de variabilidad interpoblacional, como queda patente en la elipse equiprobable del 95% realizada junto con los análisis MDS (Figura 57). Así, se ha observado que la variabilidad del grupo asiático es muy superior a la del europeo hasta el punto incluir la elipse de este último grupo. La gran variabilidad genética existente en el continente asiático ha sido muy analizada, aunque en menor medida que en el continente europeo (Cavalli-Sforza y Piazza, 1997; Der Sarkissian et al., 2013; Navarro-López et al., 2021), con estudios como el de Pan y Xu (2020) o el de Stoneking y Delfin (2010), que ponen su foco en la alta variabilidad de este continente y sus posibles causas.

Con el fin de profundizar en los procesos microevolutivos experimentados por la población romaní del País Vasco, se realizó un análisis del centroide, en el que, dicha población se situó en una posición central (Figura 65), con valores medios de variabilidad y heterogeneidad; este resultado indica que, posiblemente, ha experimentado flujo génico desde ambos grupos continentales. Dado el reducido número de poblaciones representativas del Sur de Asia incluidas en estos análisis, se decidió reducir el número de inserciones analizadas para poder aumentar el número de poblaciones asiáticas, por lo que hubo que descartar la inserción A25; las poblaciones añadidas para los análisis fueron, en su mayor parte, procedentes de la India. Con el grupo ampliado de poblaciones se repitieron los análisis multivariantes. En el MDS (Figura 66) se puede observar una disposición de las poblaciones similar a la encontrada en los análisis realizados con 8 inserciones (Figuras 56-60), siendo de nuevo las poblaciones de Emiratos Árabes y Siria las más próximas a la población romaní del País Vasco, que, a su vez, se encuentra muy alejada de las poblaciones indias y europeas. El análisis AFC muestra una mayor dispersión de las poblaciones del Norte de la India y un desplazamiento de la población de

Emiratos Árabes, que sea próxima a poblaciones de la India y se aleja de la Siria. La población romaní se encuentra de nuevo alejada de ambos grupos continentales. Estos resultados pueden considerarse indicios de una importante acción de la deriva sobre el patrimonio genético romaní.

Para profundizar en las relaciones de parentesco genético de los romaní del País Vasco con otras poblaciones, se realizó un dendrograma basado en las inserciones *Alu*, dado que éstas permiten reconstruir una población ancestral; ésta se incluyó como una población con frecuencia 0 para todas las inserciones (Figura 71). También se añadió para este análisis una población africana (iKung) y otra asiática (Han), con el fin de tener dos poblaciones control de África y del Este de Asia. Se observó una disposición lógica, con la población ancestral en la raíz, la población iKung próxima a ella y la Han próxima a las poblaciones indias; entre ellas se encontraron las poblaciones europeas. Esta disposición coincide con lo esperado de acuerdo a nuestros conocimientos actuales sobre el proceso de colonización de los diferentes continentes. La población romaní se encuentra junto al grupo del Próximo Oriente, lo que concuerda con los resultados de los MDS previos, en una posición sobre la que puede haber influido tanto el flujo génico como la deriva. Por último, la población pakistaní se encuentra en el grupo del Próximo Oriente, en una posición inusual, siendo una posible explicación que la procedencia de la muestra es de trabajadores pakistaníes que residen en Dubái (Stoneking et al., 1997), y posiblemente no son la representación más óptima de dicha población. El análisis de clinas ha revelado 5 clinas con una orientación nor-noroeste sur-sudeste (Figura 68), de manera similar a los resultados obtenidos mediante los SNPs (Figura 51). En el análisis pormenorizado de cada clina se ha observado un cierto solapamiento en las frecuencias de ambos continentes, confirmándose, asimismo, que los romaníes analizados se sitúan fuera del rango de las poblaciones asiáticas y europeas para la inserción ACE (Figura 69).

Se ha abordado también el análisis de la acción potencial de la deriva mediante $-\log(P)$. En este análisis se han observado valores significativos para 3 inserciones, ACE, TPA25 y FXIIB, lo que implicaría una acción perceptible de la deriva en estos marcadores (Tabla 14). En la Figura 70, se muestra la disposición de los valores $-\log(P)$ una vez ordenados y, pese a su bajo número, mantienen la forma de una curva exponencial negativa, lo que coincide con lo observado en los SNPs (Figura 54) y, como ya se explicó anteriormente, esta distribución se corresponde con la esperada por efecto de la deriva genética (Fisher, 1930; Wright, 1931).

Por último, la estimación de mestizaje obtenida a partir de las inserciones *Alu* sin evidencia de deriva ha mostrado valores similares entre los componentes ibérico y del norte de

la India (tabla 15). La población romaní del País Vasco ha mostrado un equilibrio entre los valores de ambos componentes continentales, no muy distintos a los observados con el análisis de SNPs, aunque algo inferiores: un 41,6% de componente europeo con inserciones *Alu* frente al 58,4% con SNPs. Este elevado componente europeo en la población romaní estudiada es similar a estudios realizados en otras poblaciones romaníes europeas (Bianco et al., 2020; Gusmão et al., 2010). Los resultados obtenidos revelan que la población romaní analizada ha sufrido un considerable grado de deriva a lo largo de su periplo por el Oeste de Eurasia, presumiblemente por una combinación de efectos fundador y cuello de botella, aunque también se han encontrado evidencias de flujo génico con poblaciones europeas y eventualmente del Próximo Oriente.

VI.3 CODIS

Los SNPs y las inserciones *Alu* son dos tipos de marcadores con una serie de características interesantes, entre las que destacan su carácter bialélico y su baja tasa de mutación, como se ha señalado a lo largo de este texto. Por otra parte, los marcadores que se han analizado y discutido previamente han sido, todos ellos, autosómicos. Con el objetivo de analizar al grupo romaní del País Vasco desde otros puntos de vista, y al mismo tiempo contrastar los resultados obtenidos con los de otras poblaciones europeas de origen romaní, se han recopilado las frecuencias de otros tipos de marcadores que incluyeran también a otros grupos de esta etnia. Se ha comenzado con una base de datos que contenía los 13 STRs autosómicos pertenecientes al sistema CODIS ya que, debido a sus altas tasas de mutación, son buenos candidatos para analizar eventos poblacionales recientes (Fan et al., 2007) y, de hecho, se usan de forma habitual como marcadores en los estudios de genética de poblaciones (Aguilar-Velázquez et al., 2016; Krzeminska-Ahmadzai et al., 2021; Soták et al., 2008, 2011; Wu et al., 2020; Yao y Wang, 2016)

Se abordó inicialmente el análisis de los STRs mediante un AMOVA para determinar si existían diferencias significativas para estos marcadores entre los dos grupos continentales de Europa y Sur de Asia. Los resultados mostraron significación en los 3 niveles de organización para todos los marcadores, con la excepción de la variación entre grupos, donde 4 de los 13 STRs no fueron significativos (Tabla 16). Esto debería dificultar, en principio, la eficacia de la comparación de las poblaciones romaníes con europeos y asiáticos. Sin embargo, por ser marcadores multialélicos, y aunque el conjunto del marcador no resulte útil, si pueden serlo algunos de sus alelos. Así, en el análisis de clinas (Tabla 17), se han observado clinas significativas en 3 de estos 4 STRs, incluyendo 2 alelos de CSF1PO, 2 de D13S317 y 2 de D5S818, lo que implica

un cierto patrón espacial de los mismos. Hay que señalar que el carácter multialélico de estos marcadores, en los que diferentes alelos pueden verse afectados por diferentes procesos, puede influir en el resultado global de los AMOVA.

El conjunto del análisis de detección de clinas para los *loci* del sistema CODIS en las poblaciones de Europa y Sur de Asia (Tabla 17 y Figura 77) mostró 43 clinas significativas, de las que 11 superaron la corrección de Bonferroni. La mayor parte de estas clinas han sido esencialmente latitudinales, en sentido nor-noroeste sur-sudeste, de forma similar a los resultados obtenidos para SNPs (Tabla 8 y Figura 51) e inserciones *Alu* (Tabla 13 y Figura 68). Analizando los resultados de los MDS y AFC, se ha observado que el grupo formado por las poblaciones romaníes europeas tiende a aparecer separado de los grupos continentales europeo y asiático. Así, dichas poblaciones se encuentran desplazadas en los MDS continentales, tanto del grupo europeo (Figura 73) como del asiático (Figura 74). En el análisis conjunto (Figuras 72 y 75), los romaníes aparecen en su mayor parte separados de ambos grupos continentales, con la excepción de los romaníes de Macedonia, que se encuentran relativamente próximos a los asiáticos, y los de Hungría, que se muestran más relacionados con los europeos. Esta posición relativa podría ser debida al efecto de la deriva genética en la etnia romaní desde el inicio de su diáspora desde la India (Gresham et al., 2001; Martínez-Cruz et al., 2016).

En lo que respecta a las posiciones relativas dentro del grupo de poblaciones romaníes considerado, los croatas muestran las posiciones más extremas, siendo el grupo más heterogéneo de las 5 poblaciones romaníes analizadas (Figuras 72-75). Esta mayor heterogeneidad podría deberse a un mayor efecto de la deriva, causado eventualmente por un menor tamaño efectivo y un mayor aislamiento (Lynch et al., 2016; Wang et al., 2020). Los posteriores análisis realizados han confirmado esta hipótesis, ya que el análisis del centroide revela para los romaníes croatas el grado más alto de aislamiento en el conjunto de poblaciones de esta etnia (Figura 76). La muestra de romaníes portugueses, que también se ha revelado en el análisis centroide con un alto grado de aislamiento, aparece claramente separada de los grupos europeo y asiático en los MDS y AFC. La población romaní croata también presenta el mayor número de alelos con significación ($n=53$) entre las poblaciones analizadas para la prueba de $-\log(P)$, que son además los mismos que en la población romaní portuguesa (Tabla 18). La curva de distribución de valores del cociente de F_{ST} en ambas poblaciones se encuentra por encima del resto, mostrando la población romaní croata los mayores valores en el rango de frecuencias alélicas más altas y la portuguesa en el rango de las frecuencias medias. Estos

resultados indicarían que entre las diversas poblaciones romaníes analizadas, la croata y la portuguesa son las que han experimentado un mayor grado de deriva genética (Figura 78).

Una vez conocida la influencia de la deriva sobre las frecuencias STR de estas poblaciones, se ha valorado la intensidad del flujo génico experimentado con poblaciones europeas, en particular con sus respectivas poblaciones huésped. El análisis del grado de mestizaje ha revelado que las poblaciones romaníes croata y portuguesa son, precisamente, aquellas que portan un mayor componente del Noroeste de la India. No obstante, su componente europeo es bastante importante, siendo en los croatas del 44,6% y en los portugueses del 36,5%. La población de Eslovaquia también tiene un componente mayoritario asiático (50,3%), siendo menor en los romaníes de Hungría (40,3%) y, sobre todo, en los macedonios (29,9%) (Tabla 20). En resumen, las poblaciones romaníes analizadas muestran para estos marcadores un amplio rango de variación en cuanto al grado de mestizaje. Algunas de ellas muestran valores similares a los observados en los romaníes del País Vasco (entre el 41,6% y el 58,4% de origen europeo), como es el caso de los romaníes croatas y eslovacos; en los húngaros y macedonios son mayores, y en los romaníes portugueses inferior.

Para comprender los resultados obtenidos, cabe señalar que la población romaní de Croacia se originó a raíz del asentamiento de un grupo originario de Rumanía tras la abolición de la esclavitud en el siglo XIX en ese país (Novokmet et al., 2007). La notable incidencia de la deriva genética observada en esta población podría haber sido causada por un efecto fundador inicial, seguido de posteriores efectos fundadores y cuellos de botella, que tuvieron como consecuencia múltiples procesos de aislamiento intra e interpoblacionales. A su vez esta población habría podido sufrir diferentes grados de endogamia (debido a su aislamiento en subpoblaciones reducidas), de mezcla diferencial y un efecto de la deriva genética a lo largo del tiempo (Novokmet et al., 2007). Es posible que el contexto social y/o político haya podido afectar al grado de flujo génico entre la población general croata y la romaní, debido a la discriminación tanto a nivel social como laboral de este grupo étnico (Minority Rights Group International, 2008).

En lo que respecta a las poblaciones romaníes húngara y eslovaca, los resultados obtenidos en los análisis MDS y AFC no muestran un elevado grado de similitud genética con los croatas (Figuras 72-75), como cabría esperar por su proximidad geográfica y su presumible origen común (Ehler y Vanek, 2017; Font-Porterías et al., 2019; Soták et al., 2008). En el artículo de Soták et al. (2008) se trata esta cuestión y se propone que las poblaciones romaníes de Europa central serían en realidad un conglomerado de poblaciones fundadoras genéticamente

aisladas. En cualquier caso, se ha observado un comportamiento notablemente heterogéneo entre los diferentes grupos romaníes en los distintos análisis realizados. Así, en los MDS y AFC los croatas muestran un elevado grado de aislamiento, pero también húngaros y eslovacos tienen un notable grado de aislamiento según el análisis del centroide (Figura 76), y manifiestan diferencias en cuanto a su grado de mestizaje (Tabla 20). Posiblemente, el aislamiento ha propiciado el efecto de la deriva genética promoviendo la diferenciación entre poblaciones romaníes vecinas.

En la población eslovaca hay indicios de que la deriva ha influido apreciablemente en su patrimonio genético, por encima de la húngara, que ha mostrado los índices más bajos del efecto de este agente microevolutivo de todas poblaciones romaníes analizadas, según los valores de $-\log(P)$ (Tabla 18 y Figura 78). A su vez, ambas poblaciones han tenido posiciones relativas más próximas al grupo europeo que lo observado en los croatas, siendo la romaní-húngara la más próxima (Figuras 72-75). Estos resultados concuerdan con su grado de mestizaje, ya que en los húngaros es cercano al 60% (siendo uno de los más altos registrados en este estudio), en tanto que la población eslovaca muestra valores muy similares de ambos componentes, el indio y el europeo, lo que indica que, pese a su aislamiento, ha sufrido cierto grado de flujo génico (Tabla 20). En resumen, se observa que, en general, aquellas poblaciones con valores más bajos de deriva genética muestran un mayor flujo génico, y viceversa.

En esta investigación se valoró también la posibilidad de que el mestizaje tuviera alguna dependencia del porcentaje de población romaní asentada en sus respectivos territorios (Figura 79). Se ha observado que, a mayor porcentaje de población romaní mayor era el grado de mestizaje, existiendo una correlación lineal positiva significativa entre ambas variables (Tabla 21). Una posible explicación a esta correlación podría ser el contexto sociopolítico anteriormente mencionado en el caso de los romaníes croatas, que conlleva a diversos tipos de discriminación y que parece ser un hecho bastante frecuente respecto a la etnia romaní (FSG Annual Report, 2020). Así, algunas de las dificultades comunes a las que se enfrentan los romaníes europeos son el acceso limitado a una educación de calidad e impedimentos para integrarse en el mercado laboral, lo que genera más pobreza y exclusión social, falta de atención médica de calidad y malas condiciones de vida (<https://www.europarl.europa.eu/news/es/headlines/society/20200918STO87401/que-hace-la-ue-para-atajar-la-discriminacion-gitana>, consultado el 16 de noviembre de 2022).

Unas mejores condiciones políticas y económicas en cada país de acogida promoverían una mayor migración y asentamiento, aumentando el porcentaje de esta etnia frente a la población general y, a su vez, estas condiciones podrían promover un mejor clima social entre ambas poblaciones facilitando su mestizaje. Minorities at Risk Project (2006) es un proyecto realizado por el Centro para el Desarrollo Internacional y Gestión de Conflictos de la universidad de Maryland, que monitoriza y analiza el estatus de grupos étnicos minoritarios, entre ellos los romaníes, en un gran número de países como es el caso de Eslovaquia. En sus informes se describen tanto eventos de discriminación como de avances en la integración, así como una cada vez mayor representación de esta etnia en la vida política del país, mediante la formación de partidos políticos romaníes. La posibilidad de participación en el proceso político ha podido favorecer cierto grado de mestizaje e incentivar la diáspora hacia estos territorios, aún sufriendo todavía muchos casos de discriminación y racismo.

En lo que compete a la población romaní húngara, y al igual que otras poblaciones romaníes, ha sido objeto de persecución hasta principios del siglo XX (Kemény y Janky, 2005). En los años posteriores a la Segunda Guerra Mundial los romaníes de Hungría comenzaron a tener un trato más igualitario que en tiempos previos. Pero no fue hasta después de la Revolución Húngara de 1956, durante la cual varios miles de romaníes participaron en un levantamiento, cuando se consiguió un mayor grado de integración tanto social como laboral con la población general. Es posible que este contexto haya posibilitado un mayor grado de mestizaje que en otras poblaciones romaníes centroeuropeas.

Por su parte, la población romaní portuguesa ha mostrado en los análisis realizados posiciones relativas cercanas a alguna de las poblaciones romaníes centroeuropeas, particularmente a la eslovaca (Figuras 72-75). También ha manifestado un grado de aislamiento muy elevado, según los resultados del análisis del centroide (Figura 76). Esto explicaría el bajo valor de su grado de mestizaje con poblaciones europeas, un 36,5%, el más bajo de los obtenidos en las poblaciones romaníes analizadas (Tabla 20). Presumiblemente este aislamiento también habría promovido un mayor efecto de la deriva genética, como se refleja con los resultados obtenidos para $-\log(P)$ (Tabla 18 y Figura 78), mostrando el número máximo de alelos afectados junto a los croatas. La distribución de los valores de $-\log(P)$ indica que esta población ha experimentado un cuello de botella intenso en tiempos recientes. Sin embargo, nuestros resultados contrastan con los publicados por Gusmão et al. (2010), que han estimado un grado de mestizaje mayoritario de la etnia romaní con la población portuguesa receptora. Rodríguez

et al. (2011) analizaron el estado de convivencia de esta etnia en Portugal. En su investigación observaron severos problemas en el marco judicial, racismo tanto desde las instituciones como desde la población general, así como políticas de inclusión enfocadas a proporcionar a los miembros de la etnia un espacio en la sociedad, pero no dirigidas realmente a la aceptación e integración de y con sus características particulares. Este ambiente podría haber reducido el flujo génico interpoblacional, dificultado la migración de otros romaníes hacia este país, reducido el asentamiento, e incluso promovido el desplazamiento hacia otros territorios con mejores condiciones de aceptación de su idiosincrasia.

En cuanto a la población romaní de Macedonia, los resultados obtenidos han sido un tanto paradójicos. Su posición en los análisis MDS (Figuras 72-74) y AFC (Figura 75) la sitúa entre las poblaciones europeas e indias, aunque más cerca de estas últimas y muy próxima a la de Irán. De hecho, es la población romaní que muestra una mayor similitud genética (aparentemente) con las poblaciones asiáticas. Y, sin embargo, es la que muestra un mayor grado de mestizaje con un componente europeo muy elevado (70,1%). En el análisis del centroide esta población no sólo no muestra aislamiento, sino un flujo génico por encima de lo esperado (Figura 76). Este resultado concuerda con el obtenido al analizar el grado de mestizaje (Tabla 20), no muy distinto del estimado por Bianco et al. (2020), donde la población romaní macedonia analizada por estos autores mostró valores de mestizaje superiores al 60%. Cabe resaltar que Minorities at Risk Project (2006) menciona en su informe un notable apoyo de la sociedad macedonia a la etnia romaní, fruto de los resultados del proyecto *Decade of Roma Inclusion* iniciado en 2005 en el que participaron 12 países europeos, entre ellos Macedonia del Norte. Además, desde los años 90, los romaníes tienen representación en el parlamento y han participado en gobiernos de coalición. Estas, entre otras variables económicas y sociales, revelan un razonable acomodo de la población romaní en la sociedad macedonia, lo que justificaría sus altos niveles de mestizaje y ejemplifican la interrelación entre los factores de tipo sociocultural e histórico y los genéticos. Como se ha mencionado anteriormente, se ha observado una relación opuesta entre el flujo génico y la deriva, por lo que cabría esperar que esta población, que ha mostrado altos índices de flujo génico, presentase también un grado bajo del efecto de la deriva genética. Sin embargo, los resultados de $-\log(P)$ la sitúan en un nivel medio de deriva, tan sólo por detrás de los romaníes croatas y portugueses. Si se observa la distribución de los valores de $-\log(P)$, la curva es relativamente plana, con un gran número de alelos afectados por deriva ($n=44$), pero con valores bajos. Según estos resultados, cabe pensar que la población romaní de Macedonia podría haber experimentado intensos episodios de deriva, que habrían sido compensados en parte por un flujo génico reciente desde la población receptora. En todo caso,

la posición de esta población se ha visto afectada, de un modo un tanto inesperado, por el efecto de la deriva en los análisis MDS y AFC realizados con las poblaciones europeas y asiáticas (Figuras 72 y 75). Así, cuando se analizan las poblaciones romaníes junto a las europeas (Figura 73), los romaníes de Macedonia aparecen más alejados de ellas que los romaníes húngaros, eslovacos y portugueses; y cuando se analizan junto a las asiáticas ocurre lo mismo, quedando más alejada de las poblaciones de la India que las poblaciones mencionadas (Figura 74).

En resumen, las poblaciones romaníes europeas han mostrado grados variables de flujo génico y aislamiento y, como consecuencia, valores desiguales de la acción de la deriva genética. Esto se refleja en los porcentajes de mestizaje que muestran un rango de variación que va desde el 36,5% de componente europeo en la población romaní portuguesa hasta el 70,1% en la macedonia (Tabla 20), y resultados asimismo desiguales en las estimaciones del efecto de la deriva. Estas importantes diferencias en la acción de ambos agentes microevolutivos coinciden con el planteamiento de que las poblaciones romaníes europeas son un conjunto de poblaciones fundadoras genéticamente aisladas entre sí, que comparten un origen común, pero sujetas a diferentes condiciones tras su asentamiento en diferentes regiones (Ehler y Vanek, 2017; Font-Porterias et al., 2019; Soták et al., 2008). Por último, queda señalar que los STRs, marcadores con gran variabilidad debido a sus altas tasas de mutación (mucho mayor que los SNPs y las inserciones *Alu*), poseen una alta capacidad para caracterizar la diversidad y estructura de las poblaciones humanas (Jobling, 2022; Silva et al., 2012) aunque son más proclives a mostrar los efectos de la deriva genética. Los STRs, como se ha mencionado a lo largo del texto, son una buena herramienta para estudiar la genética de poblaciones en una escala temporal reciente, si bien su capacidad para describir procesos en una escala temporal dilatada es mucho menor. Esto justifica la realización de análisis mediante familias de marcadores con diferentes características, para obtener una mejor representación del origen del patrimonio genético de un grupo humano de tanto interés como la población romaní.

VI.4 Marcadores de linaje

La población romaní cuenta con algunas características culturales que la hacen especialmente idónea para los estudios de marcadores de linajes. En primer lugar, la posición jerárquica de la mujer dentro del núcleo familiar está supeditada a los varones (Kalaydjieva et al., 2001a), por lo que la posibilidad de mestizaje puede verse afectada por este desigual estatus. Igualmente, esta discrepancia social también predispone a los miembros de esta población a experimentar diferentes contextos socio-culturales que pueden afectar a sus relaciones dentro y fuera de su población (Gouva et al., 2015). En segundo lugar, los romaníes presentan un alto

grado de matrimonios concertados en función del estatus de las familias y/o del clan al que pertenecen; como consecuencia presentan un elevado nivel de matrimonios entre parientes (Kalaydjieva et al., 2001a). Ambas características posibilitan que varones y mujeres tengan una historia significativamente diferente, por lo que el análisis de ambos linajes por separado resulta fundamental para el conocimiento de sus orígenes.

Análisis con haplogrupos de cromosoma Y

En primer lugar, se abordó el análisis de los linajes desde el punto de vista de los varones, mediante el estudio de haplogrupos del cromosoma Y tomados de diferentes fuentes. Con el fin de comparar la heterogeneidad entre las muestras continentales para estos haplogrupos se realizó un AMOVA, que mostró diferencias significativas elevadas en todos los niveles organizativos (Tabla 22). Los resultados obtenidos, además de indicar que hay una gran heterogeneidad en los haplogrupos seleccionados, también han mostrado que estos marcadores tienen capacidad para diferenciar ambos grupos continentales. A continuación, se realizaron varios análisis de MDS y AFC para visualizar con mayor detalle las diferencias interpopulacionales. La figura 80 muestra los resultados obtenidos mediante un MDS al comparar ambos grupos continentales junto a las poblaciones romaníes. La Dimensión I diferencia a una buena parte de las poblaciones del subcontinente indio y a las romaníes, en el extremo positivo, respecto a las europeas y, excepcionalmente, a las poblaciones asiáticas de Punjab y pavagadhi chaudhari, todas ellas emplazadas en el extremo negativo. La Dimensión II caracteriza especialmente a las poblaciones del Próximo Oriente, en el extremo positivo, frente al resto de poblaciones consideradas.

El extremo positivo del eje o Dimensión I se caracteriza por la ubicación de las poblaciones koya (del Sur de la India), vasava, gamit, nana chaudhari y mota chaudhari (de Gujarati, Oeste de la India), todas ellas tribus, es decir, no pertenecientes a castas. El término tribu en la India se refiere en general a grupos étnicos no asimilados al grupo de castas y por tanto considerados inferiores a todas ellas. Supuestamente este eje incluye a todas aquellas etnias representantes de los pobladores de la India anteriores a la llegada de los indoeuropeos, algunas de ellas con culturas cazadoras-recolectoras. No obstante, entre ellos se encuentran grupos que hablan idiomas dravídicos, los cuales habrían sido empujados por los indoeuropeos desde regiones situadas más al Oeste, posiblemente desde Irán (Cavalli-Sforza et al., 1994). También se encuentran entre ellos grupos que hablan idiomas austroasiáticos y tibeto-birmanos. Más cerca del centro de la Dimensión I se encuentran las poblaciones dubla, dhodia, konkana (indoeuropeos del Oeste de la India) y chenchu (dravidianos del Sur, cazadores-

recolectores), así como las poblaciones del Próximo Oriente iraquí, libanesa, siria y jordana. Las poblaciones romaníes se ubican en su mayor parte en la mitad positiva de esta dimensión, excepto la española, que se encuentra ligeramente por debajo del valor 0. Es de destacar la proximidad de las poblaciones romaníes de Grecia, Eslovaquia, Rumanía y Bulgaria a las tribus del Sur de la India. Por otra parte, es sorprendente la posición de la población punjabí, que representa en este análisis la región del Noroeste de la India de donde presumiblemente proceden los romaníes, la cual se sitúa en el extremo negativo de la Dimensión I, junto a los europeos. El AFC (Figura 83) ha mostrado unos resultados muy similares, con la salvedad de que ambos ejes (o Dimensiones) se encuentran invertidos. Los análisis por continentes (Figuras 81 y 82) subrayan la posición relativa de las poblaciones romaníes, de tal modo que la griega es la más distante genéticamente de las europeas y a la vez la más emparentada con las tribus koya, vasava, gamit, nana chaudhari y mota chaudhari. Las poblaciones de etnia romaní eslovaca, búlgara y rumana adoptan posiciones intermedias, mientras que la ucraniana, la española y la húngara aparecen menos emparentadas con las tribus, aunque en todos los casos se diferencian claramente de las poblaciones europeas.

Para explicar estos resultados conviene revisar las frecuencias haplotípicas de los diferentes grupos poblacionales considerados (Tabla 8 del Anexo). Como es conocido, las poblaciones romaníes muestran frecuencias elevadas del haplogrupo H, con frecuencias que oscilan entre 0,200 en los romaníes españoles y 0,672 en los griegos. Entre las poblaciones europeas, este haplogrupo solo aparece con frecuencias bajas en la húngara (0,053), la búlgara (0,010) y la eslovaca (0,006), no habiendo sido detectado en el resto. En Próximo Oriente no alcanza en ningún caso el valor 0,05, como tampoco en la muestra de población punjabí. Por el contrario, en las poblaciones koya, vasava, gamit, nana chaudhari y mota chaudhari los valores de este haplogrupo se sitúan en todos los casos por encima de 0,6, con un valor máximo en los koya de 0,707. Otras poblaciones de la India, tanto tribus como castas, tienen valores intermedios. El gráfico centroide nos ha permitido profundizar en los resultados observados en las poblaciones romaníes (Figura 84). De hecho, se ha observado una relación bastante clara entre el aislamiento detectado por este método de análisis y la proximidad genética a las tribus de la India mencionadas anteriormente. Así, la población romaní griega es la que ha mostrado un mayor grado de aislamiento y también la más próxima a las poblaciones tribales; asimismo, tiene la frecuencia más alta del haplogrupo H (0,672). Las poblaciones romaníes eslovaca y rumana revelan también un alto grado de aislamiento relativo, así como un notable parentesco con las poblaciones tribales, y frecuencias elevadas de H (0,542 y 0,500, respectivamente). Finalmente, las poblaciones romaníes búlgara, húngara, ucraniana y española no muestran un

grado de aislamiento importante, aparecen más distantes de las poblaciones tribales y sus frecuencias de H tienen valores inferiores (0,457, 0,342, 0,286 y 0,200, respectivamente) respecto a las poblaciones romaníes antes mencionadas.

La distribución diferencial de frecuencias del haplogrupo H entre poblaciones europeas y asiáticas se observa perfectamente en el gráfico de la correspondiente clina (Figura 86). En él se aprecia el reducido rango de frecuencias en poblaciones europeas y del Próximo Oriente, con valores cercanos a 0 y el mayor rango de frecuencias de las poblaciones del Sur de Asia. A su vez, el rango de frecuencias de las poblaciones romaníes es muy amplio y se solapa con el de las poblaciones asiáticas, concretamente con las poblaciones tribales de la India. El análisis de $-\log(P)$ ha detectado una cierta acción de la deriva, que afecta concretamente al haplogrupo J en todas las poblaciones, y al R en alguna de ellas. En el primer caso, puede ser el resultado de un efecto fundador común a todas las subpoblaciones romaníes y, en el segundo, podría deberse un proceso mucho más reciente. En todo caso, es importante resaltar que no se han encontrado evidencias de deriva genética en relación al haplogrupo H.

Finalmente, los resultados del análisis del mestizaje han sido concordantes con el resto de resultados observados para los haplogrupos del cromosoma Y, pero al mismo tiempo resultan sorprendentes. Así, no se observa flujo génico desde las poblaciones europeas hacia los varones romaníes griegos y eslovacos, es apenas perceptible en los rumanos (0,098), bajo en los romaníes españoles (0,282), búlgaros (0,302) y húngaros (0,347), y relativamente alto en los ucranianos (0,559); en este último caso el valor obtenido se corresponde con su posición en el centroide, ya que es la única población romaní situada por encima de la línea de los valores esperados. Por otra parte, los griegos son la población romaní con un mayor grado de aislamiento según este mismo análisis. Podemos señalar además que, a grandes rasgos, y según los análisis MDS y AFC, las poblaciones con un menor grado de mestizaje se han ubicado más próximas a las poblaciones tribales de la India. La población romaní española, en todo caso, parece mostrar un grado de mestizaje menor del que cabría esperar en los análisis MDS, AFC y del centroide, e incluso por su frecuencia del haplogrupo H, la menor entre los romaníes. Cabe pensar en algún proceso importante de efecto fundador en esta población, ya que si bien el número de haplogrupos afectados por deriva en todas las poblaciones ha sido escaso, el valor de $-\log(P)$ observado en el haplogrupo J de los romaníes españoles es con diferencia el más alto de los observados en todos los romaníes (>323) y, por otra parte, es la población con el menor número de haplogrupos presente, tan sólo 4.

Las diferencias observadas tanto en el grado de aislamiento como en el flujo génico de las poblaciones romaníes europeas son presumiblemente un reflejo de las diferentes condiciones a las que se han visto expuestas en cada territorio y a los posibles cuellos de botella sucesivos que ha experimentado la etnia durante su diáspora (Font-Porterías et al., 2019; Martínez-Cruz et al., 2016; Soták et al., 2008). Existe un cierto paralelismo con los resultados previamente obtenidos con los STRs autosómicos del sistema CODIS, en el sentido de que las poblaciones romaníes europeas serían en realidad un conjunto de poblaciones genéticamente aisladas, como han apuntado diversos autores (Font-Porterías et al., 2021; Kalaydjieva et al., 2001b; Soták et al., 2008), ya que la nube de puntos romaní tiende a aparecer muy dispersa en los sucesivos análisis realizados. Al mismo tiempo, su diversidad se diferencia claramente de la de otras poblaciones, de modo que aunque amplia, la nube de puntos romaní es en general identificable en relación a las de Europa o del Sur de Asia. Esto implicaría un mismo origen, por lo que, a pesar de la subdivisión posterior y el elevado grado de aislamiento, los romaníes habrían mantenido en cierta medida sus similitudes genéticas a lo largo del tiempo (Gresham et al., 2001; Martínez-Cruz et al., 2016).

Se ha planteado en esta investigación la posibilidad de que, como han revelado otros marcadores, el grado de mestizaje tuviera alguna relación con el porcentaje de población romaní asentada en cada territorio (Figura 87 y Tabla 27). Sin embargo, no se ha encontrado ninguna correlación significativa entre ambas variables, a diferencia de los resultados obtenidos mediante los marcadores autosómicos de STR. Por otra parte, la especificidad de los resultados del cromosoma Y radica en el escaso componente de origen europeo en el patrimonio genético de los romaníes en la mayor parte de sus poblaciones, que oscila entre el 55,9% en Ucrania y el 0% en Grecia y Eslovaquia. Esto difiere notablemente de los resultados obtenidos por Martínez-Cruz et al. (2016), que mencionan unos valores que van desde el 46% en Hungría al 17% en Rumanía. En nuestro caso, los valores para estas dos poblaciones han sido más bajos: 34,7% y 9,8%, respectivamente. Más curioso resulta que el grupo de Martínez-Cruz obtuviera una probabilidad máxima para el Norte de la India como lugar de origen de los linajes masculinos romaníes, cuando es precisamente en esta región donde más bajas son las frecuencias del haplogrupo H, que es el mayoritario en los romaníes, como se ha visto anteriormente.

Análisis con haplogrupos de ADNmt

Con el objetivo de describir la estructura genética de las poblaciones romaníes desde el punto de vista de los linajes femeninos, se ha analizado una base de datos de frecuencias de haplogrupos del ADNmt, para lo cual se recopilaron las frecuencias de 28 haplogrupos en 43

poblaciones euroasiáticas. De este modo, junto a los resultados del cromosoma Y, hemos podido obtener una imagen más completa de la historia de esta etnia. Se realizó un AMOVA, separando las poblaciones en sus dos grupos continentales, que dio como resultado una alta significación en todos los niveles organizativos, confirmando la capacidad de estos haplogrupos para discriminar a las poblaciones seleccionadas, y un primer indicio de la alta heterogeneidad presentada por este tipo de marcadores (Tabla 28). En los sucesivos análisis MDS y AFC, la población romaní ucraniana destacó como la más heterogénea. Así, tanto en el MDS que incluyó a poblaciones de ambos continentes (Figura 88), como en los MDS que analizaron por separado a grupos europeos o asiáticos (Figuras 89 y 90), y también en el AFC (Figura 91), esta población se ha situado claramente alejada de todas las demás. Como ya se ha mencionado, los romaníes de Ucrania presentan una diversidad genética del ADN mitocondrial muy baja, con tan sólo 4 haplogrupos presentes y uno de ellos claramente predominante, el haplogrupo J, con una frecuencia de 0,765. Por tanto resulta evidente, según los diferentes análisis realizados, que la deriva genética en los linajes femeninos de las romaníes ucranianas ha sido muy elevada.

Es preciso mencionar también a las romaníes eslovacas que, si en los MDS se situaban dentro de la nube de puntos romaní, especialmente cercana a la de las poblaciones europeas (Figura 89), en el análisis AFC han aparecido alejadas del grupo principal de poblaciones, quedando más próximas a los romaníes de Ucrania (Figura 91). Este tipo de discordancias justifica el interés de realizar análisis multivariantes redundantes mediante diferentes métodos, con el fin de poner de manifiesto posibles artefactos estadísticos. En este caso, la posición de las romaníes eslovacas en el AFC queda caracterizada por el haplotipo I y en menor medida por el J (Figura 92). Esta población presenta el valor más alto del haplogrupo I (0,319), seguida por la población romaní ucraniana (0,125) (Tabla 9 del Anexo). Por su parte, las romaníes ucranianas tienen el valor más alto del haplogrupo J (0,765), muy por encima del resto de poblaciones, ya que ninguna supera el valor 0,3. Estos valores tan elevados han generado un eje 2 en el AFC que en buena medida refleja la varianza residual originada por estas dos poblaciones de etnia romaní, debido en última instancia a la acción de la deriva. De hecho, el haplogrupo I ha tenido un valor significativo para $-\log(P)$ en ambas poblaciones y el haplogrupo J en la eslovaca (Tablas 30 y 31).

Por lo que respecta al resto de poblaciones romaníes, se sitúan en el MDS global dentro de la nube de puntos de las poblaciones europeas y del Sur de Asia (Figura 88). Considerando por separado ambos continentes, la población más próxima a las europeas es la eslovaca, seguida de la húngara y la búlgara, estando más separadas la rumana, la española y la griega

(Figura 89). Este es, precisamente, el orden en el grado de mestizaje que hemos obtenido con poblaciones europeas, con un valor máximo para la eslovaca (0,735) y un valor mínimo para la griega (0,107) (Tabla 32). En el análisis MDS con poblaciones asiáticas el orden de proximidad a estas ha sido el contrario, con la excepción relativa de las romaníes españolas, que aparecen en una posición próxima a las húngaras y búlgaras (Figura 90). En el análisis AFC, con la excepción mencionada de las españolas, la ordenación de las romaníes entre europeas y asiáticas ha sido similar, observándose de nuevo una menor relación con las poblaciones asiáticas de lo que cabría esperar por su grado de mestizaje. El grado de flujo génico procedente de Europa calculado para los linajes del ADN mitocondrial ha sido muy heterogéneo, con valores que oscilan entre el 73,5% en las mujeres eslovacas y el 10,7% de las griegas. Entre las poblaciones romaníes con un componente asiático mayoritario se encuentran la griega (89,3%), la española (70,1%) y la rumana (63,0%). Con una proporción menor se encuentran la búlgara (43,7%), la ucraniana (38,5%), la húngara (36,4%) y la eslovaca (26,5%).

Por otra parte, se ha observado un desequilibrio importante en las tasas de mestizaje entre linajes masculinos y femeninos. Si bien las poblaciones analizadas para los diferentes marcadores no son coincidentes en general, se ha podido estimar el grado de mestizaje para los romaníes húngaros y eslovacos a partir de haplogrupos del ADN mitocondrial, del cromosoma Y y de marcadores STR autosómicos del sistema CODIS. En ambos casos, la proporción de origen asiático ha sido mayor en los linajes masculinos (100% en eslovacos y 65,3% en húngaros, para el cromosoma Y), respecto a los femeninos (26,5% en eslovacas y 36,4% en húngaras, para el ADN mitocondrial). El rango observado en el conjunto de poblaciones para el cromosoma Y se encuentra entre el 44,1% y el 100%, con un valor medio del 77,3%, mientras que para el ADN mitocondrial varía entre el 26,5% y el 89,3%, con un valor medio del 52,5%.

Los valores estimados para cromosomas autosómicos se encuentran, lógicamente, en valores intermedios tanto en eslovacos (50,3%) como en húngaros (40,3%). Ambos resultados se encuentran más cerca de los estimados mediante linajes femeninos que con masculinos, lo que induce a pensar en un mayor número de mujeres fundadoras y, por ende, en un número menor de varones fundadores. Las diferencias observadas entre linajes masculinos y femeninos podrían indicar la presencia de un factor diferencial relacionado con el sexo (y género en sentido más social y de costumbres). Algunos autores han propuesto que, como consecuencia de la dominancia de los varones frente a las mujeres dentro de la etnia, se puede haber permitido a los primeros una mayor libertad en sus enlaces (Mendizábal et al., 2011; Weyrauch, 2001). Tan sólo en periodos de cambios en las políticas de exclusión varones y mujeres habrían tenido una

libertad de elección similar (García-Fernández et al., 2020; Mendizábal et al., 2011; Pereira et al., 2012; Weyrauch, 2001). En general se habría producido un mayor número de enlaces entre varones romaníes y mujeres externas a la etnia, siendo menos habitual el caso de mujeres romaníes unidas con varones externos, ya que no tendrían (salvo excepciones) esta posibilidad. De esta forma, habría una mayor entrada de haplogrupos del ADNmt dentro de la población romaní procedentes de su población huésped, que de haplogrupos del cromosoma Y. En consecuencia, el flujo génico observado en las poblaciones romaníes con marcadores autosómicos correspondería en mayor medida a la entrada de mujeres dentro de las poblaciones romaníes. Finalmente, hay que señalar que la discrepancia en los valores de mestizaje entre ambos sexos ha sido estudiada en otras poblaciones (Alfonso-Sánchez et al., 2018, 2019; Bajić et al., 2018; Ma et al., 2021; Ongaro et al., 2021; Resano y Moral, 2018). En el caso de la etnia romaní, García-Fernández et al. (2020) han obtenido resultados que sustentan la hipótesis de que la inclusión de mujeres europeas no-romaníes dentro de la población romaní ha sido tradicionalmente más habitual que la de varones, lo que apoya sin duda los resultados obtenidos en la presente investigación.

VI.5 Recapitulación

En definitiva, considerando los distintos tipos de marcadores utilizados en el presente estudio, los diferentes análisis realizados y las diferentes poblaciones consideradas, hemos observado que la población romaní del País Vasco, durante su periodo migratorio, ha estado sometida a períodos de aislamiento y tamaños efectivos reducidos, como consecuencia de efectos fundadores y cuellos de botella sucesivos que han podido propiciar la acción de la deriva genética, encontrándose evidencias de la misma tanto en un gran número de los SNPs estudiados como de inserciones *Alu*. Este mismo aislamiento ha podido mantenerse durante las primeras etapas tras su asentamiento en la actual área del País Vasco durante el siglo XV (Font-Porterías et al., 2019; Lermo et al., 2006), promovido tanto por las características culturales propias de este grupo étnico, como por el contexto socio-político del periodo en que se fueron asentando, que llegó a su extremo en el siglo XVIII (Kalaydjieva et al., 2001a; Melegh et al., 2017; Martínez-Martínez, 2014).

Por otra parte, hemos observado un notable impacto del flujo génico en la población romaní estudiada procedente tanto de su población huésped como de otras europeas y, eventualmente, del Próximo Oriente. Un elevado componente europeo también se ha observado en otras poblaciones romaníes europeas, situándose la población romaní del País Vasco en un valor intermedio (Bianco et al., 2020; Gusmão et al., 2010). Este cambio de

tendencia en relación a la población huésped ha podido ser causado por variaciones en el Estado de derecho y en el contexto social, como se refleja en los diversos proyectos de integración y en el reconocimiento de los derechos de esta etnia por el Parlamento Europeo durante la última década (Sordé et al., 2013), si bien, aún se mantiene en la actualidad cierto grado de discriminación tanto en el ámbito social como laboral (Minorities at Risk Project, 2006). Los cambios señalados han podido reducir la barrera cultural entre ambas poblaciones promoviendo una mayor permeabilidad en sus enlaces interpoblacionales.

En lo que compete a las poblaciones romaníes europeas analizadas mediante STRs, hemos detectado una notable variabilidad en las acciones de los agentes microevolutivos estudiados. Dicha heterogeneidad puede ser la demostración de las diferentes condiciones a las que han estado expuestas estas poblaciones y al escaso contacto entre ellas, que ha podido generar una subestructuración de las mismas posterior a su entrada en el continente europeo (Ehler y Vanek, 2017; Font-Porterías et al., 2019, 2021; Gusmão et al., 2010; Kalaydjieva et al., 2001b; Soták et al., 2008). Esta subdivisión promovió sin duda la aparición de subgrupos tras el asentamiento de las diferentes poblaciones romaníes, como sería el caso de los valaco romaníes, descendientes de las poblaciones romaníes asentadas en la región de Valaquia (actual Rumanía) y sometidas a esclavitud hasta el siglo XIX (Fraser, 1992; Gresham et al., 2001; Hancock, 1987). Otro ejemplo sería el de los zhutanos, descendientes del mestizaje entre judíos y romaníes en los campos de concentración durante la Segunda Guerra Mundial (Kalaydjieva et al., 2005).

Por otra parte, se ha observado la misma heterogeneidad en cuanto a la acción de los agentes microevolutivos (detectada mediante el uso de marcadores autosómicos) en los haplogrupos de ambos linajes (comosoma Y y ADNmt). Además de exhibir un amplio rango de actuación tanto del flujo génico como de la deriva genética, los haplogrupos también han demostrado un reducido contacto entre las distintas poblaciones romaníes europeas. Estos resultados coinciden con lo planteado con marcadores autosómicos y con lo observado por otros autores, siendo indicio de la subestructuración de las poblaciones romaníes tras la entrada en el continente (Gusmão et al., 2010; Soták et al., 2008). Además, se ha observado un mestizaje diferencial según el sexo en las poblaciones romaníes europeas, con mayores indicios de mestizaje en los marcadores del linaje femenino respecto al masculino, y un valor intermedio en el caso de marcadores autosómicos, pero más próximos al linaje femenino. Varios autores explican esta discrepancia en función de la tradicional dominancia de los varones frente a las mujeres, característica sociocultural de esta etnia. Esta jerarquía podría promover una mayor permisividad en los enlaces interpoblacionales entre varones romaníes y mujeres de otro origen

(pero no de mujeres romaníes con varones extrínsecos a esta etnia), lo que se traduciría en una mayor entrada de haplogrupos de ADNmt con origen externo que de haplogrupos del cromosoma Y (García-Fernández et al., 2020; Mendizábal et al., 2011; Weyrauch, 2001).

Por último, los análisis con ambos linajes han contribuido a explicar el origen indio de esta etnia, pero a su vez han mostrado discordancias en cuanto a la región o área geográfica. Por un lado, los análisis del linaje masculino han situado su origen alejado del Noroeste del subcontinente indio debido a la presencia del haplogrupo H, mayoritario en las poblaciones romaníes y en las tribus del Sur de la India, pero ausente tanto en las poblaciones europeas (por lo que no puede proceder del mestizaje con dichas poblaciones durante su diáspora), como en las poblaciones del Punjab en el Noroeste, origen teórico de la etnia romaní. Sin embargo, los haplogrupos del ADNmt localizan su origen en esta misma región norteña (el Punjab), coincidiendo con lo señalado por otros autores (Mastana et al., 1992; Mendizábal et al., 2011, 2012). Un evento de cuello de botella o efecto fundador posterior al inicio de la diáspora romaní podría haber provocado la desaparición del haplogrupo H del acervo genético de las poblaciones del Noroeste de la India. El mantenimiento de dicho haplogrupo en las tribus y castas del Sur podría haber ocurrido debido al aislamiento entre ambos extremos del subcontinente, provocado por la estructuración en castas y por la barrera al flujo génico, lo que generó un gradiente Sur-Norte de la presencia de este haplogrupo (Kivisild et al., 2003). O también el origen de los linajes masculinos de los romaníes podría encontrarse en algún grupo todavía no identificado.



Conclusiones

VII. Conclusiones

Según los resultados obtenidos en el presente manuscrito, se han extraído las siguientes conclusiones:

- I. Se observa un notable grado de flujo génico acumulado en la población romaní del País Vasco, tanto a partir de marcadores SNP como de inserciones *Alu*, con unas proporciones similares en su patrimonio genético provenientes tanto de Europa como del Sur de Asia.
- II. La acción de la deriva genética también es importante en esta población, en la que se ha detectado un efecto potencial de la misma en más del 40% de los SNPs e inserciones *Alu* analizados.
- III. El análisis de marcadores STRs en poblaciones romaníes de diferentes países europeos revela una notable heterogeneidad en la intensidad de las presiones microevolutivas experimentadas por dichas poblaciones, lo que sustenta la hipótesis de un alto grado de subdivisión entre los diferentes grupos de la etnia romaní.
- IV. Los valores estimados para el grado de mestizaje de las diversas poblaciones romaníes analizadas se distribuyen en un amplio rango de variación, que va desde un componente mayoritariamente europeo, superior al 70% en la población romaní de Macedonia, hasta un componente fundamentalmente asiático, superior al 63% en la población romaní de Portugal, quedando la población romaní del País Vasco en un lugar intermedio entre ambas.
- V. En el contexto de una notable acción de la deriva genética en todas las poblaciones romaníes consideradas, se han identificado episodios especialmente intensos en los romaníes croatas y portugueses en cuanto a los marcadores autosómicos, y en los romaníes ucranianos en el ADN mitocondrial.
- VI. El análisis de las frecuencias de haplogrupos del cromosoma Y y del ADN mitocondrial muestra un mestizaje diferencial en relación al sexo, existiendo un componente asiático mayoritario en los linajes masculinos y unas proporciones similares de origen asiático y europeo en los femeninos.
- VII. Los resultados obtenidos a partir del ADN mitocondrial confirman el parentesco de los romaníes con poblaciones del noroeste de la India, pero los haplogrupos del cromosoma Y indican un alto grado de parentesco con poblaciones tribales del Oeste y el Sur de la India.



Bibliografía

VIII. Bibliografía

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature*. 2015 Oct 1;526(7571):68-74. doi: 10.1038/nature15393. PMID: 26432245; PMCID: PMC4750478.

Abdelmohsen K, Panda AC, Kang MJ, Guo R, Kim J, Grammatikakis I, Yoon JH, Dudekula DB, Noh JH, Yang X, Martindale JL, Gorospe M. 7SL RNA represses p53 translation by competing with HuR. *Nucleic Acids Res*. 2014 Sep;42(15):10099-111. doi: 10.1093/nar/gku686. Epub 2014 Aug 14. PMID: 25123665; PMCID: PMC4150789

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogväli EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A. 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*. 2004 Nov;75(5):910-8. doi: 10.1086/425590. Epub 2004 Sep 20. PMID: 15382008; PMCID: PMC1182122.

ADN no codificante. National Human Genome Research Institute, NIH. Recuperado de: <https://www.genome.gov/es/genetics-glossary/ADN-no-codificante> [27/01/2022]

Aguiar VR, de Castro AM, Almeida VC, Malta FS, Ferreira AC, Louro ID. New CODIS core loci allele frequencies for 96,400 Brazilian individuals. *Forensic Sci Int Genet*. 2014 Nov;13:e6-e12. doi: 10.1016/j.fsigen.2014.04.015. Epub 2014 May 10. PMID: 24888518.

Aguilar-Velázquez JA, Martínez-Cortés G, Inclán-Sánchez A, Romero-Rentería O, Díaz-Navarro XX, Rangel-Villalobos H. Population data of 23 STR loci (PowerPlex® FusionSystem) in Mexican Mestizos from the West Region. *Int J Legal Med*. 2016 Nov;130(6):1489-1491. doi: 10.1007/s00414-016-1361-0. Epub 2016 Apr 5. PMID: 27048213.

Ahmed W, Malik M, Saeed I, Khan AA, Sadeque A, Kaleem U, Ahmed N, Ajmal M, Azam M, Qamar R. Role of tissue plasminogen activator and plasminogen activator inhibitor polymorphism in myocardial infarction. *Mol Biol Rep*. 2011 Apr;38(4):2541-8. doi: 10.1007/s11033-010-0392-8. Epub 2010 Nov 17. PMID: 21082259.

Aktaş T, Avşarlılık İ, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, Manke T, Backofen R, Akhtar A. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature*. 2017 Apr 6;544(7648):115-119. doi: 10.1038/nature21715. Epub 2017 Mar 29. PMID: 28355180.

Alfonso-Sánchez, M. A, Espinosa, I, Gómez-Pérez, L, Poveda, A, Rebato, E, & Peña, J. A. (2018). Tau haplotypes support the Asian ancestry of the Roma population settled in the Basque Country. *Heredity*, 120(2), 91.

Alfonso-Sánchez MA, Gómez-Pérez L, Dipierri JE, Peña JA. Paternal heritage in Jujuy province (Northwest Argentina): Evidence for sex-biased gene flow and genetic drift effects. *Am J Hum Biol.* 2019 Jul;31(4):e23262. doi: 10.1002/ajhb.23262. Epub 2019 May 30. PMID: 31145506.

Algee-Hewitt BF, Edge MD, Kim J, Li JZ, Rosenberg NA. Individual Identifiability Predicts Population Identifiability in Forensic Microsatellite Markers. *Curr Biol.* 2016 Apr 4;26(7):935-42. doi: 10.1016/j.cub.2016.01.065. Epub 2016 Mar 17. PMID: 26996508.

Álvarez Z P, Verdugo L P, Carvajal K L, Múhlhausen M G, Ríos A P, Rodríguez V D. Activador del plasminógeno tisular recombinante en el manejo de trombos cardiacos en recién nacidos [Recombinant tissue plasminogen activator for the management of intracardiac thrombi in newborns]. *Rev Chil Pediatr.* 2015 May-Jun;86(3):194-9. Spanish. doi: 10.1016/j.rchipe.2015.03.004. Epub 2015 Jul 30. PMID: 26235268.

Amigo J, Phillips C, Salas T, Fernández Formoso L, Carracedo A, Lareu M. pop.STR - An online population frequency browser for established and new forensic STRs. *Forensic Sci. Int. Gene. Suppl.* 2009 (10.1016/j.fsigss.2009.08.178)

Andreeva AV, Kutuzov MA. Cadherin 13 in cancer. *Genes Chromosomes Cancer.* 2010 Sep;49(9):775-90. doi: 10.1002/gcc.20787. PMID: 20607704.

Angst BD, Marozzi C, Magee AI. The cadherin superfamily: diversity in form and function. *J Cell Sci.* 2001 Feb;114(Pt 4):629-41. PMID: 11171368.

Arakaki N, Nishihama T, Owaki H, Kuramoto Y, Suenaga M, Miyoshi E, Emoto Y, Shibata H, Shono M, Higuti T. Dynamics of mitochondria during the cell cycle. *Biol Pharm Bull.* 2006 Sep;29(9):1962-5. doi: 10.1248/bpb.29.1962. PMID: 16946518.

Arcot SS, Fontius JJ, Deininger PL, Batzer MA. Identification and analysis of a 'young' polymorphic Alu element. *Biochim Biophys Acta.* 1995a Jul 25;1263(1):99-102. doi: 10.1016/0167-4781(95)00080-z. PMID: 7632743.

Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA. Alu repeats: a source for the genesis of primate microsatellites. *Genomics.* 1995b Sep 1;29(1):136-44. doi: 10.1006/geno.1995.1224. PMID: 8530063.

Arnold SJ, Pfrender ME, Jones AG. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica.* 2001;112-113:9-32. PMID: 11838790.

Asari M, Umetsu K, Adachi N, Azumi J, Shimizu K, Shiono H. Utility of haplogroup determination for forensic mtDNA analysis in the Japanese population. *Leg Med (Tokyo).* 2007 Sep;9(5):237-40. doi: 10.1016/j.legalmed.2007.01.007. Epub 2007 Apr 27. PMID: 17467322.

Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, Reynolds A, Indap A, Wright MH, Degenhardt JD, Gutenkunst RN, King KS, Nelson MR, Bustamante CD. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res.* 2009 May;19(5):795-803. doi: 10.1101/gr.088898.108. Epub 2009 Feb 13. PMID: 19218534; PMCID: PMC2675968.

Bajić V, Barbieri C, Hübner A, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, Mpoloka SW, Roewer L, Purps J, Stoneking M, Pakendorf B. Genetic structure and sex-biased gene flow in the history of southern African populations. *Am J Phys Anthropol*. 2018 Nov;167(3):656-671. doi: 10.1002/ajpa.23694. Epub 2018 Sep 7. PMID: 30192370; PMCID: PMC6667921.

Barbosa FB, Cagnin NF, Simioni M, Farias AA, Torres FR, Molck MC, Araujo TK, Gil-Da-Silva-Lopes VL, Donadi EA, Simões AL. Ancestry Informative Marker Panel to Estimate Population Stratification Using Genome-wide Human Array. *Ann Hum Genet*. 2017 Nov;81(6):225-233. doi: 10.1111/ahg.12208. Epub 2017 Sep 11. PMID: 28895130.

Barbujani G, Colonna V. Human genome diversity: frequently asked questions. *Trends Genet*. 2010 Jul;26(7):285-95. doi: 10.1016/j.tig.2010.04.002. Epub 2010 May 13. PMID: 20471132.

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M. Genetic variation of recent Alu insertions in human populations. *J Mol Evol*. 1996a Jan;42(1):22-9. doi: 10.1007/BF00163207. PMID: 8576959.

Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E. Standardized nomenclature for Alu repeats. *J Mol Evol*. 1996b Jan;42(1):3-6. doi: 10.1007/BF00163204. PMID: 8576960.

Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet*. 2002 May;3(5):370-9. doi: 10.1038/nrg798. PMID: 11988762.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, et al. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci U S A*. 1994 Dec 6;91(25):12288-92. doi: 10.1073/pnas.91.25.12288. PMID: 7991620; PMCID: PMC45422.

Batzer MA, Rubin CM, Hellmann-Blumberg U, Alegria-Hartman M, Leeflang EP, Stern JD, Bazan HA, Shaikh TH, Deininger PL, Schmid CW. Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. *J Mol Biol*. 1995 Mar 31;247(3):418-27. doi: 10.1006/jmbi.1994.0150. PMID: 7714898.

Bellance N, Lestienne P, Rossignol R. Mitochondria: from bioenergetics to the metabolic regulation of carcinogenesis. *Front Biosci (Landmark Ed)*. 2009 Jan 1;14(11):4015-34. doi: 10.2741/3509. PMID: 19273331.

Benham FJ, Spurr N, Povey S, Brinton BT, Goodfellow PN, Solomon E, Harris TJ. Assignment of tissue-type plasminogen activator to chromosome 8 in man and identification of a common restriction length polymorphism within the gene. *Mol Biol Med*. 1984 Aug;2(4):251-9. PMID: 6100560.

Berkman CC, Dinc H, Sekeryapan C, Togan I. Alu insertion polymorphisms and an assessment of the genetic contribution of Central Asia to Anatolia with respect to the Balkans. *Am J Phys Anthropol.* 2008 May;136(1):11-8. doi: 10.1002/ajpa.20772. PMID: 18161848.

Bhatti S, Abbas S, Aslamkhan M, Attimonelli M, Trinidad MS, Aydin HH, de Souza EMS, Gonzalez GR. Genetic perspective of uniparental mitochondrial DNA landscape on the Punjabi population, Pakistan. *Mitochondrial DNA A DNA Mapp Seq Anal.* 2018 Jul;29(5):714-726. doi: 10.1080/24701394.2017.1350951. Epub 2017 Jul 26. PMID: 28745560.

Bianco E, Laval G, Font-Porterías N, García-Fernández C, Dobon B, Sabido-Vera R, SukarovaStefanovska E, Kučinskas V, Makukh H, Pamjav H, Quintana-Murci L, Netea MG, Bertranpetit J, Calafell F, Comas D. Recent Common Origin, Reduced Population Size, and Marked Admixture Have Shaped European Roma Genomes. *Mol Biol Evol.* 2020 Nov 1;37(11):3175-3187. doi: 10.1093/molbev/msaa156. PMID: 32589725.

Bolnick DI, Nosil P. Natural selection in populations subject to a migration load. *Evolution.* 2007 Sep;61(9):2229-43. doi: 10.1111/j.1558-5646.2007.00179.x. PMID: 17767592.

Bottenus RE, Ichinose A, Davie EW. Nucleotide sequence of the gene for the b subunit of human factor XIII. *Biochemistry.* 1990 Dec 25;29(51):11195-209. doi: 10.1021/bi00503a007. PMID: 2271707.

Bouzekri N, Zhu X, Jiang Y, McKenzie CA, Luke A, Forrester T, Adeyemo A, Kan D, Farrall M, Anderson S, Cooper RS, Ward R. Angiotensin I-converting enzyme polymorphisms, ACE level and blood pressure among Nigerians, Jamaicans and African-Americans. *Eur J Hum Genet.* 2004 Jun;12(6):460-8. doi: 10.1038/sj.ejhg.5201166. PMID: 14970846.

Budowle B, Chakraborty R. Population variation at the CODIS core short tandem repeat loci in Europeans. *Leg Med (Tokyo).* 2001 Mar;3(1):29-33. doi: 10.1016/s1344-6223(01)00008-6. PMID: 12935730.

Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *J Forensic Sci.* 1999 Nov;44(6):1277-86. Erratum in: *J Forensic Sci.* 2015 Jul;60(4):1114-6. PMID: 10582369.

Budowle B, Moretti T.R, Niezgoda S.J, Brown B.L. 1998. CODIS and PCR-Based Short Tandem Repeat Loci: Law Enforcement Tools. Second european symposium on human identification. From:

<https://www.promega.com/~media/files/resources/conference%20proceedings/ishi%2002/oral%20presentations/17.pdf> (06/03/2022)

Butler, J. M, Budowle B, Gill P, Kidd KK, Phillips C, Schneider PM, Vallone PM, Morling N. 2008. Report on ISFG SNP panel discussion. *Forensic Science International: Genetics Supplement Series*, 1, 471–472.

Butler J.M, 2011. *Advanced Topics in Forensic DNA Typing: Methodology*. Academic Press.

Butler JM, Coble MD, Vallone PM. STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol*. 2007 Sep;3(3):200-5. doi: 10.1007/s12024-007-0018-1. Epub 2007 Sep 12. PMID: 25869164.

Butler J.M, 2004. *Forensic DNA Typing. Biology, Technology, and Genetics of STR Markers*. Elsevier

Butler JM, John M. 2011. *Advanced Topics in Forensic DNA Typing: Methodology*. San Diego: Elsevier Academic Press. pp. 99–100.

Bulbul O, Speed WC, Gurkan C, Soundararajan U, Rajeevan H, Pakstis AJ, Kidd KK. Improving ancestry distinctions among South west Asian populations. *Forensic Sci Int Genet*. 2018 Jul;35:14-20. doi: 10.1016/j.fsigen.2018.03.010. Epub 2018 Mar 23. PMID: 29625264.

Cabezas Silva R, Ribeiro T, Lucas I, Porto MJ, Costa Santos J, Dario P. Analysis of 17 STR data on 5362 southern Portuguese individuals-an update on reference database. *Forensic Sci Int Genet*. 2016 Mar;21:e10-2. doi: 10.1016/j.fsigen.2015.11.007. Epub 2015 Nov 22. PMID: 26651434.

Cabrera VM. Human molecular evolutionary rate, time dependency and transient polymorphism effects viewed through ancient and modern mitochondrial DNA genomes. *Sci Rep*. 2021 Mar 3;11(1):5036. doi: 10.1038/s41598-021-84583-1. PMID: 33658608; PMCID: PMC7930196.

Cao L, Chen C, Zhu H, Gu X, Deng D, Tian X, Liu J, Xiao Q. MMP16 is a marker of poor prognosis in gastric cancer promoting proliferation and invasion. *Oncotarget*. 2016 Aug 9;7(32):51865-51874. doi: 10.18632/oncotarget.10177. PMID: 27340864; PMCID: PMC5239520.

Cavalli-Sforza, L. L., Cavalli-Sforza, L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton University Press.

Cavalli-Sforza LL, Piazza A. Human genomic diversity in Europe: a summary of recent research and prospects for the future. *Eur J Hum Genet*. 1993;1(1):3-18. doi: 10.1159/000472383. PMID: 7520820.

Central Statistical Office (CSO) Hungary-census 2011.
<https://www.ksh.hu/nepszamlalas/?lang=en> (23/05/2022)

Chaix, R, Austerlitz, F, Morar, B, Kalaydjieva, L, & Heyer, E. (2004). Vlach Roma history: what do coalescent-based methods tell us? *European Journal of Human Genetics*, 12(4), 285-292.

Chambers GK, MacAvoy ES. Microsatellites: consensus and controversy. *Comp BiochemPhysiol B Biochem Mol Biol*. 2000 Aug;126(4):455-76. doi: 10.1016/s0305-0491(00)00233-9. PMID: 11026658.

Chaubey G (2010) The demographic history of India: A perspective based on genetic evidence. *Universitatis Tartuensis, Estonia*. Ph. D Thesis.185p.

Chen JQ, Cammarata PR, Baines CP, Yager JD. Regulation of mitochondrial respiratory chain biogenesis by estrogens/estrogen receptors and physiological pathological and pharmacological implications. *BiochimBiophys Acta*. 2009 Oct;1793(10):1540-70. doi: 10.1016/j.bbamcr.2009.06.001. Epub 2009 Jun 23. PMID: 19559056; PMCID: PMC2744640.

Chen K, Guo MR, Zhang Y, Li G, Liu Y, Zhang B. Association between MMP16 rs60298754 and clinical phenotypes of Parkinson's disease in southern Chinese. *Neurol Sci*. 2021 Aug;42(8):3211-3215. doi: 10.1007/s10072-020-04894-5. Epub 2020 Nov 25. PMID: 33241532.

Chen K, Zhou YX, Li K, Qi LX, Zhang QF, Wang MC, Xiao JH. A novel three-round multiplex PCR for SNP genotyping with next generation sequencing. *AnalBioanal Chem*. 2016 Jun;408(16):4371-7. doi: 10.1007/s00216-016-9536-6. Epub 2016 Apr 25. PMID: 27113460.

Chen YF, Chen WJ, Lin XZ, Zhang QJ, Cai JP, Liou CW, Wang N. Mitochondrial DNA Haplogroups and the Risk of Sporadic Parkinson's Disease in Han Chinese. *Chin Med J (Engl)*. 2015 Jul 5;128(13):1748-54. doi: 10.4103/0366-6999.159348. PMID: 26112715; PMCID: PMC4733725.

Cheung, K. H, Miller, P. L, Kidd, J. R, Kidd, K. K, Osier, M. V, Pakstis, A. J. (1999). ALFRED: a Web-accessible allele frequency database. In *Biocomputing 2000* (pp. 639-650).

Claude C, Elspeth G. (2008) .Recent Migration of Roma in Europe 2nd Edition October 2010

Ciccone L, Shi C, di Lorenzo D, Van Baelen AC, Tonali N. The Positive Side of the Alzheimer's Disease Amyloid Cross-Interactions: The Case of the A β 1-42 Peptide with Tau, TTR, CysC, and ApoA1. *Molecules*. 2020 May 23;25(10):2439. doi: 10.3390/molecules25102439. PMID: 32456156; PMCID: PMC7288020.

Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet*. 2000 Oct;107(4):312-9. doi: 10.1007/s004390000370. PMID: 11129330.

Comas, D, Schmid, H, Braeuer, S, Flaiz, C, Busquets, A, Calafell, F, Bertranpetit, J, Scheil, H.G, Huckenbeck, W, Efremovska, L, Schmidt, H, 2004, Alu insertion polymorphisms in the Balkans and the origins of the Aromuns. *Ann. Hum. Genet*, 68,120–127

Cooper RS, Tayo B, Zhu X. Genome-wide association studies: implications for multiethnic samples. *Hum Mol Genet*. 2008 Oct 15;17(R2):R151-5. doi: 10.1093/hmg/ddn263. PMID: 18852204; PMCID: PMC2782359.

Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 2009 Oct;10(10):691-703. doi: 10.1038/nrg2640. PMID: 19763152; PMCID: PMC2884099.

Cotrim NH, Auricchio MT, Vicente JP, Otto PA, Mingroni-Netto RC. Polymorphic Alu insertions in six Brazilian African-derived populations. *Am J Hum Biol.* 2004 May-Jun;16(3):264-77. doi: 10.1002/ajhb.20024. PMID: 15101052.

Croatian Bureau of statistics-census 2011. [https://web.dzs.hr/arhiva_e.htm\(24/05/2022\)](https://web.dzs.hr/arhiva_e.htm(24/05/2022))

Decade of Roma Inclusion. Various reports. 2005-2015.
<https://www.rcc.int/romaintegration2020/romadecade> (27/05/2022)

Degen SJ, Rajput B, Reich E. The human tissue plasminogen activator gene. *J Biol Chem.* 1986 May 25;261(15):6972-85. PMID: 3009482.

Deininger PL, Batzer MA. Alu repeats and human disease. *Mol Genet Metab.* 1999 Jul;67(3):183-93. doi: 10.1006/mgme.1999.2864. PMID: 10381326.

de la Puente M, Santos C, Fondevila M, Manzo L; EUROFORGEN-NoE Consortium, Carracedo Á, Lareu MV, Phillips C. The Global AIMs Nano set: A 31-plex SNaPshot assay of ancestry-informative SNPs. *Forensic Sci Int Genet.* 2016 May;22:81-88. doi: 10.1016/j.fsigen.2016.01.015. Epub 2016 Jan 25. PMID: 26881328.

Der Sarkissian C, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, Koshelev S, Zaporozhchenko V, Gronenborn D, Moiseyev V, Kolpakov E, Shumkin V, Alt KW, Balanovska E, Cooper A, Haak W; Genographic Consortium. Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. *PLoS Genet.* 2013;9(2):e1003296. doi: 10.1371/journal.pgen.1003296. Epub 2013 Feb 14. PMID: 23459685; PMCID: PMC3573127.

Desagher S, Martinou JC. Mitochondria as the central control point of apoptosis. *Trends Cell Biol.* 2000 Sep;10(9):369-77. doi: 10.1016/s0962-8924(00)01803-1. PMID: 10932094.

Di Pasquale P, Cannizzaro S, Paterna S (2004) Does angiotensin-converting enzyme gene polymorphism affect blood pressure? Findings after 6 years of follow-up in healthy subjects. *Eur J Heart Fail* 6: 11–16.

Dmitrenko OP, Karpova NS, Nurbekov MK, Papysheva OV. I/D Polymorphism Gene ACE and Risk of Preeclampsia in Women with Gestational Diabetes Mellitus. *Dis Markers.* 2020;2020:8875230. Published 2020 Dec 28. doi:10.1155/2020/8875230

DNA Analysis Backlog Elimination Act of 2000, Pub Law 106–546, 114 Stat. 2726–37

DNA Fingerprinting Act of 2005, Pub Law 109–162, 119 Stat. 2960, 3085

DNA Identification Act of 1994, Pub Law 103–322, 108 Stat. 1796, 2065–71.

Egyed B, Füredi S, Angyal M, Boutrand L, Vandenberghe A, Woller J, Pádár Z. Analysis of eight STR loci in two Hungarian populations. *Forensic Sci Int.* 2000 Sep 11;113(1-3):25-7. doi: 10.1016/s0379-0738(00)00191-2. PMID: 10978596.

Egyed B, Füredi S, Angyal M, Balogh I, Kalmar L, Padar Z. Analysis of the population heterogeneity in Hungary using fifteen forensically informative STR markers. *Forensic Sci Int.* 2006 May 10;158(2-3):244-9. doi: 10.1016/j.forsciint.2005.07.004. Epub 2005 Sep 6. PMID: 16143479.

Ehler E, Vanek D. Forensic genetic analyses in isolated populations with examples of central European Valachs and Roma. *J Forensic Leg Med.* 2017 May;48:46-52. doi: 10.1016/j.jflm.2017.04.001. Epub 2017 Apr 13. PMID: 28454050.

Eisenmann JC, Sarzynski MA, Glenn K, Rothschild M, Heelan KA (2009) ACE I/D genotype, adiposity, and blood pressure in children. *Cardiovasc Diabetol* 8: 14.

Elfenbein, J. H. (1987). A periploous of the 'Brahui problem'. *Studia Iranica.* 16: 215-233

Ellegren H. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat Genet.* 2000 Apr;24(4):400-2. doi: 10.1038/74249. PMID: 10742106.

EuropeanCommission. Greece-census 2011. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/roma-eu/roma-equality-inclusion-and-participation-eu-country/greece_en (23/05/2022)

EuropeanComission-Portugal 2021. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/roma-eu/roma-equality-inclusion-and-participation-eu-country/portugal_en (24/05/2022)

EuropeanCommission. Spain-census 2011. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/roma-eu/roma-equality-inclusion-and-participation-eu-country/spain_en#:~:text=ERDF%20operational%20programmes,-,Facts%20and%20figures,1.57%25%20of%20the%20population (23/05/2022)

Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol EcolResour.* 2010 May;10(3):564-7. doi: 10.1111/j.1755-0998.2010.02847.x.Epub 2010 Mar 1. PMID: 21565059.

Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* 1992 Jun;131(2):479-91. doi: 10.1093/genetics/131.2.479. PMID: 1644282; PMCID: PMC1205020.

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci U S A.* 2007 Nov 6;104(45):17614-9. doi: 10.1073/pnas.0708280104. Epub 2007 Oct 31. PMID: 17978179; PMCID: PMC2077041.

- Fan H, Chu JY. A brief review of short tandem repeat mutation. *Genomics Proteomics Bioinformatics*. 2007 Feb;5(1):7-14. doi: 10.1016/S1672-0229(07)60009-6. PMID: 17572359; PMCID: PMC5054066.
- Fan H, Wang X, Chen H, Li W, Wang W, Deng J. The Ong Be language-speaking population in Hainan Island: genetic diversity, phylogenetic characteristics and reflections on ethnicity. *Mol Biol Rep*. 2019 Aug;46(4):4095-4103. doi: 10.1007/s11033-019-04859-8. Epub 2019 May 16. PMID: 31098804.
- Farha S, Hu B, Comhair S, Zein J, Dweik R, Erzurum SC, Aldred MA. Mitochondrial Haplogroups and Risk of Pulmonary Arterial Hypertension. *PLoS One*. 2016 May 25;11(5):e0156042. doi: 10.1371/journal.pone.0156042. PMID: 27224443; PMCID: PMC4880300.
- Fehér T, Németh E, Vándor A, Kornienko IV, Csáji LK, Pamjav H. Y-SNP L1034: limited genetic link between Mansi and Hungarian-speaking populations. *Mol Genet Genomics*. 2015 Feb;290(1):377-86. doi: 10.1007/s00438-014-0925-2. Epub 2014 Sep 26. PMID: 25258186.
- Feng Y, Niu T, Xu X, Chen C, Li Q, et al. (2002) Insertion/deletion polymorphism of the ACE gene is associated with type 2 diabetes. *Diabetes* 51: 1986–1988.
- Fisher R.A. 1930. *The genetical theory of natural selection*. Oxford: Clarendon.
- Flores-Bello A, Bauduer F, Salaberria J, Oyharçabal B, Calafell F, Bertranpetit J, Quintana-Murci L, Comas D. Genetic origins, singularity, and heterogeneity of Basques. *Curr Biol*. 2021 May 24;31(10):2167-2177.e4. doi: 10.1016/j.cub.2021.03.010. Epub 2021 Mar 25. PMID: 33770488.
- Flores C, Maca-Meyer N, Larruga JM, Cabrera VM, Karadsheh N, Gonzalez AM. Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan. *J Hum Genet*. 2005;50(9):435-441. doi: 10.1007/s10038-005-0274-4. Epub 2005 Sep 2. PMID: 16142507.
- Foley RA, Martin L, Mirazón Lahr M, Stringer C. Major transitions in human evolution. *Philos Trans R Soc Lond B Biol Sci*. 2016 Jul 5;371(1698):20150229. doi: 10.1098/rstb.2015.0229. PMID: 27298461; PMCID: PMC4920287.
- Font-Porterías N, Arauna LR, Poveda A, Bianco E, Rebato E, Prata MJ, Calafell F, Comas D. European Roma groups show complex West Eurasian admixture footprints and a common South Asian genetic origin. *PLoS Genet*. 2019 Sep 23;15(9):e1008417. doi: 10.1371/journal.pgen.1008417. PMID: 31545809; PMCID: PMC6779411.
- Font-Porterías N, Giménez A, Carballo-Mesa A, Calafell F, Comas D. Admixture Has Shaped Romani Genetic Diversity in Clinically Relevant Variants. *Front Genet*. 2021 Jun 16;12:683880. doi: 10.3389/fgene.2021.683880. PMID: 34220960; PMCID: PMC8244592.
- Fraser A (1992) *The Gypsies*. Oxford: Blackwell Publishers Ltd. 359p.
- FSG Annual Report (2020). *Discrimination and the Roma Community*. Fundación Secretariado Gitano (FSG). Serie Cuadernos Técnicos nº 133. Madrid, 2020. 79 pgs.

- Füredi S, Angyal M, Kozma Z, Sétáló J, Woller J, Pádár Z. Semi-automatic DNA profiling in a Hungarian Romany population using the STR loci HumVWFA31, HumTH01, HumTPOX, and HumCSF1PO. *Int J Legal Med.* 1997;110(4):184-7. doi: 10.1007/s004140050064. PMID: 9274941.
- Füredi S, Budowle B, Woller J, Pádár Z. Hungarian population data on six STR loci--HUMVWFA31, HUMTH01, HUMCSF1PO, HUMFES/FPS, HUMTPOX, and HUMHPRTB--derived using multiplex PCR amplification and manual typing. *Int J Legal Med.* 1996;109(2):100-1. doi: 10.1007/BF01355526. PMID: 8912057.
- Gal-Mark N, Schwartz S, Ast G. Alternative splicing of Alu exons--two arms are better than one. *Nucleic Acids Res.* 2008 Apr;36(6):2012-23. doi: 10.1093/nar/gkn024. Epub 2008 Feb 14. PMID: 18276646; PMCID: PMC2330237.
- García-Fernández C, Font-Porterías N, Kučinskas V, Sukarova-Stefanovska E, Pamjav H, Makukh H, Dobon B, Bertranpetit J, Netea MG, Calafell F, Comas D. Sex-biased patterns shaped the genetic history of Roma. *Sci Rep.* 2020 Sep 2;10(1):14464. doi: 10.1038/s41598-020-71066-y. Erratum in: *Sci Rep.* 2020 Oct 20;10(1):18142. PMID: 32879340; PMCID: PMC7468237.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, de Pancorbo MM, Peña JA. Polymorphic Alu insertions and the genetic structure of Iberian Basques. *J Hum Genet.* 2007;52(4):317-327. doi: 10.1007/s10038-007-0114-9. Epub 2007 Feb 3. PMID: 17277899.
- García-Obregón S, Alfonso-Sánchez M.A, Pérez-Miranda, A.M, Vidales, C, Arroyo, D, Peña, J.A, 2006, Genetic position of Valencia (Spain) in the Mediterranean basin according to Alu insertions. *Am. J. Hum. Biol.* 18,187–195
- Ge J, Eisenberg A, Budowle B. Developing criteria and data to determine best options for expanding the core CODIS loci. *Investig Genet.* 2012 Jan 6;3:1. doi: 10.1186/2041-2223-3-1. PMID: 22226306; PMCID: PMC3314575.
- Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. The NCBI BioSystems database. *Nucleic Acids Res.* 2010 Jan;38(Database issue):D492-6. doi: 10.1093/nar/gkp858. Epub 2009 Oct 23. PMID: 19854944; PMCID: PMC2808896.
- Genome. National Human Genome Research Institute,NIH. Recuperado de: <https://www.genome.gov/es/genetics-glossary/Genoma#:~:text=En%20los%20seres%20humanos%2C%20el,de%20la%20secuencia%20de%20ADN.> [27/01/2022]
- Geppert M, Baeta M, Núñez C, Martínez-Jarreta B, Zweynert S, Cruz OW, González-Andrade F, González-Solorzano J, Nagy M, Roewer L. Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. *Forensic Sci Int Genet.* 2011 Mar;5(2):100-4. doi: 10.1016/j.fsigen.2010.08.016. Epub 2010 Oct 6. PMID: 20932815.
- Gharesouran J, Hosseinzadeh H, Ghafouri-Fard S, Taheri M, Rezazadeh M. STRs: Ancient Architectures of the Genome beyond the Sequence. *J Mol Neurosci.* 2021 Dec;71(12):2441-2455. doi: 10.1007/s12031-021-01850-6. Epub 2021 May 30. PMID: 34056692.

- Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A*. 1980 Nov;77(11):6715-9. doi: 10.1073/pnas.77.11.6715. PMID: 6256757; PMCID: PMC350359.
- Glazko G, Veeramachaneni V, Nei M, Makiłowski W. Eighty percent of proteins are different between humans and chimpanzees. *Gene*. 2005 Feb 14;346:215-9. doi: 10.1016/j.gene.2004.11.003. PMID: 15716009.
- Gould SJ (2002). "Chapter 7, section "Synthesis as Hardening"". *The Structure of Evolutionary Theory*.
- Gouva M, Mentis M, Kotrotsiou S, Paralikas T, Kotrotsiou E. Shame and Anxiety Feelings of a Roma Population in Greece. *J ImmigrMinorHealth*. 2015 Dec;17(6):1765-70. doi: 10.1007/s10903-014-9979-9. PMID: 24845934.
- Graves JA, Wakefield MJ, Toder R. The origin and evolution of the pseudoautosomal regions of human sex chromosomes. *Hum Mol Genet*. 1998 Dec;7(13):1991-6. doi: 10.1093/hmg/7.13.1991. PMID: 9817914.
- Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kučinskas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L. Origins and divergence of the Roma (gypsies). *Am J Hum Genet*. 2001 Dec;69(6):1314-31. doi: 10.1086/324681. Epub 2001 Nov 9. PMID: 11704928; PMCID: PMC1235543.
- Gusmão A, Valente C, Gomes V, Alves C, Amorim A, Prata MJ, Gusmão L. A genetic historical sketch of European Gypsies: The perspective from autosomal markers. *Am J Phys Anthropol*. 2010 Apr;141(4):507-14. doi: 10.1002/ajpa.21166. PMID: 19918999.
- Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q, Mittnik A, Bánffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szécsényi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2015 Jun 11;522(7555):207-11. doi: 10.1038/nature14317. Epub 2015 Mar 2. PMID: 25731166; PMCID: PMC5048219.
- Habbane M, Montoya J, Rhouda T, Sbaoui Y, Radallah D, Emperador S. Human Mitochondrial DNA: Particularities and Diseases. *Biomedicines*. 2021 Oct 1;9(10):1364. doi: 10.3390/biomedicines9101364. PMID: 34680481; PMCID: PMC8533111.
- Haber M, Platt DE, AshrafianBonab M, Youhanna SC, Soria-Hernanz DF, Martínez-Cruz B, Douaihy B, Ghassibe-Sabbagh M, Rafatpanah H, Ghanbari M, Whale J, Balanovsky O, Wells RS, Comas D, Tyler-Smith C, Zalloua PA; Genographic Consortium. Afghanistan's ethnic groups share a Y-chromosomal heritage structured by historical events. *PLoS One*. 2012;7(3):e34288.

doi: 10.1371/journal.pone.0034288. Epub 2012 Mar 28. PMID: 22470552; PMCID: PMC3314501.

Hallast P, Agdzhoyan A, Balanovsky O, Xue Y, Tyler-Smith C. A SoutheastAsianoriginforpresent-day non-African human Y chromosomes. *Hum Genet.* 2021 Feb;140(2):299-307. doi: 10.1007/s00439-020-02204-9. Epub 2020 Jul 14. PMID: 32666166; PMCID: PMC7864842.

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet.* 1999 Jul;22(3):239-47. doi: 10.1038/10297. PMID: 10391210.

Hammer Ø, Harper DAT, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. 2001 *Palaeontologia Electronica* 4, 9.

Hamilton RL, Williams MC, Fielding CJ, Havel RJ. Discoidal bilayer structure of nascent high density lipoproteins from perfused rat liver. *J Clin Invest.* 1976 Sep;58(3):667-80. doi: 10.1172/JCI108513. PMID: 182724; PMCID: PMC333225.

Hancock IF (2002) *We Are the Romani People*. United Kingdom: University Of Hertfordshire Press. 288p.

Hansen MM, Skaala O, Jensen LF, Bekkevold D, Mensberg KL. Gene flow, effective population size and selection at major histocompatibility complex genes: brown trout in the Hardanger Fjord, Norway. *Mol Ecol.* 2007 Apr;16(7):1413-25. doi: 10.1111/j.1365-294X.2007.03255.x. PMID: 17391266.

Haliti N, Carapina M, Masić M, Strinović D, Klarić IM, Kubat M. Evaluation of population variation at 17 autosomal STR and 16 Y-STR haplotype loci in Croatians. *Forensic Sci Int Genet.* 2009 Sep;3(4):e137-8. doi: 10.1016/j.fsigen.2008.11.004. Epub 2008 Dec 27. PMID: 19647700.

Hares DR. Expanding the CODIS core loci in the United States. *Forensic Sci Int Genet.* 2012 Jan;6(1):e52-4. doi: 10.1016/j.fsigen.2011.04.012. Epub 2011 May 4. Erratum in: *Forensic Sci Int Genet.* 2012 Sep;6(5):e135. PMID: 21543275.

Harpending HC, Ward RH. Chemical systematics and human populations. En: Nitecki M, editor. *Biochemical aspects of evolutionary biology*. 1982. Chicago: University of Chicago Press. p 213-256

Harpending HC, Jenkins T. Genetic distance among southern African population. 1973. En: Crawford M, Workman P (Eps). *Method and Theory in Anthropological Genetics*. Albuquerque. USA: University of New Mexico Press. Pp 177-199

Harvati K, Röding C, Bosman AM, Karakostis FA, Grün R, Stringer C, Karkanas P, Thompson NC, Koutoulidis V, Mouloupoulos LA, Gorgoulis VG, Kouloukoussa M. Apidima Cave fossils provide earliest evidence of *Homo sapiens* in Eurasia. *Nature.* 2019 Jul;571(7766):500-504. doi: 10.1038/s41586-019-1376-z. Epub 2019 Jul 10. PMID: 31292546.

Häsler J, Strub K. Alu elements as regulators of gene expression [published correction appears in *Nucleic Acids Res.* 2007;35(4):1389]. *Nucleic Acids Res.* 2006;34(19):5491-5497.

doi:10.1093/nar/gkl706

Havas D, Jeran N, Efremovska L, Dordević D, Rudan P. Population genetics of 15 Ampfl STRI dentifiler loci in Macedonians and Macedonian Romani (Gypsy). *Forensic Sci Int.* 2007 Dec 20;173(2-3):220-4. doi: 10.1016/j.forsciint.2006.10.027. Epub 2007 Feb 20. PMID: 17307318.

Hawks J, Hunley K, Lee SH, Wolpoff M. Population bottlenecks and Pleistocene human evolution. *Mol Biol Evol.* 2000 Jan;17(1):2-22. doi: 10.1093/oxfordjournals.molbev.a026233. PMID: 10666702.

Haywood, J. (2008). *The great migrations: From the earliest humans to the age of globalization.* Toledo (USA): Quercus.

Hedges DJ, Callinan PA, Cordaux R, Xing J, Barnes E, Batzer MA. Differential alu mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* 2004 Jun;14(6):1068-75. doi: 10.1101/gr.2530404. PMID: 15173113; PMCID: PMC419785.

Helgason A, Einarsson AW, Guðmundsdóttir VB, Sigurðsson Á, Gunnarsdóttir ED, Jagadeesan A, Ebenesersdóttir SS, Kong A, Stefánsson K. The Y-chromosome point mutation rate in humans. *Nat Genet.* 2015 May;47(5):453-7. doi: 10.1038/ng.3171. Epub 2015 Mar 25. PMID: 25807285.

Hidalgo PC, Mut P, Ackermann E, Figueiro G, Sans M. Questioning the "melting pot": analysis of Alu inserts in three population samples from Uruguay. *Hum Biol.* 2014 Spring;86(2):83-92. doi: 10.3378/027.086.0202. PMID: 25397699.

Hopkins C, Taylor D, Hill K, Henry J. Analysis of the South Australian Aboriginal population using the Global AIMS Nano ancestry test. *Forensic Sci Int Genet.* 2019 Jul;41:34-41. doi: 10.1016/j.fsigen.2019.03.020. Epub 2019 Mar 23. PMID: 30952105.

Hout MC, Papesh MH, Goldinger SD. Multidimensional scaling. *Wiley Interdiscip Rev Cogn Sci.* 2013 Jan;4(1):93-103. doi: 10.1002/wcs.1203. Epub 2012 Oct 8. PMID: 23359318; PMCID: PMC3555222.

Huerta-Chagoya A, Moreno-Macías H, Fernández-López JC, Ordóñez-Sánchez ML, Rodríguez-Guillén R, Contreras A, Hidalgo-Miranda A, Alfaro-Ruíz LA, Salazar-Fernandez EP, Moreno-Estrada A, Aguilar-Salinas CA, Tusié-Luna T. A panel of 32 AIMS suitable for population stratification correction and global ancestry estimation in Mexican mestizos. *BMC Genet.* 2019 Jan 8;20(1):5. doi: 10.1186/s12863-018-0707-7. PMID: 30621578; PMCID: PMC6323778.

Hu XS, Yeh FC, Hu Y, Deng LT, Ennos RA, Chen X. High mutation rates explain low population genetic divergence at copy-number-variable loci in *Homo sapiens*. *Sci Rep.* 2017 Feb 22;7:43178. doi: 10.1038/srep43178. PMID: 28225073; PMCID: PMC5320550.

Hwa HL, Wu LS, Lin CY, Huang TY, Yin HI, Tseng LH, Lee JC. Genotyping of 75 SNPs using arrays for individual identification in five population groups. *Int J Legal Med.* 2016 Jan;130(1):81-9. doi: 10.1007/s00414-015-1250-y. Epub 2015 Aug 22. PMID: 26297200.

Iovita RP, Schurr TG (2004) Reconstructing the Origins and Migrations of Diasporic Populations: The Case of the European Gypsies. *American Anthropologist* 106: 267–281.

Itano MS, Arnion H, Wolin SL, Simon SM. Recruitment of 7SL RNA to assembling HIV-1 virus-like particles. *Traffic*. 2018 Jan;19(1):36-43. doi: 10.1111/tra.12536. Epub 2017 Dec 1. PMID: 29044909; PMCID: PMC6781622.

Jiang L, Wei YL, Zhao L, Li N, Liu T, Liu HB, Ren LJ, Li JL, Hao HF, Li Q, Li CX. Global analysis of population stratification using a smart panel of 27 continental ancestry-informative SNPs. *Forensic Sci Int Genet*. 2018 Jul;35:e10-e12. doi: 10.1016/j.fsigen.2018.05.006. Epub 2018 May 18. PMID: 29803513.

Jin XY, Cui W, Chen C, Guo YX, Tao YW, Lan Q, Kong TT, Zhu BF. Biogeographic origin prediction of three continental populations through 42 ancestry informative SNPs. *Electrophoresis*. 2020 Feb;41(3-4):235-245. doi: 10.1002/elps.201900241. Epub 2019 Nov 29. PMID: 31743462.

Jin X, Dimitriadis EK, Liu Y, Combs CA, Chang J, Varsano N, Stempinski E, Flores R, Jackson SN, Muller L, Woods AS, Addadi L, Kruth HS. Macrophages Shed Excess Cholesterol in Unique Extracellular Structures Containing Cholesterol Microdomains. *ArteriosclerThrombVasc Biol*. 2018 Jul;38(7):1504-1518. doi: 10.1161/ATVBAHA.118.311269. Epub 2018 May 31. PMID: 29853567; PMCID: PMC6023747.

Jobling MA. Forensic genetics through the lens of Lewontin: population structure, ancestry and race. *Philos Trans R Soc Lond B Biol Sci*. 2022 Jun 6;377(1852):20200422. doi: 10.1098/rstb.2020.0422. Epub 2022 Apr 18. PMID: 35430883; PMCID: PMC9014189.

Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet*. 2003 Aug;4(8):598-612. doi: 10.1038/nrg1124. PMID: 12897772.

Johnson DC, Shrestha S, Wiener HW, Makowsky R, Kurundkar A, Wilson CM, Aissani B. Mitochondrial DNA diversity in the African American population. *Mitochondrial DNA*. 2015 Jun;26(3):445-51. doi: 10.3109/19401736.2013.840591. Epub 2013 Oct 9. PMID: 24102597; PMCID: PMC4048334.

Jorde PE, Ryman N. Unbiased estimator for genetic drift and effective population size. *Genetics*. 2007 Oct;177(2):927-35. doi: 10.1534/genetics.107.075481. Epub 2007 Aug 24. PMID: 17720927; PMCID: PMC2034655.

Jung JY, Kang PW, Kim E, Chacon D, Beck D, McNevin D. Ancestry informative markers (AIMs) for Korean and other East Asian and South East Asian populations. *Int J Legal Med*. 2019 Nov;133(6):1711-1719. doi: 10.1007/s00414-019-02129-7. Epub 2019 Aug 7. PMID: 31388795.

Justice for All Act of 2004, Pub Law 108–405, 118 Stat. 2260.

Kajikawa M, Okada N. LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell*. 2002 Nov 1;111(3):433-44. doi: 10.1016/s0092-8674(02)01041-3. PMID: 12419252.

- Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH, Hurles ME, Underhill P, Tournev I, Marushiakova E, Popov V. Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur J Hum Genet.* 2001a Feb;9(2):97-104. doi: 10.1038/sj.ejhg.5200597. PMID: 11313742.
- Kalaydjieva, L, Gresham, D,&Calafell, F. (2001b). Genetic studies of the Roma (Gypsies): a review. *BMC Medical Genetics*, 2(1).
- Kalaydjieva, L, Morar, B, Chaix, R, & Tang, H. (2005). A newly discovered founder population: the Roma/Gypsies. *BioEssays*, 27(10), 1084-1094.
- Kapitonov V, Jurka J. The age of Alu subfamilies. *J Mol Evol.* 1996 Jan;42(1):59-65. doi: 10.1007/BF00163212. PMID: 8576965.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 2008 May;18(5):830-8. doi: 10.1101/gr.7172008. Epub 2008 Apr 2. PMID: 18385274; PMCID: PMC2336805.
- Karathanasis SK. Apolipoprotein multigene family: tandem organization of human apolipoprotein AI, CIII, and AIV genes. *Proc Natl Acad Sci U S A.* 1985 Oct;82(19):6374-8. doi: 10.1073/pnas.82.19.6374. PMID: 3931073; PMCID: PMC390718.
- Kass DH, Aleman C, Batzer MA, Deininger PL. Identification of a human specific Alu insertion in the factor XIII B gene. *Genetica.* 1994;94(1):1-8. doi: 10.1007/BF01429214. PMID: 7729692.
- Katsanis SH, Wagner JK. Characterization of the standard and recommended CODIS markers. *J Forensic Sci.* 2013 Jan;58 Suppl 1(Suppl 1):S169-72. doi: 10.1111/j.1556-4029.2012.02253.x. Epub 2012 Aug 24. PMID: 22925064; PMCID: PMC8591976.
- Kee BP, Chua KH, Lee PC, Lian LH. Population data of six Alu insertions in indigenous groups from Sabah, Malaysia. *Ann Hum Biol.* 2012 Nov-Dec;39(6):505-10. doi: 10.3109/03014460.2012.719548. Epub 2012 Sep 18. PMID: 22989108.
- Keene SE, Telesnitsky A. cis-Acting determinants of 7SL RNA packaging by HIV-1. *J Virol.* 2012 Aug;86(15):7934-42. doi: 10.1128/JVI.00856-12. Epub 2012 May 16. PMID: 22593161; PMCID: PMC3421668.
- Kemény, I., Janky, B. (2005). Roma Population of Hungary 1971-2003. In: Roma of Hungary, edited by Kemény, I., Atlantic Research and Publication, Highland Lakes, 2005, pp. 70–225
- Khurana P, Aggarwal A, Mitra S, Italia YM, Saraswathy KN, Chandrasekar A, Kshatriya GK. Y chromosome haplogroup distribution in Indo-European speaking tribes of Gujarat, western India. *PLoS One.* 2014 Mar 10;9(3):e90414. doi: 10.1371/journal.pone.0090414. PMID: 24614885; PMCID: PMC3948632.
- Kidd KK, Bulbul O, Gurkan C, Dogan M, Dogan S, Neophytou PI, Cherni L, Gurwitz D, Speed WC, Murtha M, Kidd JR, Pakstis AJ. Genetic relationships of Southwest Asian and Mediterranean

populations. *Forensic Sci Int Genet.* 2021 Jul;53:102528. doi: 10.1016/j.fsigen.2021.102528. Epub 2021 May 14. PMID: 34020230.

Kidd KK, Speed WC, Pakstis AJ, Furtado MR, Fang R, Madbouly A, Maiers M, Middha M, Friedlaender FR, Kidd JR. Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Sci Int Genet.* 2014 May;10:23-32. doi: 10.1016/j.fsigen.2014.01.002. Epub 2014 Jan 15. PMID: 24508742.

Kim SH, Shahani N, Bae BI, Sbodio JI, Chung Y, Nakaso K, Paul BD, Sawa A. Allele-specific regulation of mutant Huntingtin by Wig1, a downstream target of p53. *Hum Mol Genet.* 2016 Jun 15;25(12):2514-2524. doi: 10.1093/hmg/ddw115. Epub 2016 May 19. PMID: 27206983; PMCID: PMC6086561.

Kivisild T. Maternal ancestry and population history from whole mitochondrial genomes. *Investig Genet.* 2015 Mar 10;6:3. doi: 10.1186/s13323-015-0022-2. PMID: 25798216; PMCID: PMC4367903.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioğlu C, King R, Cavalli-Sforza L, Underhill PA, Villems R. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet.* 2003 Feb;72(2):313-32. doi: 10.1086/346068. Epub 2003 Jan 20. PMID: 12536373; PMCID: PMC379225.

Kojima KK. Hagfish genome reveals parallel evolution of 7SL RNA-derived SINEs. *Mob DNA.* 2020;11:18. Published 2020 May 22. doi:10.1186/s13100-020-00210-2

Kong DD, Wang MH, Yang J, Li L, Wang W, Wang SB, Zhou YZ. T-cadherin is associated with prognosis in triple-negative breast cancer. *Oncol Lett.* 2017 Sep;14(3):2975-2981. doi: 10.3892/ol.2017.6505. Epub 2017 Jun 30. PMID: 28928835; PMCID: PMC5588130.

Kong DD, Yang J, Li L, Wang W, Chen YN, Wang SB, Zhou YZ. T-cadherin association with clinicopathological features and prognosis in axillary lymph node-positive breast cancer. *Breast Cancer Res Treat.* 2015 Feb;150(1):119-26. doi: 10.1007/s10549-015-3302-x. Epub 2015 Feb 13. PMID: 25677746.

Konkel MK, Walker JA, Batzer MA. LINEs and SINEs of primate evolution. *EvolAnthropol.* 2010;19(6):236-249. doi:10.1002/evan.20283

Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet.* 2007 Apr;23(4):158-61. doi: 10.1016/j.tig.2007.02.002. Epub 2007 Feb 20. PMID: 17307271.

Krzeminska-Ahmadzai U, Buckley B, Loake T, Nicholson C, Beesley D, Randall C. Population data for 23 autosomal STR loci in White British population. *Leg Med (Tokyo).* 2021 May;50:101863. doi: 10.1016/j.legalmed.2021.101863. Epub 2021 Feb 13. PMID: 33618085.

Kumar A, Kumar R, Kumawat RK, Shrivastava P, Yadav R, Chaubey G. Population genetic data of 22 autosomal STR loci for the Scheduled Caste population of Rajasthan India. *Ann Hum Biol.*

2021 Dec;48(7-8):598-604. doi: 10.1080/03014460.2021.2009030. Epub 2021 Dec 20. PMID: 34802358.

Kuzniar P, Ploski R. STR data for the power plex-16 loci in a population from Central Poland. *Forensic Sci Int.* 2004 Jan 28;139(2-3):261-3. doi: 10.1016/j.forsciint.2003.11.011. PMID: 15040927.

Lan Q, Fang Y, Mei S, Xie T, Liu Y, Jin X, Yang G, Zhu B. Next generation sequencing of a set of ancestry-informative SNPs: ancestry assignment of three continental populations and estimating ancestry composition for Mongolians. *Mol Genet Genomics.* 2020 Jul;295(4):1027-1038. doi: 10.1007/s00438-020-01660-2. Epub 2020 Mar 23. PMID: 32206883.

Langella, O (1999) Populations <http://bioinformatics.org/~tryphon/populations/>

Lantermann A, Hampe J, Kim WH, Winter TA, Kidd M, Nagy M, Fölsch UR, Schreiber S. Investigation of HLA-DPA1 genotypes as predictors of inflammatory bowel disease in the German, South African, and South Korean populations. *Int J Colorectal Dis.* 2002 Jul;17(4):238-44. doi: 10.1007/s00384-001-0382-3. Epub 2001 Dec 21. PMID: 12073072.

Larmuseau MHD, Otten GPPL, Decorte R, Van Damme P, Moisse M. Defining Y-SNP variation among the Flemish population (Western Europe) by full genome sequencing. *Forensic Sci Int Genet.* 2017 Nov;31:e12-e16. doi: 10.1016/j.fsigen.2017.10.008. Epub 2017 Oct 28. PMID: 29089250.

Laybourn S, Akam EC, Cox N, Singh P, Mastana SS. Genetic analysis of novel Alu insertion polymorphisms in selected Indian populations. *Am J Hum Biol.* 2016 Nov;28(6):941-944. doi: 10.1002/ajhb.22881. Epub 2016 Jun 13. PMID: 27292586

Leakey RE. Early homo sapiens remains from the Omo River region of south-west Ethiopia. *Nature.* 1969 Jun 21;222(5199):1132-4. doi: 10.1038/2221132a0. PMID: 5788977.

Lee AC, Kamalam A, Adams SM, Jobling MA. Molecular evidence for absence of Y-linkage of Hares DR. Expanding the CODIS core loci in the United States. *Forensic Sci Int Genet* 2011;6(1):e52-4.

Lermo, J, Román, J, Marrodán, M.D, Mesa, M.S. 2006, Modelos de distribución de apellidos en la población gitana española. *Antropo*, 13, 69-87. www.didac.ehu.es/antropo

LeVan TD, Bloom JW, Bailey TJ, Karp CL, Halonen M, Martinez FD, Vercelli D. A common single nucleotide polymorphism in the CD14 promoter decreases the affinity of Sp protein binding

and enhances transcriptional activity. *J Immunol.* 2001 Nov 15;167(10):5838-44. doi: 10.4049/jimmunol.167.10.5838. PMID: 11698458.

Li J, Zhang S, Wu L, Pei M, Jiang Y. Berberine inhibited metastasis through miR-145/MMP16 axis in vitro. *J Ovarian Res.* 2021 Jan 6;14(1):4. doi: 10.1186/s13048-020-00752-2. PMID: 33407764; PMCID: PMC7789793.

Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW. Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *Am J Hum Biol.* 2010 Mar-Apr;22(2):187-92. doi: 10.1002/ajhb.20976. PMID: 19639555.

Li X, Sun X, Jin L, Xue F (2011) Worldwide spatial genetic structure of angiotensin-converting enzyme gene: a new evolutionary ecological evidence for the thrifty genotype hypothesis. *Eur J Hum Genet* 19: 1002–1008.

Liégeois J-P. .Roma, Gypsies, Travellers, Council of Europe, Strasbourg (1994)

Liu, L, Johnson, H. L, Cousens, S, Perin, J, Scott, S, Lawn, J. E, Mathers, C. (2012). Global regional and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *The Lancet*, 379(9832), 2151-2161.

Liu Y, Sun CC, Ma M, Wang L, Zhao WT, Ma Q, Ji AQ, Liu J, Li CX. The ancestry inference of Chinese populations using 74-plex SNPs system. *Yi Chuan.* 2020 Mar 20;42(3):296-308. doi: 10.16288/j.ycz.19-252. PMID: 32217515.

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, Wang Z, Chinwalla AT, Minx P, Mitreva M, Cook L, Delehaunty KD, Fronick C, Schmidt H, Fulton LA, Fulton RS, Nelson JO, Magrini V, Pohl C, Graves TA, Markovic C, Cree A, Dinh HH, Hume J, Kovar CL, Fowler GR, Lunter G, Meader S, Heger A, Ponting CP, Marques-Bonet T, Alkan C, Chen L, Cheng Z, Kidd JM, Eichler EE, White S, Searle S, Vilella AJ, Chen Y, Flicek P, Ma J, Raney B, Suh B, Burhans R, Herrero J, Haussler D, Faria R, Fernando O, Darré F, Farré D, Gazave E, Oliva M, Navarro A, Roberto R, Capozzi O, Archidiacono N, Della Valle G, Purgato S, Rocchi M, Konkel MK, Walker JA, Ullmer B, Batzer MA, Smit AF, Hubley R, Casola C, Schridder DR, Hahn MW, Quesada V, Puente XS, Ordoñez GR, López-Otín C, Vinar T, Brejova B, Ratan A, Harris RS, Miller W, Kosiol C, Lawson HA, Taliwal V, Martins ALSiepel A, Roychoudhury A, Ma X, Degenhardt J, Bustamante CD, Gutenkunst RN, Mailund T, Dutheil JY, Hobolth A, Schierup MH, Ryder OA, Yoshinaga Y, de Jong PJ, Weinstock GM,

Lohrer HD, Tangen U. Investigations into the molecular effects of single nucleotide polymorphism. *Pathobiology.* 2000;68(6):283-90. doi: 10.1159/000055939. PMID: 11493762.

Luizon MR, Mendes-Junior CT, De Oliveira SF, Simões AL. Ancestry informative markers in Amerindians from Brazilian Amazon. *Am J Hum Biol.* 2008 Jan-Feb;20(1):86-90. doi: 10.1002/ajhb.20681. PMID: 17963227.

Luo S, Valencia CA, Zhang J, Lee NC, Slone J, Gui B, Wang X, Li Z, Dell S, Brown J, Chen SM, Chien YH, Hwu WL, Fan PC, Wong LJ, Atwal PS, Huang T. Biparental Inheritance of

- Mitochondrial DNA in Humans. *Proc Natl Acad Sci U S A*. 2018 Dec 18;115(51):13039-13044. doi: 10.1073/pnas.1810946115. Epub 2018 Nov 26. PMID: 30478036; PMCID: PMC6304937.
- Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet*. 2016 Oct 14;17(11):704-714. doi: 10.1038/nrg.2016.104. PMID: 27739533.
- Madar S, Harel E, Goldstein I, Stein Y, Kogan-Sakin I, Kamer I, Solomon H, Dekel E, Tal P, Goldfinger N, Friedlander G, Rotter V. Mutant p53 attenuates the anti-tumorigenic activity of fibroblasts-secreted interferon beta. *PLoS One*. 2013 Apr 22;8(4):e61353. doi: 10.1371/journal.pone.0061353. PMID: 23630584; PMCID: PMC3632588.
- Madsen BE, Villesen P, Wiuf C. Short tandem repeats in human exons: a target for disease mutations. *BMC Genomics*. 2008 Sep 12;9:410. doi: 10.1186/1471-2164-9-410. PMID: 18789129; PMCID: PMC2543027.
- Majumder PP, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N, Roy M, Thakurta PG, Sil SK. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur J Hum Genet*. 1999 May-Jun;7(4):435-46. doi: 10.1038/sj.ejhg.5200317. PMID: 10352934.
- Makarla PB, Saboorian MH, Ashfaq R, Toyooka KO, Toyooka S, Minna JD, Gazdar AF, Schorge JO. Promoter hypermethylation profile of ovarian epithelial neoplasms. *Clin Cancer Res*. 2005 Aug 1;11(15):5365-9. doi: 10.1158/1078-0432.CCR-04-2455. PMID: 16061849.
- Mangla A, Hamad H, Kumar A. Factor XIII Deficiency. 2021 Oct 10. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan—. PMID: 32491399.
- Mao X, Bigham AW, Mei R, Gutierrez G, Weiss KM, Brutsaert TD, Leon-Velarde F, Moore LG, Vargas E, McKeigue PM, Shriver MD, Parra EJ. A genomewide admixture mapping panel for Hispanic/Latino populations. *Am J Hum Genet*. 2007 Jun;80(6):1171-8. doi: 10.1086/518564. Epub 2007 Apr 20. PMID: 17503334; PMCID: PMC1867104.
- Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitzel NO, Hillier L, Kwok PY, Gish WR. A general approach to single-nucleotide polymorphism discovery. *Nat Genet*. 1999 Dec;23(4):452-6. doi: 10.1038/70570. PMID: 10581034.
- Martínez-Cruz B, Mendizábal I, Harmant C, de Pablo R, Ioana M, Angelicheva D, Kouvatsi A, Makukh H, Netea MG, Pamjav H, Zalán A, Tournev I, Marushiakova E, Popov V, Bertranpetit J, Kalaydjieva L, Quintana-Murci L, Comas D. Origins, admixture and founder lineages in European Roma. *Eur J Hum Genet*. 2016 Jun;24(6):937-43. doi: 10.1038/ejhg.2015.201. Epub 2015 Sep 16. Erratum in: *Eur J Hum Genet*. 2021 Dec 24;; PMID: 26374132; PMCID: PMC4867443.
- Martínez-Cruz B, Ziegler J, Sanz P et al.; Multiplex single-nucleotide polymorphism typing of the human Y chromosome using TaqMan probes. *Invest Genet* 2011; 2: 13.

Martínez-Cruz B, Harmant C, Platt DE et al.: Evidence of pre-Roman tribal genetic structure in Basques from uniparentally inherited markers. *Mol BiolEvol* 2012; 29: 2211–2222.

Martínez Martínez, M. 2014. Los gitanos y gitanas de España a mediados del siglo XVIII. El fracaso de un proyecto de "exterminio" (1748-1765). Universidad de Almería. p. 30-33. ISBN 978-84-16027-31-6.

Marushiakova E, Popov V. 1997. Gypsies (Roma) in Bulgaria. Frankfurtam Main: Peter Lang.

Marushiakova E. and Popov V. .W Guy (Ed.), *Between past and future: the Roma of Central and Eastern Europe*, University of Hertfordshire Press, Hatfield, England (2001c), pp. 33-53

Marushiakova E. , Popov V. (2008) Ethnonyms and Professionyms. Gyps groups in the Balkans(Etnonimyiprofesjonimy. GrupyCyganownaBalkanach) .*StudiaRomologica* 177-198

Marushiakova, E, Popov V. "'Gypsy' groups in Eastern Europe: ethnonyms vs. professionyms." *Romani Studies*, vol. 23, no. 1, Apr. 2013, pp. 61+. Gale Academic

Masel J. Genetic drift. *Curr Biol.* 2011 Oct 25;21(20):R837-8. doi: 10.1016/j.cub.2011.08.007. PMID: 22032182.

Mastana SS, Papiha SS (1992) Origin of the Romany gypsies—genetic evidence. *Z MorpholAnthropol* 79: 43–51.

Matera AG, Hellmann U, Hintz MF, Schmid CW. Recently transposed Alu repeats result from multiple source genes. *Nucleic Acids Res.* 1990 Oct 25;18(20):6019-23. doi: 10.1093/nar/18.20.6019. PMID: 2172925; PMCID: PMC332399.

Ma X, Yang W, Gao Y, Pan Y, Lu Y, Chen H, Lu D, Xu S. Genetic Origins and Sex-Biased Admixture of the Huis. *Mol Biol Evol.* 2021 Aug 23;38(9):3804-3819. doi: 10.1093/molbev/msab158. PMID: 34021754; PMCID: PMC8382924.

Metspalu, M,Kivisild, T,Metspalu, E. et al,. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5, 26 (2004). <https://doi.org/10.1186/1471-2156-5-26>

Melegh, B. I,Banfai, Z,Hadzsiev, K,Miseta, A,&Melegh, B. (2017). Refining the South Asian Origin of the Romani people. *BMC Genetics*, 18(1).

Mendizábal I, Lao O, Marigorta UM, Wollstein A, Gusmão L, Ferak V, Ioana M, Jordanova A, Kaneva R, Kouvatsi A, Kučinskas V, Makukh H, Metspalu A, Netea MG, de Pablo R, Pamjav H, Radojkovic D, Rolleston SJ, Sertic J, Macek M Jr, Comas D, Kayser M. Reconstructing the population history of European Romani from genome-wide data. *Curr Biol.* 2012 Dec 18;22(24):2342-9. doi: 10.1016/j.cub.2012.10.039. Epub 2012 Dec 6. PMID: 23219723.

Mendizábal I, Valente C, Gusmão A, Alves C, Gomes V, Goios A, Parson W, Calafell F, Alvarez L, Amorim A, Gusmão L, Comas D, Prata MJ. Reconstructing the Indian origin and dispersal of the European Roma: a maternal genetic perspective. *PLoS One.* 2011 Jan 10;6(1):e15988. doi: 10.1371/journal.pone.0015988. PMID: 21264345; PMCID: PMC3018485.

Mighell AJ, Markham AF, Robinson PA. Alu sequences. *FEBS Lett.* 1997 Nov 3;417(1):1-5. doi: 10.1016/s0014-5793(97)01259-3. PMID: 9395063.

Millar JS, Maugeais C, Fuki IV, Rader DJ. Normal production rate of apolipoprotein B in LDL receptor-deficient mice. *ArteriosclerThrombVasc Biol.* 2002 Jun 1;22(6):989-94. doi: 10.1161/01.atv.0000018304.30943.06. PMID: 12067909.

Mills RE, Bennett EA, Iskow RC, Luttig CT, Tsui C, Pittard WS, Devine SE. Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet.* 2006 Apr;78(4):671-9. doi: 10.1086/501028. Epub 2006 Feb 2. PMID: 16532396; PMCID: PMC1424692.

Minorities at Risk Project. Assessment for Roma in North Macedonia 2006. Center for International Development and Conflict Management. University of Maryland. <http://www.mar.umd.edu/assessment.asp?groupid=34303> (27/05/2022)

Minorities at Risk Project. Assessment for Roma in Slovakia 2006. Center for International Development and Conflict Management. University of Maryland. <http://www.mar.umd.edu/assessment.asp?groupid=31702>(25/05/2022)

Minorities at Risk Project. Assessment for Roma in Spain 2006. Center for International Development and Conflict Management. University of Maryland. <http://www.mar.umd.edu/assessment.asp?groupid=23003> (10/10/2022)

MinorityRightsGroup International- Croatia /Roma
2008.<https://minorityrights.org/minorities/roma-4/> (25/05/2022)

Mishra P, Chan DC. Metabolic regulation of mitochondrial dynamics. *J Cell Biol.* 2016 Feb 15;212(4):379-87. doi: 10.1083/jcb.201511036. Epub 2016 Feb 8. PMID: 26858267; PMCID: PMC4754720.

Mishra VC, Gautam KA, Chaudhary R, Solanki H, Pokhariyal S, Saxena V, Dorwal P, Deshpande T, Singh P, Anthwal A, Dey N, Chandra D, Raina V. mtDNA Analysis: A Valuable Tool to Establish Relationships in Live Related Organ Transplants. *Indian J Nephrol.* 2020 Jan-Feb;30(1):14-20. doi: 10.4103/ijn.IJN_346_18. Epub 2019 Sep 6. PMID: 32015594; PMCID: PMC6977378.

Mitchell SL, Goodloe R, Brown-Gentry K, Pendergrass SA, Murdock DG, Crawford DC. Characterization of mitochondrial haplogroups in a large population-based sample from the United States. *Hum Genet.* 2014 Jul;133(7):861-8. doi: 10.1007/s00439-014-1421-9. Epub 2014 Feb 1. PMID: 24488180; PMCID: PMC4113317.

Morar B, Gresham D, Angelicheva D, Tournev I, Gooding R, Guerguelcheva V, Schmidt C, Abicht A, Lochmuller H, Tordai A, Kalmar L, Nagy M, Karcagi V, Jeanpierre M, Herczegfalvi A, Beeson D, Venkataraman V, Warwick Carter K, Reeve J, de Pablo R, Kucinskas V, Kalaydjieva L. Mutation history of the roma/gypsies. *Am J Hum Genet.* 2004 Oct;75(4):596-609. doi: 10.1086/424759. Epub 2004 Aug 20. PMID: 15322984; PMCID: PMC1182047.

Moreau R (2002) *The Rom: walking in the paths of the Gypsies.* FireflyBooks 301p.

Muscella A, Stefàno E, Marsigliante S. The effects of exercise training on lipid metabolism and coronary heart disease. *Am J Physiol Heart Circ Physiol*. 2020 Jul 1;319(1):H76-H88. doi: 10.1152/ajpheart.00708.2019. Epub 2020 May 22. PMID: 32442027.

Nachman MW. Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet*. 2001 Sep;17(9):481-5. doi: 10.1016/s0168-9525(01)02409-x. PMID: 11525814.

Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, Stoneking M. Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur J Hum Genet*. 2001 Apr;9(4):267-72. doi: 10.1038/sj.ejhg.5200615. PMID: 11313770.

National Statistical Institute.Republic of Bulgaria-census 2011.
[https://www.nsi.bg/en/content/3077/census-2011\(23/05/2022\)](https://www.nsi.bg/en/content/3077/census-2011(23/05/2022))

National Statistical Institute.Republic of Romania-census 2011.
[https://www.recensamantromania.ro/ \(23/05/2022\)](https://www.recensamantromania.ro/ (23/05/2022))

Navarrete HP, Soler LH, Mares RE, Ramos MA. Frequency of Alu insertions within the ACE and PR loci in Northwestern Mexicans. *BMC Res Notes*. 2017 Jul 27;10(1):339. doi: 10.1186/s13104-017-2673-y. PMID: 28750672; PMCID: PMC5530943.

Navarro-López B, Granizo-Rodríguez E, Palencia-Madrid L, Raffone C, Baeta M, de Pancorbo MM. Phylogeographic review of Y chromosome haplogroups in Europe. *Int J Legal Med*. 2021 Sep;135(5):1675-1684. doi: 10.1007/s00414-021-02644-6. Epub 2021 Jul 3. PMID: 34216266.

Nei M, Tajima F. Genetic drift and estimation of effective population size. *Genetics*. 1981 Jul;98(3):625-40. doi: 10.1093/genetics/98.3.625. PMID: 17249104; PMCID: PMC1214463.

Nicholson, G, Smith, A.V,Jónsson, F,Gústafsson, Ó,Stefánsson, K. and Donnelly, P. (2002), Assessing population differentiation and isolation from single-nucleotide polymorphism data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64: 695-715. <https://doi.org/10.1111/1467-9868.00357>

Noris G, Santana C, Meraz-Ríos MA, de Lourdes Munoz M, Majluf-Cruz A, Magaña JJ, Granados J, Quezada R, Revilla MC, Martínez-Salas S, Xihuitl S, Martínez de la Escalera G, Díaz-Badillo A, Calderon-Aranda ES, Gómez R. Mexican mestizo population sub-structure: effects on genetic

- and forensic statistical parameters. *Mol Biol Rep.* 2012 Dec;39(12):10139-56. doi: 10.1007/s11033-012-1888-1. Epub 2012 Sep 15. PMID: 22983890.
- Noticias Parlamento Europeo, (2020). ¿Qué hace la UE para atajar la discriminación gitana?. <https://www.europarl.europa.eu/news/es/headlines/society/20200918STO87401/que-hace-la-ue-para-atajar-la-discriminacion-gitana> (consultado el 16 de noviembre de 2022)
- Novick GE, Batzer MA, Deininger PL, Herrera RJ. The mobile genetic element Alu in the human genome. *Bioscience* 1996, 46: 32-41.
- Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ. Polymorphic Alu insertions and the Asian origin of Native American populations. *Hum Biol.* 1998 Feb;70(1):23-39. PMID: 9489232.
- Novokmet N, Pavcec Z. Genetic polymorphisms of 15 AmpFISTRidentifiler loci in Romani population from Northwestern Croatia. *Forensic Sci Int.* 2007 May 24;168(2-3):e43-6. doi: 10.1016/j.forsciint.2007.01.028. Epub 2007 Mar 8. PMID: 17349759.
- Ochoa A, Storey JD. Estimating FST and kinship for arbitrary population structures. *PLoS Genet.* 2021 Jan 19;17(1):e1009241. doi: 10.1371/journal.pgen.1009241. PMID: 33465078; PMCID: PMC7846127.
- Ongaro L, Molinaro L, Flores R, Marnetto D, Capodiferro MR, Alarcón-Riquelme ME, Moreno-Estrada A, Mabunda N, Ventura M, Tambets K, Achilli A, Capelli C, Metspalu M, Pagani L, Montinaro F. Evaluating the Impact of Sex-Biased Genetic Admixture in the Americas through the Analysis of Haplotype Data. *Genes (Basel).* 2021 Oct 7;12(10):1580. doi: 10.3390/genes12101580. PMID: 34680976; PMCID: PMC8535939.
- Otoni C, Ricaut FX, Vanderheyden N, Brucato N, Waelkens M, Decorte R. Mitochondrial analysis of a Byzantine population reveals the differential impact of multiple historical events in South Anatolia. *Eur J Hum Genet.* 2011 May;19(5):571-6. doi: 10.1038/ejhg.2010.230. Epub 2011 Jan 12. PMID: 21224890; PMCID: PMC3083616.
- Pakendorf B, Stoneking M. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet.* 2005;6:165-83. doi: 10.1146/annurev.genom.6.080604.162249. PMID: 16124858.
- Palencia-Madrid L, Baeta M, Kleinbielen T, Toro-Delgado N, Villaescusa P, Sanchez-Bustamante E, de Pancorbo MM, Luis JR, Ware KE, Somarelli JA, Garcia-Bertrand R, Herrera RJ. Post

- Austronesian migrational wave of West Polynesians to Micronesia. *Gene*. 2022 May 20;823:146357. doi: 10.1016/j.gene.2022.146357. Epub 2022 Feb 18. PMID: 35189246.
- Pan Z, Xu S. Population genomics of East Asian ethnic groups. *Hereditas*. 2020 Dec 8;157(1):49. doi: 10.1186/s41065-020-00162-w. PMID: 33292737; PMCID: PMC7724877.
- Park S, Cho S, Seo HJ, Lee JH, Kim MY, Lee SD. Entire Mitochondrial DNA Sequencing on Massively Parallel Sequencing for the Korean Population. *J Korean Med Sci*. 2017 Apr;32(4):587-592. doi: 10.3346/jkms.2017.32.4.587. PMID: 28244283; PMCID: PMC5334155.
- Peter BM, Slatkin M. The effective founder effect in a spatially expanding population. *Evolution*. 2015 Mar;69(3):721-34. doi: 10.1111/evo.12609. Epub 2015 Mar 10. PMID: 25656983; PMCID: PMC5104172.
- Peña JA, Alfonso-Sánchez MA, Pérez-Miranda AM, Garcia-Obregón S, Gómez-Pérez L GeDis: Un programa para análisis de datos en Antropogenética. 2009 *Antropo*,20: 40-59
- Peña, J. A., Gómez-Pérez, L., y Alfonso-Sánchez, M. A. (2022). On the Trail of Spatial Patterns of Genetic Variation. *Evolutionary Biology*, 49(1), 84-91. doi: 10.1007/s11692-021-09552-y.
- Peña JA, Morales B, Calderón R. New method for comparing levels of microdifferentiation: application to migration matrices of two populations from the Basque Country (Spain). *Hum Biol*. 1997 Jun;69(3):329-44. PMID: 9164044.
- Pereira V, Gusmão L, Valente C, Pereira R, Carneiro J, Gomes I, Morling N, Amorim A, João Prata M. Refining the genetic portrait of Portuguese Roma through X-chromosomal markers. *Am J Phys Anthropol*. 2012 Jul;148(3):389-94. doi: 10.1002/ajpa.22061. Epub 2012 May 11. PMID: 22576185.
- Petrejčíková E, Carnogurská J, Hronská D, Bernasovská J, Boronová I, Gabriková D, Bôziková A, Maceková S. Y-SNP analysis versus Y-haplogroup predictor in the Slovak population. *AnthropolAnz*. 2014;71(3):275-85. doi: 10.1127/0003-5548/2014/0368. PMID: 25065120.
- Petulengro J. *Gypsy Lore Soc*, 9 (1915-16), pp. 1-109
- Picard M, Sandi C. The social nature of mitochondria: Implications for human health. *NeurosciBiobehav Rev*. 2021 Jan;120:595-610. doi: 10.1016/j.neubiorev.2020.04.017. Epub 2020 Jul 8. PMID: 32651001; PMCID: PMC8058501.
- Piras IS, De Montis A, Calò CM, Marini M, Atzori M, Corrias L, Sazzini M, Boattini A, Vona G, Contu L. Genome-wide scan with nearly 700,000 SNPs in two Sardinian sub-populations suggests some regions as candidate targets for positive selection. *Eur J Hum Genet*. 2012 Nov;20(11):1155-61. doi: 10.1038/ejhg.2012.65. Epub 2012 Apr 25. PMID: 22535185; PMCID: PMC3476709.
- Price AL, Butler J, Patterson N, Capelli C, Pascali VL, Scarnicci F, Ruiz-Linares A, Groop L, Saetta AA, Korkolopoulou P, Seligsohn U, Waliszewska A, Schirmer C, Ardlie K, Ramos A, Nemes J, Arbeitman L, Goldstein DB, Reich D, Hirschhorn JN. Discerning the ancestry of European

- Americans in genetic association studies. *PLoS Genet.* 2008 Jan;4(1):e236. doi: 10.1371/journal.pgen.0030236. Epub 2007 Nov 19. PMID: 18208327; PMCID: PMC2211542.
- Quentin Y. Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Res.* 1992 Feb 11;20(3):487-93. doi: 10.1093/nar/20.3.487. PMID: 1741283; PMCID: PMC310412.
- Raff JA, Bolnick DA, Tackney J, O'Rourke DH. Ancient DNA perspectives on American colonization and population history. *Am J Phys Anthropol.* 2011 Dec;146(4):503-14. doi: 10.1002/ajpa.21594. Epub 2011 Sep 13. PMID: 21913177.
- Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK. ALFRED: an allele frequency resource for research and teaching. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D1010-5. doi: 10.1093/nar/gkr924. Epub 2011 Oct 28. PMID: 22039151; PMCID: PMC3245092.
- Relethford JH. *Human Population Genetics: 5 (Foundation of Human Biology)*. 2012. Hoboken, New Jersey, USA. Editorial: Wiley-Blackwell
- Raslan A, Dalia AA. Tissue Plasminogen Activator Thrombolysis of Entrapped Intra-Aortic Balloon Pump. *J CardiothoracVascAnesth.* 2020 May;34(5):1285-1287. doi: 10.1053/j.jvca.2019.12.002. Epub 2019 Dec 6. PMID: 31901465.
- Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ. Iran: tricontinental nexus for Y-chromosome driven migration. *Hum Hered.* 2006;61(3):132-43. doi: 10.1159/000093774. Epub 2006 Jun 12. PMID: 16770078.
- Ren WH, Li XH, Zhang HG, Deng FM, Liao WQ, Pang Y, Liu YH, Qiu MJ, Zhang GY, Zhang YG. Mitochondrial DNA haplogroups in a Chinese Uygur population and their potential association with longevity. *Clin Exp Pharmacol Physiol.* 2008 Dec;35(12):1477-81. doi: 10.1111/j.1440-1681.2008.05028.x. Epub 2008 Aug 26. PMID: 18759861.
- Resano, M., Moral, P., 2018. Mestizaje genético en las poblaciones humanas actuales de Argentina. *Revisión. Antropo*, 39, 77-96. www.didac.ehu.es/antropo
- Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, Marszalek JD, Pyntikova T, van der Veen F, Skaletsky H, Page DC, Rozen S. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet.* 2006 Apr;38(4):463-7. doi: 10.1038/ng1754. Epub 2006 Feb 26. PMID: 16501575.
- Reyniers A. *International migration and labour market policies: occasional papers No 1*, Organisation for Economic Co-operation and Development, Paris (1995), p. 8
- Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics.* 1983 Nov;105(3):767-79. doi: 10.1093/genetics/105.3.767. PMID: 17246175; PMCID: PMC1202185.
- Reznick DN, Ricklefs RE. Darwin's bridge between microevolution and macroevolution. *Nature.* 2009 Feb 12;457(7231):837-42. doi: 10.1038/nature07894. PMID: 19212402.

Ridker PM, Baker MT, Hennekens CH, Stampfer MJ, Vaughan DE. Alu-repeat polymorphism in the gene coding for tissue-type plasminogen activator (t-PA) and risks of myocardial infarction among middle-aged men. *ArteriosclerThrombVasc Biol.* 1997 Sep;17(9):1687-90. PMID: 9327764.

Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, et al., (1990) An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest* 86: 1343–1346

Rochow I. and K MatschkeK.Jahrbuch der ÖsterreichischenByzantinistik, 41 (1998), pp. 241-254

Rodvalho RG, Santos GS, Cavalcanti LM, Moura BF, Rodrigues EL, Lima PR, Gigonzac MA, Vieira TC. Allele frequencies of combined DNA index system (CODIS) and non-CODIS short tandem repeat loci in Goiás, Central Brazil. *Genet Mol Res.* 2015 Jul 3;14(3):7310-4. doi: 10.4238/2015.July.3.6. PMID: 26214409.

Rodríguez Maeso S, Araújo M. 'Civilising' the Roma/Gypsies. Public policies, 'employability' and the depoliticisation of (anti- 'employability' and the depoliticisation of (anti-) racism in Portugal. 2011.Centre for Social Studies, University of Coimbra (CES/UC)

Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M. Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet.* 2005 Mar;116(4):279-91. doi: 10.1007/s00439-004-1201-z. Epub 2005 Jan 20. PMID: 15660227.

Rossmann MP, Dubois SM, Agarwal S, Zon LI. Mitochondrial function in development and disease. *Dis Model Mech.* 2021 Jun 1;14(6):dmm048912. doi: 10.1242/dmm.048912. Epub 2021 Jun 11. PMID: 34114603; PMCID: PMC8214736.

Roy-Engel AM, Batzer MA, Deininger PL. Evolution of Human Retrosequences: Alu. In: *Encyclopedia of Life Sciences (ELS).* 2008 March 14th.

Roy S, Eaaswarkhanth M, Dubey B, Haque I. Autosomal STR variations in three endogamous populations of West Bengal, India. *Leg Med (Tokyo).* 2008 Nov;10(6):326-32. doi: 10.1016/j.legalmed.2008.06.001. Epub 2008 Jul 26. PMID: 18658003.

Rupert JL, Kidd KK, Norman LE, Monsalve MV, Hochachka PW, Devine DV. Genetic polymorphisms in the Renin-Angiotensin system in high-altitude and low-altitude Native American populations. *Ann Hum Genet.* 2003 Jan;67(Pt 1):17-25. doi: 10.1046/j.1469-1809.2003.00004.x. PMID: 12556231.

Saha T, Roy S, Chakraborty R, Biswas A, Das SK, Ray K, Ray J, Sengupta M. Mitochondrial DNA Haplogroups and Three Independent Polymorphisms have no Association with the Risk of Parkinson's Disease in East Indian Population. *Neurol India.* 2021 Mar-Apr;69(2):461-465. doi: 10.4103/0028-3886.314553. PMID: 33904476.

Sahakyan H, Margaryan A, Saag L, Karmin M, Flores R, Haber M, Kushniarevich A, Khachatryan Z, Bahmanimehr A, Parik J, Karafet T, Yunusbayev B, Reisberg T, Solnik A, Metspalu E,

Hovhannisyán A, Khusnutdinova EK, Behar DM, Metspalu M, Yepiskoposyan L, Rootsi S, VILLEMS R. Origin and diffusion of human Y chromosome haplogroup J1-M267. *Sci Rep*. 2021 Mar 23;11(1):6659. doi: 10.1038/s41598-021-85883-2. PMID: 33758277; PMCID: PMC7987999.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987 Jul;4(4):406-25. doi: 10.1093/oxfordjournals.molbev.a040454. PMID: 3447015.

Salem AH, Kilroy GE, Watkins WS, Jorde LB, Batzer MA. Recently integrated Alu elements and human genomic diversity. *Mol Biol Evol*. 2003 Aug;20(8):1349-61. doi: 10.1093/molbev/msg150. Epub 2003 May 30. PMID: 12777511.

Salem AH, Ray DA, Batzer MA. Identity by descent and DNA sequence variation of human SINE and LINE elements. *Cytogenet Genome Res*. 2005;108(1-3):63-72. doi: 10.1159/000080803. PMID: 15545717.

Salihović MP, Barešić A, Klarić IM, Cukrov S, Lauc LB, Jančićević B. The role of the Vlax Roma in shaping the European Romani maternal genetic history. *Am J Phys Anthropol*. 2011 Oct;146(2):262-70. doi: 10.1002/ajpa.21566. Epub 2011 Aug 16. PMID: 21915846.

Sato H, Okada Y, Seiki M. Membrane-type matrix metalloproteinases (MT-MMPs) in cell invasion. *Thromb Haemost*. 1997 Jul;78(1):497-500. PMID: 9198203.

Santiago MJ, López-Herce J, Zarzoso S, Sánchez A, Mencía S, Carrillo A. Trombólisis con dosis bajas de factor activador de plasminógeno en niños [Thrombolytic therapy using a low dose of tissue plasminogen activator in children]. *An Pediatr (Barc)*. 2012 Feb;76(2):77-82. Spanish. doi: 10.1016/j.anpedi.2011.07.031. Epub 2011 Oct 5. PMID: 21982548.

Santos HC, Horimoto AV, Tarazona-Santos E, Rodrigues-Soares F, Barreto ML, Horta BL, Lima-Costa MF, Gouveia MH, Machado M, Silva TM, Sanches JM, Esteban N, Magalhaes WC, Rodrigues MR, Kehdy FS, Pereira AC; Brazilian EPIGEN Project Consortium. A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: the Brazilian set. *Eur J Hum Genet*. 2016 May;24(5):725-31. doi: 10.1038/ejhg.2015.187. Epub 2015 Sep 23. PMID: 26395555; PMCID: PMC4930091.

Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg*. 2018 May;126(5):1763-1768. doi: 10.1213/ANE.0000000000002864. PMID: 29481436.

Schunkert H, Hense HW, Holmer SR, Stender M, Perz S, et al. (1994) Association between a deletion polymorphism of the angiotensin-converting-enzyme gene and left ventricular hypertrophy. *N Engl J Med* 330: 1634–1638.

Schwartz M, Vissing J. Paternal inheritance of mitochondrial DNA. *N Engl J Med*. 2002 Aug 22;347(8):576-80. doi: 10.1056/NEJMoa020350. PMID: 12192017.

Schwartz M, Vissing J. No evidence for paternal inheritance of mtDNA in patients with sporadic mtDNA mutations. *J Neurol Sci*. 2004 Mar 15;218(1-2):99-101. doi: 10.1016/j.jns.2003.11.008. PMID: 14759640.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science*. 2000 Nov 10;290(5494):1155-9. doi: 10.1126/science.290.5494.1155. PMID: 11073453.

Shen MR, Batzer MA, Deininger PL. Evolution of the master Alu gene(s). *J Mol Evol*. 1991 Oct;33(4):311-20. doi: 10.1007/BF02102862. PMID: 1774786.

Shen Z, Wang X, Yu X, Zhang Y, Qin L. MMP16 promotes tumor metastasis and indicates poor prognosis in hepatocellular carcinoma. *Oncotarget*. 2017 Aug 7;8(42):72197-72204. doi: 10.18632/oncotarget.20060. PMID: 29069779; PMCID: PMC5641122.

Sherier AJ, Woerner AE, Budowle B. Population Informative Markers Selected Using Wright's Fixation Index and Machine Learning Improves Human Identification Using the Skin Microbiome. *Appl Environ Microbiol*. 2021 Sep 28;87(20):e0120821. doi: 10.1128/AEM.01208-21. Epub 2021 Aug 11. PMID: 34379455; PMCID: PMC8478459.

Silva NM, Pereira L, Poloni ES, Currat M. Human neutral genetic variation and forensic STR data. *PLoS One*. 2012;7(11):e49666. doi: 10.1371/journal.pone.0049666. Epub 2012 Nov 21. PMID: 23185401; PMCID: PMC3504113.

Simão F, Ferreira AP, de Carvalho EF, Parson W, Gusmão L. Defining mtDNA origins and population stratification in Rio de Janeiro. *Forensic Sci Int Genet*. 2018 May;34:97-104. doi: 10.1016/j.fsigen.2018.02.003. Epub 2018 Feb 6. PMID: 29433058.

Singh G, Talwar I, Sandhu HS, Matharoo K, Bhanwer AJS. Genetic dissection of five ethnic groups from Punjab, North-West India-A study based on Autosomal Markers. *Leg Med (Tokyo)*. 2017 May;26:25-32. doi: 10.1016/j.legalmed.2017.02.003. Epub 2017 Mar 1. PMID: 28549544.

Singh G, Talwar I, Sharma R, Matharoo K, Bhanwer AJ. Genetic differentiation and population structure of five ethnic groups of Punjab (North-West India). *Mol Genet Genomics*. 2016 Dec;291(6):2055-2063. doi: 10.1007/s00438-016-1239-3. Epub 2016 Aug 9. PMID: 27507701.

Single nucleotide polymorphism (SNPS). National Human Genome Research Institute,NIH.Recuperado de: <https://www.genome.gov/genetics-glossary/Single-Nucleotide-Polymorphisms> (12/05/22)

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T,

Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. 2003 Jun 19;423(6942):825-37. doi: 10.1038/nature01722. PMID: 12815422.

Slatkin M. Gene flow and the geographic structure of natural populations. *Science*. 1987 May 15;236(4803):787-92. doi: 10.1126/science.3576198. PMID: 3576198.

Soták M, Petrejčíková E, Bernasovská J, Bernasovský I, Sovicová A, Boronová I, Svicková P, Bôžiková A, Gabriková D. Genetic variation analysis of 15 autosomal STR loci in Eastern Slovak Caucasian and Romany (Gypsy) population. *Forensic Sci Int Genet*. 2008 Dec;3(1):e21-5. doi: 10.1016/j.fsigen.2008.04.002. Epub 2008 May 20. PMID: 19083857.

Soták M, Petrejčíková E, Siváková D, Řeřaba K, Bôžiková A, Bernasovský I, Carnogurská J, Boronová I, Mačeková S, Homol'ová L, Sovičová A, Gabriková D, Rusínová L, Bernasovská J. Historical sketch of Slovak Haban (Hutterite) population based on autosomal STR analysis. *Hum Biol*. 2011 Oct;83(5):599-609. doi: 10.3378/027.083.0502. PMID: 22146063.

Sordé, T., Flecha, R., & Mircea, T. (2013). El pueblo gitano: una identidad global sin territorio (The Roma people: a global identity without territory) (Special Issue). *Scripta Nova*, 427(XVII).

Staadig A, Tillmar A. Evaluation of microhaplotypes in forensic kinship analysis from a Swedish population perspective. *Int J Legal Med*. 2021 Jul;135(4):1151-1160. doi: 10.1007/s00414-021-02509-y. Epub 2021 Jan 28. PMID: 33506298; PMCID: PMC8205927.

State Statistics Committee of Ukraine-census 2001. <http://www.ukrcensus.gov.ua/eng/> (23/05/2022)

State Statistical Office-Republic of North Macedonia-census 2021. https://www.stat.gov.mk/OblastOpsto_en.aspx?id=31 (24/05/2022)

Statistical Office of the Slovak Republic-census 2011.

https://slovak.statistics.sk/wps/portal/ext/themes/demography/census/indicators/!ut/p/z1/lZ HdjolwEIWfxRegU35Ke1lYKZiKhQLr9sawG2NIBL0wGt_exhiTze6Ce-4m-c6c UEGrZEZ2nO3a0_dYWj3tv4wZNNIRaMlc6DpEkOWSNWoeYXLCND7HYgFT_1QAIpAsh4Wpes8 DzgHjKv-OEPcXjN_wREWFkgz3Xd0MIrmPtP_0_Ajl83nl-SqfwFMt1n71y-egecgNkLhgSzAMAnOLDdzfcGv-x3BxKtXM58Eb-VqwSyKnaplsQFQSYAwA9g7IVTR8jTQ79FejugY19bra8669SOz2Y3KCS08A!/dz/d5/L2dBISEvZ0F BIS9nQSEh/ (23/05/2022)

Stefano GB, Kream RM. Mitochondrial DNA heteroplasmy in human health and disease. *Biomed Rep*. 2016 Mar;4(3):259-262. doi: 10.3892/br.2016.590. Epub 2016 Feb 4. PMID: 26998260; PMCID: PMC4774312

Stoneking M, Delfin F. The human genetic history of East Asia: weaving a complex tapestry. *Curr Biol*. 2010 Feb 23;20(4):R188-93. doi: 10.1016/j.cub.2009.11.052. PMID: 20178766.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* 1997 Nov;7(11):1061-71. doi: 10.1101/gr.7.11.1061. PMID: 9371742; PMCID: PMC310683.

Stringer C. The origin and evolution of *Homo sapiens*. *Philos Trans R Soc Lond B Biol Sci.* 2016 Jul 5;371(1698):20150237. doi: 10.1098/rstb.2015.0237. PMID: 27298468; PMCID: PMC4920294.

Tamburrini C, de Saint Pierre M, Bravi CM, Bailliet G, Jurado Medina L, Velázquez IF, Real LE, Holley A, Tedeschi CM, Basso NG, Parolin ML. Uniparental origins of the admixed Argentine Patagonia. *Am J Hum Biol.* 2021 Sep 17:e23682. doi: 10.1002/ajhb.23682. Epub ahead of print. PMID: 34533260.

Tang Y, Dai Y, Huo J. Decreased expression of T-cadherin is associated with gastric cancer prognosis. *Hepatogastroenterology.* 2012 Jun;59(116):1294-8. doi: 10.5754/hge12016. PMID: 22414547.

Tätte K, Metspalu E, Post H, Palencia-Madrid L, Luis JR, Reidla M, Rea A, Tamm E, Moding EJ, de Pancorbo MM, Garcia-Bertrand R, Metspalu M, Herrera RJ. The Ami and Yami aborigines of Taiwan and their genetic relationship to East Asian and Pacific populations. *Eur J Hum Genet.* 2021 Jul;29(7):1092-1102. doi: 10.1038/s41431-021-00837-6. Epub 2021 Mar 23. PMID: 33753914; PMCID: PMC8298601.

Thaler DS. Toward a microbial Neolithic revolution in buildings. *Microbiome.* 2016 Mar 29;4:14. doi: 10.1186/s40168-016-0157-2. PMID: 27021307; PMCID: PMC4810507.

Thalmann O, Fischer A, Lankester F, Pääbo S, Vigilant L. The complex evolutionary history of gorillas: insights from genomic data. *Mol Biol Evol.* 2007 Jan;24(1):146-58. doi: 10.1093/molbev/msl160. Epub 2006 Oct 25. PMID: 17065595.

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci U S A.* 2000 Jun 20;97(13):7360-5. doi: 10.1073/pnas.97.13.7360. PMID: 10861004; PMCID: PMC16550.

Tian C, Hinds DA, Shigeta R, Adler SG, Lee A, Pahl MV, Silva G, Belmont JW, Hanson RL, Knowler WC, Gregersen PK, Ballinger DG, Seldin MF. A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. *Am J Hum Genet.* 2007 Jun;80(6):1014-23. doi: 10.1086/513522. PMID: 17557415; PMCID: PMC1867091.

Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF. A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. *Am J Hum Genet.* 2006 Oct;79(4):640-9. doi: 10.1086/507954. Epub 2006 Aug 15. PMID: 16960800; PMCID: PMC1592561.

Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C, Klareskog L, Pulver AE, Qi L, Gregersen PK, Seldin MF. Analysis and application of European genetic substructure using 300

K SNP information. PLoS Genet. 2008 Jan;4(1):e4. doi: 10.1371/journal.pgen.0040004. PMID: 18208329; PMCID: PMC2211544.

Tigano A, Friesen VL. Genomics of local adaptation with gene flow. Mol Ecol. 2016 May;25(10):2144-64. doi: 10.1111/mec.13606. Epub 2016 Apr 6. PMID: 26946320.

Tishkoff SA, Pakstis AJ, Stoneking M, Kidd JR, Destro-Bisol G, Sanjantila A, Lu RB, Deinard AS, Sirugo G, Jenkins T, Kidd KK, Clark AG. Short tandem-repeat polymorphism/alu haplotype variation at the PLAT locus: implications for modern human origins. Am J Hum Genet. 2000 Oct;67(4):901-25. doi: 10.1086/303068. Epub 2000 Sep 13. PMID: 10986042; PMCID: PMC1287905.

Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 2000 Jul;10(7):967-81. doi: 10.1101/gr.10.7.967. PMID: 10899146; PMCID: PMC310925.

Tripathi M, Tripathi P, Chauhan UK, Herrera RJ, Agrawal S. Alu polymorphic insertions reveal genetic structure of north Indian populations. Hum Biol. 2008 Oct;80(5):483-99. doi: 10.3378/1534-6617-80.5.483. PMID: 19341319.

Tvedebrink T, Eriksen PS. Inference of admixed ancestry with Ancestry Informative Markers. Forensic Sci Int Genet. 2019 Sep;42:147-153. doi: 10.1016/j.fsigen.2019.06.013. Epub 2019 Jul 8. PMID: 31323537.

Ullu E, Tschudi C. *Alu* sequences are processed 7SL RNA genes. Nature. 1984;312:171-172. doi: 10.1038/312171a0.

Untoro E, Atmadja DS, Pu CE, Wu FC. Allele frequency of CODIS 13 in Indonesian population. Leg Med (Tokyo). 2009 Apr;11 Suppl 1:S203-5. doi: 10.1016/j.legalmed.2009.01.007. Epub 2009 Mar 3. PMID: 19261522.

Urabe F, Kosaka N, Sawa Y, Yamamoto Y, Ito K, Yamamoto T, Kimura T, Egawa S, Ochiya T. miR-26a regulates extracellular vesicle secretion from prostate cancer cells via targeting SHC4, PFDN4, and CHORDC1. Sci Adv. 2020 Apr 29;6(18):eaay3051. doi: 10.1126/sciadv.aay3051. PMID: 32494663; PMCID: PMC7190312.

Van der Bom JG, de Knijff P, Haverkate F, Bots ML, Meijer P, de Jong PT, Hofman A, Kluft C, Grobbee DE. Tissue plasminogen activator and risk of myocardial infarction. The Rotterdam Study. Circulation. 1997 Jun 17;95(12):2623-7. doi: 10.1161/01.cir.95.12.2623. PMID: 9193430.

Varela MA, Amos W. Evidence for nonindependent evolution of adjacent microsatellites in the human genome. J Mol Evol. 2009 Feb;68(2):160-70. doi: 10.1007/s00239-008-9192-3. Epub 2009 Jan 30. PMID: 19184164.

Varela MA, Amos W. Heterogeneous distribution of SNPs in the human genome: microsatellites as predictors of nucleotide diversity and divergence. Genomics. 2010 Mar;95(3):151-9. doi: 10.1016/j.ygeno.2009.12.003. Epub 2009 Dec 21. PMID: 20026267.

Varzari, A, Stephan, W, Stepanov, V, Raicu, F, Cojocaru, R, Roschin, Y, Glavce, C, Dergachev, V, Spiridonova, M, Schmidt, H.D, Weiss, E, 2007, Population history of the Dniester–Carpathians: evidence from Alu markers. *J. Hum. Genet*, 52, 308-316

Vega-Requena YV, Hidalgo PC, Ackermann E, Flores-Gutierrez S, Sans M. Genetic Admixture Analysis in the Population of Tacuarembó-Uruguay Using Alu Insertions. *Hum Biol*. 2020 Aug 6;91(4):249-256. doi: 10.13110/humanbiology.91.4.03. PMID: 32767895.

Vilborg A, Bersani C, Wilhelm MT, Wiman KG. The p53 target Wig-1: a regulator of mRNA stability and stem cell fate? *Cell Death Differ*. 2011 Sep;18(9):1434-40. doi: 10.1038/cdd.2011.20. Epub 2011 Mar 11. PMID: 21394102; PMCID: PMC3178426.

Viorel .On the enslavement of Roma in Romania, Achim, *The Roma in Romanian History*, Budapest: Central European University Press, 2004, pp.27-132.

Vissing J. Paternal comeback in mitochondrial DNA inheritance. *Proc Natl Acad Sci U S A*. 2019 Jan 29;116(5):1475-1476. doi: 10.1073/pnas.1821192116. Epub 2019 Jan 11. PMID: 30635426; PMCID: PMC6358698.

Wallace DC. Bioenergetics in human evolution and disease: implications for the origins of biological complexity and the missing genetic variation of common diseases. *Philos Trans R Soc Lond B Biol Sci*. 2013 Jun 10;368(1622):20120267. doi: 10.1098/rstb.2012.0267. PMID: 23754818; PMCID: PMC3685467.

Wallace DC, Chalkia D. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb Perspect Biol*. 2013 Nov 1;5(11):a021220. doi: 10.1101/cshperspect.a021220. PMID: 24186072; PMCID: PMC3809581.

Wang CM, Lin CJ, Feng HY. How to teach genetic drift. *Yi Chuan*. 2020 Dec 17;42(12):1211-1220. doi: 10.16288/j.ycz.20-310. PMID: 33509785.

Wang CZ, Li S, Hao X, Wang CC, Bai RF, Li H, Ma SH. Forensic characteristics and phylogenetic analysis of the Chinese Han population from Chongqing Municipality, Southwest China. *Leg Med (Tokyo)*. 2021 Nov;53:101954. doi: 10.1016/j.legalmed.2021.101954. Epub 2021 Aug 8. PMID: 34399366.

Wang F, Song F, Song M, Luo H, Hou Y. Genetic structure and paternal admixture of the modern Chinese Zhuang population based on 37 Y-STRs and 233 Y-SNPs. *Forensic Sci Int Genet*. 2022 Feb 23;58:102681. doi: 10.1016/j.fsigen.2022.102681. Epub ahead of print. PMID: 35263703.

Wang J, Pang GS, Chong SS, Lee CG. SNP web resources and their potential applications in personalized medicine. *Curr Drug Metab*. 2012 Sep 1;13(7):978-90. doi: 10.2174/138920012802138552. PMID: 22591348.

Wang LJ, Zhang CW, Su SC, Chen HH, Chiu YC, Lai Z, Bouamar H, Ramirez AG, Cigarroa FG, Sun LZ, Chen Y. An ancestry informative marker panel design for individual ancestry estimation of Hispanic population using whole exome sequencing data. *BMC Genomics*. 2019 Dec

30;20(Suppl 12):1007. doi: 10.1186/s12864-019-6333-6. PMID: 31888480; PMCID: PMC6936141.

Wang S, Smith JD. ABCA1 and nascent HDL biogenesis. *Biofactors*. 2014 Nov-Dec;40(6):547-54. doi: 10.1002/biof.1187. Epub 2014 Oct 30. PMID: 25359426; PMCID: PMC4294467.

Wang Z, He G, Luo T, Zhao X, Liu J, Wang M, Zhou D, Chen X, Li C, Hou Y. Massively parallel sequencing of 165 ancestry informative SNPs in two Chinese Tibetan-Burmese minority ethnicities. *Forensic Sci Int Genet*. 2018 May;34:141-147. doi: 10.1016/j.fsigen.2018.02.009. Epub 2018 Feb 13. PMID: 29477877.

Weaver TD. Did a discrete event 200,000-100,000 years ago produce modern humans? *J Hum Evol*. 2012 Jul;63(1):121-6. doi: 10.1016/j.jhevol.2012.04.003. Epub 2012 Jun 2. PMID: 22658331.

Wei YL, Qin CJ, Liu HB, Jia J, Hu L, Li CX. Validation of 58 autosomal individual identification SNPs in three Chinese populations. *Croat Med J*. 2014 Feb;55(1):10-3. doi: 10.3325/cmj.2014.55.10. PMID: 24577821; PMCID: PMC3944413.

Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG. Measures of human population structure show heterogeneity among genomic regions. *Genome Res*. 2005 Nov;15(11):1468-76. doi: 10.1101/gr.4398405. PMID: 16251456; PMCID: PMC1310634.

Weyrauch WO. *Gypsy law: Romani legal traditions and culture*: University of California Press. 2001. 284

Wright S. Evolution in Mendelian Populations. *Genetics*. 1931 Mar;16(2):97-159. doi: 10.1093/genetics/16.2.97. PMID: 17246615; PMCID: PMC1201091.

WRIGHT S. The genetical structure of populations. *Ann Eugen*. 1951 Mar;15(4):323-54. doi: 10.1111/j.1469-1809.1949.tb02451.x. PMID: 24540312.

Wu W, Chen D, Fu Y, Hao H, Nan H, Lu D. Population structure of Han population in China revealed by 41 STR loci. *Ann Hum Biol*. 2020 Feb;47(1):65-69. doi: 10.1080/03014460.2019.1705391. Epub 2020 Jan 5. PMID: 31905040.

Wutke S, Ludwig A. Targeted PCR Amplification and Multiplex Sequencing of Ancient DNA for SNP Analysis. *Methods Mol Biol*. 2019;1963:141-147. doi: 10.1007/978-1-4939-9176-1_15. PMID: 30875052.

Xiao, F.X, Yang, J.F, Cassiman, J.J, Decorte, R, 2002, Diversity at eight polymorphic Alu insertion loci in Chinese populations shows evidence for European admixture in an ethnic minority population from northwest China, *Hum. Biol*, 74,555-68

Yahya P, Sulong S, Harun A, Wangkumhang P, Wilantho A, Ngamphiw C, Tongsimma S, Zilfalil BA. Ancestry-informative marker (AIM) SNP panel for the Malay population. *Int J Legal Med*. 2020 Jan;134(1):123-134. doi: 10.1007/s00414-019-02184-0. Epub 2019 Nov 23. PMID: 31760471.

Yahya P, Sulong S, Harun A, Wan Isa H, Ab Rajab NS, Wangkumhang P, Wilantho A, Ngamphiw C, Tongsimma S, Zilfalil BA. Analysis of the genetic structure of the Malay population: Ancestry-informative marker SNPs in the Malay of Peninsular Malaysia. *Forensic Sci Int Genet*. 2017 Sep;30:152-159. doi: 10.1016/j.fsigen.2017.07.005. Epub 2017 Jul 14. PMID: 28743033.

Yan C, Duanmu X, Zeng L, Liu B, Song Z. Mitochondrial DNA: Distribution, Mutations, and Elimination. *Cells*. 2019 Apr 25;8(4):379. doi: 10.3390/cells8040379. PMID: 31027297; PMCID: PMC6523345.

Yao J, Wang BJ. Genetic Variation of 25 Y-Chromosomal and 15 Autosomal STR Loci in the Han Chinese Population of Liaoning Province, Northeast China. *PLoS One*. 2016 Aug 2;11(8):e0160415. doi: 10.1371/journal.pone.0160415. PMID: 27483472; PMCID: PMC4970702.

Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res*. 2002 Feb;12(2):339-48. doi: 10.1101/gr.217602. PMID: 11827954; PMCID: PMC155271.

Yokoyama S. Assembly of high-density lipoprotein. *ArteriosclerThrombVasc Biol*. 2006 Jan;26(1):20-7. doi: 10.1161/01.ATV.0000195789.39418.e8. Epub 2005 Nov 10. PMID: 16284193.

Zeder MA. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci U S A*. 2008 Aug 19;105(33):11597-604. doi: 10.1073/pnas.0801317105. Epub 2008 Aug 12. PMID: 18697943; PMCID: PMC2575338.

Zeng X, Chakraborty R, King JL, LaRue B, Moura-Neto RS, Budowle B. Selection of highly informative SNP markers for population affiliation of major US populations. *Int J Legal Med*. 2016 Mar;130(2):341-52. doi: 10.1007/s00414-015-1297-9. Epub 2015 Dec 8. PMID: 26645290.

Zhou Z, Shao C, Xie J, Xu H, Liu Y, Zhou Y, Liu Z, Zhao Z, Tang Q, Sun K. Genetic polymorphism and phylogenetic analyses of 21 non-CODIS STR loci in a Chinese Han population from

Shanghai. *Mol Genet Genomic Med*. 2020 Feb;8(2):e1083. doi: 10.1002/mgg3.1083. Epub 2019 Dec 8. PMID: 31814334; PMCID: PMC7005660.

Zhu X, Bouzekri N, Southam L et al,: Linkage and association analysis of angiotensin I-converting enzyme (ACE)-gene polymorphisms with ACE concentration and blood pressure. *Am J Hum Genet* 2001; 68: 1139–1148.

Zmasek CM, (2018) <https://sites.google.com/site/cmzmasek/christian-zmasek/software/forester>



Anexo

IX. Anexo

Tabla 1: Resumen de poblaciones recopiladas para las diferentes familias de marcadores analizadas, organizadas por continentes junto a su respectiva bibliografía.

AIMs		
Poblaciones	Continente	Bibliografía
Vasca		
Española		
Francesa		
Británica		
Irlandesa		
Italiana		
Sarda	Europa	Cheung et al., 1999
Húngara		
Danesa		
Finlandesa		
Estonia		
Bielorrusa		
Ucraniana		
Burusho		
Kalash		
Brahui		
Sindhi		
Balochi		
Hazara	Asia	Cheung et al., 1999
Telugu		
Punjabí		
Gujarati		
Bengalí		
Tamil		

Al

Poblaciones	Continente	Bibliografía		
Vasca-Navarra (BSQN)				
Vasca-Guipuzcoana (BSQG)				
Catalana (CTLN)				
Andaluza (NDLC)				
Valenciana (VLNC)				
Canaria (CNRS)				
Francesa (BRTN)				
Suiza (SWTZ)	Europa	Comas et al., 2000, 2004; García-Obregón et al., 2006; Nasidze et al., 2001; Stoneking et al., 1997; Varzari et al., 2007		
Ucraniana (UKRN)				
Moldava (MLDV)				
Rumana (ROMN)				
Macedonia (MCDN)				
Albanesa (ALBN)				
Griega (GREC)				
Turca (TRKS)				
<hr/>				
Emiratos Árabes (EMRT)				
Sirios (SYR)				
Pakistaníes (PKST)				
Tamil (TAML)				
Indo-cristianos (IND1)				
Indo-hindús (IND2)				
Indo-musulmanes (IND3)				
Pashtunes (PSHT)	Asia	<hr/>		
Kayasthas				
Mathurs				
Rastogies				
Vaish				
Brahmins			Tripathi et al., 2008	
Bhargavas				
Chaturvedis				
Sunitas				
Chiitas				
Han		Xiao et al., 2002		
<hr/>				
!Kung	África	Stoneking et al., 1997		
<hr/>				

CODIS			
Poblaciones	Continente	Bibliografía	
Romaní-macedonia	Romaníes Europa	Cabezas Silva et al., 2016; Egyed et al., 2000, 2006; Füredi et al., 1996, 1997; Gusmão et al., 2010; Haliti et al., 2009; Havaš et al., 2007; Soták et al., 2008; ; Novokmet et al., 2007;	
Romaní-eslovaca			
Romaní-croata			
Romaní-portuguesa			
Romaní-húngara			
Macedonios	Europa	Cabezas Silva et al., 2016; Egyed et al., 2000, 2006; Füredi et al., 1996, 1997; Gusmão et al., 2010; Haliti et al., 2009; Havaš et al., 2007; Soták et al., 2008; ; Novokmet et al., 2007;	
Eslovacos			
Croatas			
Portugueses			
Húngaros			
Franceses			
Vasco-franceses			
Italianos			Amigo et al., 2009; Budowle et al., 2001; Kuzniar et al., 2004
Sardos			
Españoles			
Polacos			
Teli	Asia	Badiye et al., 2021; Hedjazi et al., 2013; Kumar et al., 2020; Sahajpal et al., 2019; Shan et al., 2021	
Rajput			
Rajashtán			
Fars			
Punjabí			
Saraiki			
Pashtunes			
Sindhi			
Balochis			

CromY		
Poblaciones	Continente	Bibliografía
Romaní-búlgara	Romaníes Europa	Martínez-Cruz et al., 2016
Romaní-húngara		
Romaní-eslovaca		
Romaní-española		
Romaní-griega		
Romaní-rumana		
Romaní-ucraniana		

Búlgara		
Húngara		
Eslovaca		
Española		Martínez-Cruz et al., 2016
Griega		
Rumana		
Ucraniana	Europa	
Francesa		
Neerlandesa		
Alemana		Semino et al., 2000; Martínez-Cruz et al., 2016
Polaca		
Macedonia		
Albana		
Turca		
Libanesa		
Siria		
Jordana		
Iraquí		
Iraní		
Afgana		
Punjabí		
Dhodia		
Dubla		
Konkana	Asia	Flores et al., 2005; Haber et al., 2012; Khurana et al., 2014; Kivisild et al., 2003; Regueiro et al., 2006
Vasava		
Gamit		
Valvi Chaudhari		
Nana Chaudhari		
Mota Chaudhari		
Pavagadhi Chaudhari		
Gujarati		
Koya		
Chenchu		

ADNmt

Poblaciones	Continente	Bibliografía
Romaní-búlgara		
Romaní-húngara		
Romaní-eslovaca		
Romaní-española	Romaníes Europa	Martínez-Cruz et al., 2016
Romaní-griega		
Romaní-rumana		
Romaní-ucraniana		

Búlgara		
Húngara		
Eslovaca		
Española		Martínez-Cruz et al., 2016
Griega		
Rumana	Europa	
Ucraniana		
Francesa		
Alemana		
Italiana		Otoni et al., 2011
Austriaca		
Polaca		
Turca		
Siria		
Libanesa		
Jordana		
Iraquí		
Iraní		
Gilaki		
Mazandaria		
Persa		
Hazara		
Baluchi	Asia	Metspalu et al., 2004; Otoni et al., 2011
Brahui		
Burusho		
Pakistaní		
Pathan		
Sindhi		
Pashtunes		
Punjabí		
Rajput		
Gujarati		
Makrani		
Parsi		

Tabla 2: Frecuencias alélicas de los AIMs analizados para la población de estudio y poblaciones euroasiáticas.

SNP	Alelo principal	Cromosoma	Romaníes del				
			País Vasco	Vascos	Bielorosos	Británicos	Daneses
rs11102524	C	1	0,519	0,48	0,518	0,494	0,663
rs4908602	A	1	0,276	0,38	0,393	0,39	0,337
rs6429223	A	1	0,576	0,63	0,643	0,736	0,721
rs1019264	C	2	0,405	0,67			0,442
rs182549	C	2	0,844	0,33		0,261	0,245
rs2920059	A	2	0,451	0,4	0,518	0,555	0,523
rs896788	A	2	0,342	0,23		0,163	0,147
rs10470510	A	3	0,532	0,54	0,571	0,527	0,512
rs1498444	A	3	0,625	0,56	0,571	0,574	0,698
rs4455309	C	3	0,667	0,77			0,651
rs9870523	A	3	0,63	0,46			0,477
rs1157106	C	4	0,37	0,29	0,464	0,352	0,361
rs116818505	G	6	0			0,352	
rs200789833	G	6	0,463			0,214	
rs2071350	C	6	0,973			0,934	
rs2770	C	6	0,222			0,478	
rs369150	A	6	0,295			0,242	
rs3823324	T	6	0,671			0,484	
rs72500564	C	6	0,838			0,764	
rs72873921	C	6	0,104			0,214	
rs72873922	A	6	0,097			0,214	
rs79244404	C	6	0,149			0,17	
rs9273352	G	6	0,638			0,445	
rs9277332	A	6	0,629			0,313	
rs9277413	A	6	0,173			0,418	
rs9277418	C	6	0,818			0,588	
rs9277498	C	6	0,16			0,302	
rs9374640	A	6	0,331	0,54	0,554	0,528	0,512
rs7783055	C	7	0,669	0,52	0,643	0,5	0,547
rs917118	C	7	0,679	0,729	0,839	0,636	0,75
rs1392121	C	8	0,453	0,17	0,321	0,39	0,384
rs10123957	C	9	0,664	0,4	0,482	0,528	0,43
rs1493489	C	12	0,539	0,48			0,523
rs2710293	A	12	0,634	0,58	0,375	0,483	0,488
rs1886510	C	13	0,697	0,462		0,437	0,42
rs2065982	C	13	0,083	0,02		0,017	0,069
rs730570	A	14	0,87	0,9		0,847	0,88
rs1834640	A	15	0,805	1		1	0,98
rs8032154	A	15	0,52	0,58			0,57
rs881929	G	16	0,48	0,46		0,648	0,412
rs2033719	C	17	0,325	0,721	0,429	0,506	0,523
rs1024116	A	18	0,493	0,571		0,602	0,54
rs16951664	G	18	0,589	0,29			0,605

Material anexo

rs2285513	A	19	0,383	0,58	0,321	0,484	
rs2303798	A	19	0,032	0		0	0
rs6057059	A	20	0,453	0,5	0,446	0,5	0,558
rs722098	A	21	0,831	0,843		0,709	0,804
rs2040411	A	22	0,712	0,771		0,653	0,64



SNP	Estonios	Finlandeses	Franceses- bretones	Húngaros	Península Ibérica	Irlandeses	Italianos
rs11102524	0,59	0,626	0,62	0,549	0,542	0,546	0,575
rs4908602	0,327	0,4	0,47	0,44	0,449	0,42	0,42
rs6429223	0,763	0,7	0,66	0,659	0,645	0,75	0,659
rs1019264		0,66	0,67	0,61		0,568	0,44
rs182549		0,425	0,55	0,64	0,429	0,149	0,87
rs2920059	0,531	0,636	0,6	0,476	0,439	0,591	0,411
rs896788		0,236	0,09	0,168	0,286	0,186	0,23
rs10470510	0,566	0,485	0,52	0,488	0,435	0,364	0,514
rs1498444	0,543	0,657	0,57	0,46	0,537	0,65	0,545
rs4455309		0,64	0,69	0,622		0,5	0,73
rs9870523		0,58	0,47	0,61		0,773	0,69
rs1157106	0,459	0,54	0,41	0,354	0,336	0,386	0,46
rs116818505		0,182			0,262		0,276
rs200789833		0,056			0,215		0,187
rs2071350		0,995			0,921		0,944
rs2770		0,384			0,449		0,435
rs369150		0,248			0,262		0,257
rs3823324		0,495			0,486		0,551
rs72500564		0,909			0,757		0,799
rs72873921		0,056			0,215		0,187
rs72873922		0,056			0,215		0,187
rs79244404		0,106			0,192		0,206
rs9273352		0,53			0,407		0,407
rs9277332		0,167			0,379		0,36
rs9277413		0,192			0,346		0,327
rs9277418		0,823			0,664		0,673
rs9277498		0,172			0,257		0,238
rs9374640	0,604	0,5	0,62	0,61	0,631	0,659	0,63
rs7783055	0,556	0,5	0,6	0,5	0,491	0,5	0,472
rs917118	0,773	0,747	0,83	0,758	0,734	0,741	0,635
rs1392121	0,341	0,308	0,41	0,232	0,332	0,386	0,411
rs10123957	0,502	0,551	0,52	0,476	0,421	0,386	0,505
rs1493489	0,463	0,54	0,57	0,524		0,682	0,42
rs2710293	0,462	0,444	0,57	0,512	0,533	0,546	0,551
rs1886510		0,468	0,48	0,517	0,5	0,438	0,516
rs2065982		0,07	0,02	0,062	0	0,084	0,071
rs730570		0,785	0,8	0,851	0,857	0,891	0,859
rs1834640	0,987	0,99	0,98	0,978	0,995	1	1

Procesos microevolutivos en la población romaní residente en el País Vasco

rs8032154		0,36	0,45	0,451		0,636	0,54
rs881929		0,667	0,48	0,37	0,607	0,67	0,533
rs2033719	0,455	0,515	0,52	0,524	0,533	0,5	0,491
rs1024116		0,516	0,66		0,429		0,525
rs16951664		0,4	0,57	0,488		0,546	0,5
rs2285513	0,46	0,404	0,43		0,5		0,491
rs2303798		0,027	0	0,006	0,036	0,009	0,04
rs6057059	0,482	0,475	0,38	0,317	0,458	0,477	0,435
rs722098		0,747	0,84	0,787	0,929	0,768	0,655
rs2040411		0,629	0,62	0,619	0,714	0,64	0,5

SNP	Rusos	Sardos	Ucranianos	Burusho	Kalash	Balochi	Bengalís
rs11102524	0,577	0,59	0,586	0,44	0,66	0,46	0,448
rs4908602	0,34	0,54	0,431	0,52	0,56	0,46	0,454
rs6429223	0,71	0,46	0,534	0,5	0,7	0,58	0,506
rs1019264	0,68	0,55		0,48	0,46	0,52	
rs182549	0,691	0,93		0,88	1	0,64	0,93
rs2920059	0,539	0,48	0,569	0,46	0,62	0,54	0,657
rs896788	0,18	0,23		0,5	0,36	0,36	0,547
rs10470510	0,5	0,45	0,621	0,56	0,46	0,5	0,581
rs1498444	0,628	0,46	0,512	0,72	0,46	0,56	0,692
rs4455309	0,62	0,68		0,46	0,72	0,56	
rs9870523	0,7	0,61		0,5	0,78	0,52	
rs1157106	0,5	0,29	0,379	0,32	0,2	0,42	0,454
rs116818505							0,32
rs200789833							0,366
rs2071350							0,855
rs2770							0,355
rs369150							0,238
rs3823324							0,704
rs72500564							0,611
rs72873921							0,366
rs72873922							0,366
rs79244404							0,209
rs9273352							0,581
rs9277332							0,581
rs9277413							0,43
rs9277418							0,57
rs9277498							0,413
rs9374640	0,6	0,68	0,638	0,48	0,54	0,42	0,442
rs7783055	0,59	0,61	0,465	0,4	0,62	0,6	0,686
rs917118	0,719	0,5	0,724	0,56	0,56	0,6	0,517
rs1392121	0,359	0,27	0,362	0,42	0,5	0,48	0,535
rs10123957	0,462	0,41	0,397	0,34	0,48	0,4	0,361
rs1493489	0,42	0,48		0,46	0,48	0,5	
rs2710293	0,346	0,45	0,448	0,66	0,54	0,58	0,674

Material anexo

rs1886510	0,511	0,543		0,7	0,58	0,58	0,837
rs2065982	0,031	0,04		0,32	0,04	0,16	0,25
rs730570	0,681	0,97		0,6	0,66	0,72	0,529
rs1834640	0,958	0,971		0,98	1	0,86	0,564
rs8032154	0,54	0,52		0,66	0,74	0,62	
rs881929	0,362	0,42		0,76	0,44	0,58	0,843
rs2033719	0,423	0,52	0,483	0,38	0,16	0,32	0,349
rs1024116	0,48	0,5		0,44	0,26	0,54	0,221
rs16951664	0,4	0,45		0,56	0,58	0,5	
rs2285513	0,462	0,52	0,379	0,36	0,34	0,4	0,366
rs2303798	0,031	0,015		0,08	0,16	0,06	0,25
rs6057059	0,487	0,52	0,5	0,56	0,54	0,56	0,628
rs722098	0,722	0,871		0,6	0,28	0,68	0,558
rs2040411	0,667	0,757		0,55	0,48	0,58	0,424

SNP	Brahui	Gujarati	Hazara	Punjabi	Sindhi	Tamil	Telugu
rs11102524	0,44	0,495	0,56	0,443	0,46	0,387	0,323
rs4908602	0,44	0,5	0,29	0,412	0,48	0,515	0,544
rs6429223	0,48	0,587	0,46	0,542	0,66	0,431	0,495
rs1019264							
rs182549	0,62	0,859	0,88	0,734	0,7	0,941	0,936
rs2920059	0,42	0,733	0,69	0,651	0,68	0,721	0,662
rs896788	0,32	0,456	0,46	0,443	0,54	0,5	0,534
rs10470510	0,38	0,558	0,67	0,604	0,46	0,642	0,539
rs1498444	0,7	0,636	0,57	0,661	0,62	0,642	0,691
rs4455309	0,5		0,56		0,72		
rs9870523	0,5		0,54		0,64		
rs1157106	0,48	0,413	0,46	0,412	0,5	0,417	0,431
rs116818505		0,214		0,219		0,279	0,26
rs200789833		0,286		0,394		0,368	0,343
rs2071350		0,918		0,885		0,828	0,833
rs2770		0,427		0,474		0,397	0,392
rs369150		0,16		0,182		0,128	0,137
rs3823324		0,646		0,552		0,672	0,578
rs72500564		0,796		0,787		0,696	0,745
rs72873921		0,286		0,349		0,368	0,343
rs72873922		0,286		0,349		0,368	0,343
rs79244404		0,185		0,219		0,226	0,177
rs9273352		0,49		0,552		0,52	0,559
rs9277332		0,495		0,568		0,598	0,632
rs9277413		0,306		0,281		0,418	0,353
rs9277418		0,699		0,729		0,588	0,652
rs9277498		0,257		0,276		0,378	0,328
rs9374640	0,42	0,335	0,46	0,328	0,34	0,314	0,348
rs7783055	0,5	0,607	0,42	0,557	0,66	0,564	0,598
rs917118	0,36	0,621	0,62	0,573	0,64	0,515	0,539

Procesos microevolutivos en la población romaní residente en el País Vasco

rs1392121	0,6	0,515	0,44	0,484	0,5	0,505	0,583
rs10123957	0,34	0,316	0,5	0,385	0,48	0,368	0,373
rs1493489	0,6		0,56		0,46		
rs2710293	0,6	0,626	0,54	0,646	0,56	0,686	0,701
rs1886510	0,66	0,801	0,83	0,823	0,76	0,809	0,824
rs2065982	0,18	0,204	0,27	0,193	0,27	0,225	0,25
rs730570	0,74	0,539	0,46	0,578	0,71	0,52	0,529
rs1834640	0,98	0,942	0,632	0,766	0,8	0,485	0,652
rs8032154	0,62		0,63		0,68		
rs881929	0,78	0,944	0,62	0,802	0,74	0,892	0,907
rs2033719	0,34	0,354	0,54	0,422	0,5	0,324	0,407
rs1024116	0,6	0,291	0,44	0,359	0,4	0,338	0,324
rs16951664	0,52		0,44		0,48		
rs2285513	0,46	0,413	0,63	0,385	0,42	0,422	0,427
rs2303798	0	0,16	0,14	0,099	0,114	0,201	0,221
rs6057059	0,52	0,621	0,65	0,609	0,5	0,647	0,672
rs722098	0,58	0,636	0,31	0,698	0,18	0,667	0,662
rs2040411	0,43	0,471	0,38	0,469	0,475	0,52	0,471

Tabla 3: Frecuencias de inserción de los elementos *Alu* de estudio en poblaciones europeas y asiáticas. Leyenda: BSR: Romaníes residentes en el País Vasco, población de estudio (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), BRTN: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucrucianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, IND1: India-Cristianos, IND2: India-Hindús, IND3: India-Musulmanes.

<i>Alu</i>	BSR	BSQG	BSQN	CTLN	NDLC	VLNC
TPA25	0,434	0,553	0,592	0,608	0,59	0,556
ACE	0,196	0,425	0,32	0,3	0,47	0,387
APO	0,903	0,95	0,97	0,983	0,985	0,94
PV92	0,265	0,239	0,154	0,175	0,194	0,232
FXIIIB	0,238	0,441	0,49	0,5	0,448	0,476
D1	0,313	0,473	0,373	0,35	0,306	0,322
B65	0,422	0,557	0,466	0,525	0,552	0,529
A25	0,155	0,156	0,124	0,125	0,142	0,104
<i>Alu</i>	CNRS	FRNC	SWTZ	MCDN	GREC	ALBN
TPA25	0,609	0,56	0,453	0,521	0,552	0,557
ACE	0,369	0,48	0,372	0,44	0,26	0,467
APO	0,949	0,99	0,942	0,976	0,977	1
PV92	0,25	0,23	0,198	0,167	0,19	0,203
FXIIIB	0,383	0,42	0,477	0,542	0,5	0,6
D1	0,348	0,46	0,337	0,38	0,409	0,267
B65	0,559	0,57	0,58	0,597	0,651	0,67
A25	0,155	0,16	0,12	0,109	0,098	0,075

Material anexo

<i>Alu</i>	MLDV	ROMN	UKRN	TRKS	EMRT	SYR
TPA25	0,569	0,577	0,518	0,427	0,44	0,51
ACE	0,486	0,469	0,404	0,441	0,33	0,4
APO	0,963	0,915	0,964	0,963	0,97	0,93
PV92	0,249	0,2	0,235	0,256	0,3	0,18
FXIIIB	0,488	0,408	0,441	0,473	0,39	0,28
D1	0,383	0,415	0,412	0,284	0,08	0,29
B65	0,571	0,569	0,53	0,706	0,41	0,31
A25	0,117	0,031	0,077	0,048	0,12	0

<i>Alu</i>	PKST	TAML	PSHT	IND1	IND2	IND3
TPA25	0,51	0,56	0,55	0,57	0,34	0,41
ACE	0,44	0,69	0,52	0,6	0,52	0,52
APO	0,72	0,81	0,86	0,67	0,85	0,86
PV92	0,3	0,56	0,33	0,48	0,52	0,3
FXIIIB	0,23	0,61	0,57	0,61	0,66	0,66
D1	0,17	0,34	0,27	0,28	0,1	0,32
B65	0,37	0,55	0,49	0,31	0,35	0,4
A25	0,07	0,17	0,18	0,14	0,05	0,12

Tabla 4: Frecuencias de inserción de 7 inserciones *Alu* de estudio en poblaciones europeas y asiáticas. Leyenda: BSR: Romaníes residentes en el País Vasco, población de estudio (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), BRTN: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucranianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, Uttar Pradesh: Kayasthas, Mathurs, Rastogies, Vaish, Brahmins, Bhargavas, Chaturvedis, Sunni: Sunitas y Shia: Chiitas.

<i>Alu</i>	BSR	BSQG	BSQN	CTLN	NDLC	VLNC
TPA25	0,434	0,553	0,592	0,608	0,590	0,556
ACE	0,196	0,425	0,320	0,300	0,470	0,387
APO	0,903	0,950	0,970	0,983	0,985	0,940
PV92	0,265	0,239	0,154	0,175	0,194	0,232
FXIIIB	0,238	0,441	0,490	0,500	0,448	0,476
D1	0,313	0,473	0,373	0,350	0,306	0,322
B65	0,422	0,557	0,466	0,525	0,552	0,529

<i>Alu</i>	CNRS	FRNC	SWTZ	MCDN	GREC	ALBN
TPA25	0,609	0,560	0,453	0,521	0,552	0,557
ACE	0,369	0,480	0,372	0,440	0,260	0,467
APO	0,949	0,990	0,942	0,976	0,977	1
PV92	0,250	0,230	0,198	0,167	0,190	0,203
FXIIIB	0,383	0,420	0,477	0,542	0,500	0,600
D1	0,348	0,460	0,337	0,380	0,409	0,267
B65	0,559	0,570	0,580	0,597	0,651	0,670

<i>Alu</i>	MLDV	ROMN	UKRN	TRKS	EMRT	SYR
TPA25	0,569	0,577	0,518	0,427	0,440	0,510
ACE	0,486	0,469	0,404	0,441	0,330	0,400
APO	0,963	0,915	0,964	0,963	0,970	0,930
PV92	0,249	0,200	0,235	0,256	0,300	0,180
FXIIIB	0,488	0,408	0,441	0,473	0,390	0,280
D1	0,383	0,415	0,412	0,284	0,080	0,290
B65	0,571	0,569	0,530	0,706	0,410	0,310

<i>Alu</i>	PKST	TAML	PSHT	Shia	Sunni	Kayastha
TPA25	0,510	0,560	0,550	0,468	0,540	0,523
ACE	0,440	0,690	0,520	0,515	0,558	0,595
APO	0,720	0,810	0,860	0,725	0,743	0,820
PV92	0,300	0,560	0,330	0,345	0,353	0,385
FXIIIB	0,230	0,610	0,570	0,433	0,405	0,433
D1	0,170	0,340	0,270	0,255	0,268	0,235
B65	0,370	0,550	0,490	0,333	0,333	0,323

<i>Alu</i>	Mathur	Rastogi	Vaish	Bhargava	Brahmin	Chaturvedi
TPA25	0,658	0,500	0,500	0,498	0,483	0,528
ACE	0,570	0,605	0,613	0,565	0,525	0,573
APO	0,910	0,888	0,900	0,885	0,910	0,880
PV92	0,388	0,335	0,340	0,293	0,283	0,300
FXIIIB	0,380	0,445	0,470	0,343	0,345	0,378
D1	0,228	0,230	0,225	0,305	0,325	0,290
B65	0,293	0,410	0,415	0,453	0,483	0,448

Tabla 5: Frecuencias de inserción de 7 inserciones *Alu* de estudio en poblaciones europeas y asiáticas. Legenda: BSR: Romaníes residentes en el País Vasco, población de estudio (España), BSQG: Vascos Guipúzcoa (España), BSQN: Vascos Navarra (España), CTLN: Catalanes (España), NDLC: Andaluces (España), VLNC: Valencianos (España), CNRS: Canarios (España), BRTN: Bretones (Francia), SWTZ: Suizos, MCDN: Macedonios, GREC: Griegos, ALBN: Albaneses, MLDV: Moldavos, ROMN: Rumanos, UKRN: Ucranianos, TRKS: Turcos, EMRT: Emiratos Árabes, SYR: Sirios, PKST: Pakistaníes, TAML: Tamil, PSHT: Pashtun, Uttar Pradesh (Tripathi et al., 2008): Kayasthas, Mathurs, Rastogies, Vaish, Brahmins, Bhargavas, Chaturvedis, Sunni: Sunitas y Shia: Chiitas. KUNG: Población ¡Kung de Angola , CHBJ: Han de China, ANCESTRAL: población ancestral hipotética.

<i>Alu</i>	BSR	BSQG	BSQN	CTLN	NDLC	VLNC
TPA25	0,434	0,553	0,592	0,608	0,590	0,556
ACE	0,196	0,425	0,320	0,300	0,470	0,387
APO	0,903	0,950	0,970	0,983	0,985	0,940
PV92	0,265	0,239	0,154	0,175	0,194	0,232
FXIIIB	0,238	0,441	0,490	0,500	0,448	0,476
D1	0,313	0,473	0,373	0,350	0,306	0,322
B65	0,422	0,557	0,466	0,525	0,552	0,529

<i>Alu</i>	CNRS	FRNC	SWTZ	MCDN	GREC	ALBN
------------	------	------	------	------	------	------

Material anexo

TPA25	0,609	0,560	0,453	0,521	0,552	0,557
ACE	0,369	0,480	0,372	0,440	0,260	0,467
APO	0,949	0,990	0,942	0,976	0,977	1
PV92	0,250	0,230	0,198	0,167	0,190	0,203
FXIIIB	0,383	0,420	0,477	0,542	0,500	0,600
D1	0,348	0,460	0,337	0,380	0,409	0,267
B65	0,559	0,570	0,580	0,597	0,651	0,670
<i>Alu</i>	MLDV	ROMN	UKRN	TRKS	EMRT	SYR
TPA25	0,569	0,577	0,518	0,427	0,440	0,510
ACE	0,486	0,469	0,404	0,441	0,330	0,400
APO	0,963	0,915	0,964	0,963	0,970	0,930
PV92	0,249	0,200	0,235	0,256	0,300	0,180
FXIIIB	0,488	0,408	0,441	0,473	0,390	0,280
D1	0,383	0,415	0,412	0,284	0,080	0,290
B65	0,571	0,569	0,530	0,706	0,410	0,310
<i>Alu</i>	PKST	TAML	PSHT	Shia	Sunni	Kayastha
TPA25	0,510	0,560	0,550	0,468	0,540	0,523
ACE	0,440	0,690	0,520	0,515	0,558	0,595
APO	0,720	0,810	0,860	0,725	0,743	0,820
PV92	0,300	0,560	0,330	0,345	0,353	0,385
FXIIIB	0,230	0,610	0,570	0,433	0,405	0,433
D1	0,170	0,340	0,270	0,255	0,268	0,235
B65	0,370	0,550	0,490	0,333	0,333	0,323
<i>Alu</i>	Mathur	Rastogi	Vaish	Bhargava	Brahmin	Chaturvedi
TPA25	0,658	0,500	0,500	0,498	0,483	0,528
ACE	0,570	0,605	0,613	0,565	0,525	0,573
APO	0,910	0,888	0,900	0,885	0,910	0,880
PV92	0,388	0,335	0,340	0,293	0,283	0,300
FXIIIB	0,380	0,445	0,470	0,343	0,345	0,378
D1	0,228	0,230	0,225	0,305	0,325	0,290
B65	0,293	0,410	0,415	0,453	0,483	0,448
<i>Alu</i>	KUNG	CHBJ	ANCESTRAL			
TPA25	0,170	0,503	0			
ACE	0,290	0,685	0			
APO	0,880	0,908	0			
PV92	0,200	0,847	0			
FXIIIB	0,170	0,781	0			
D1	0,160	0,259	0			
B65	0,500	0,390	0			

Tabla 6: Número de habitantes pertenecientes a la etnia romaní en el censo de cada país huésped, % de la población romaní frente al total, grado de mestizaje en función de los marcadores de CODIS obtenidos en este estudio. (Fuentes de los datos poblacionales: R. Croacia: Croatian Bureau of Statistics-census, 2011, R. Macedonia: State Statistical Office-Republic of North Macedonia-census, 2021, R. Eslovaquia: Statistical Office of the Slovak Republic-census, 2011, R. Portugal: European Commission-Portugal, 2021., R. Hungría: The Hungarian Central Statistical Office-census, 2011).

	R. Croacia	R. Macedonia	R. Eslovaquia	R. Portugal	R. Hungría
Nº					
Habitantes	16675	46433	105738	50000	315583
%					
Población total	0,4	2,53	2	0,52	3,2
%Mestizaje					
CODIS	44,6	70,1	49,7	36,5	59,7

Tabla 7: Frecuencias alélicas para los 13 STRs pertenecientes al CODIS en las 30 poblaciones de estudio.

	Romaní-croata	Croata	Macedonia	Romaní-macedonia	Eslovaca	Romaní-eslovaca	Portuguesa
CSF1PO							
9	0	0,020	0,175	0,127	0,033	0,007	0,120
10	0,290	0,307	0,285	0,240	0,326	0,272	0,059
11	0,135	0,292	0,235	0,265	0,246	0,228	0,304
12	0,500	0,317	0,140	0,172	0,326	0,402	0,283
13	0,035	0,050	0,005	0,064	0,058	0,051	0,179
14	0	0,005	0,050	0	0,004	0,029	0,029
15	0	0,002	0	0	0,007	0,007	0,001
TPOX							
7	0	0	0	0	0,004	0	0,006
8	0,435	0,527	0,490	0,343	0,599	0,489	0,511
9	0,080	0,087	0,120	0,255	0,084	0,084	0,108
10	0,145	0,042	0,060	0,049	0,073	0,066	0,064
11	0,335	0,300	0,295	0,238	0,219	0,347	0,275
12	0,005	0,037	0,035	0,025	0,022	0,007	0,030
13	0	0	0	0	0	0,007	0,001
D16S539							
8	0	0,007	0,035	0,191	0,007	0,014	0,024
9	0,225	0,085	0,120	0,064	0,098	0,123	0,120
10	0,175	0,050	0,060	0,113	0,062	0,178	0,059
11	0,195	0,330	0,320	0,167	0,330	0,326	0,304
12	0,250	0,327	0,295	0,368	0,337	0,275	0,283
13	0,145	0,162	0,140	0,083	0,156	0,083	0,179
14	0,010	0,037	0,030	0,015	0,011	0	0,029
FGA							
18	0,025	0,005	0	0,025	0,007	0,008	0,014
19	0,100	0,067	0,090	0,103	0,073	0,038	0,066
20	0,080	0,135	0,135	0,157	0,120	0,178	0,139
21	0,145	0,182	0,190	0,186	0,201	0,186	0,181
22	0,190	0,207	0,160	0,088	0,172	0,125	0,171
23	0,075	0,152	0,175	0,113	0,095	0,186	0,151
24	0,295	0,130	0,145	0,191	0,168	0,167	0,137
25	0,030	0,077	0,060	0,044	0,088	0,030	0,078
26	0,060	0,020	0,040	0,020	0,029	0,027	0,034
VWA							
14	0,175	0,117	0,080	0,113	0,144	0,201	0,102
15	0,075	0,070	0,095	0,029	0,114	0,073	0,123
16	0,350	0,177	0,205	0,255	0,197	0,120	0,236
17	0,185	0,325	0,265	0,368	0,261	0,350	0,264
18	0,210	0,222	0,240	0,132	0,201	0,164	0,184
19	0,005	0,075	0,090	0,088	0,064	0,066	0,073
D7S820							
7	0	0,025	0,020	0,015	0,018	0,043	0,017
8	0,070	0,140	0,135	0,118	0,156	0,109	0,155

Procesos microevolutivos en la población romaní residente en el País Vasco

9	0,085	0,197	0,175	0,127	0,145	0,065	0,129
10	0,180	0,300	0,285	0,240	0,264	0,254	0,272
11	0,645	0,162	0,235	0,265	0,221	0,348	0,227
12	0,015	0,147	0,140	0,172	0,159	0,127	0,162
13	0,005	0,022	0,005	0,064	0,033	0,051	0,031
14	0	0	0,005	0	0,004	0,004	0,005
D8S1179							
8	0	0,012	0,020	0,010	0,022	0,011	0,011
9	0	0,007	0,015	0,029	0,011	0,004	0,010
10	0	0,075	0,070	0,039	0,065	0,065	0,086
11	0,185	0,070	0,090	0,118	0,076	0,167	0,097
12	0,205	0,162	0,090	0,049	0,163	0,098	0,126
13	0,255	0,335	0,260	0,245	0,312	0,301	0,289
14	0,140	0,210	0,240	0,250	0,207	0,203	0,232
15	0,120	0,102	0,170	0,201	0,112	0,130	0,116
16	0,095	0,025	0,040	0,059	0,025	0,011	0,028
17	0	0	0,005	0	0,004	0,011	0,003
D13S317							
8	0,265	0,105	0,150	0,191	0,159	0,167	0,009
9	0,210	0,085	0,095	0,064	0,091	0,138	0,031
10	0,040	0,055	0,040	0,113	0,065	0,076	0,061
11	0,110	0,380	0,315	0,167	0,308	0,188	0,340
12	0,270	0,235	0,295	0,368	0,243	0,308	0,377
13	0,090	0,092	0,080	0,083	0,098	0,098	0,171
14	0,015	0,047	0,025	0,015	0,033	0,018	0,010
D21S11							
27	0	0,025	0,010	0,015	0,026	0,004	0,024
28	0,060	0,165	0,135	0,196	0,182	0,123	0,147
29	0,210	0,205	0,265	0,221	0,223	0,268	0,230
30	0,205	0,222	0,185	0,142	0,182	0,196	0,244
30,2	0,055	0,047	0,045	0,064	0,066	0,043	0,037
31	0	0,060	0,045	0,059	0,066	0,047	0,060
31,2	0,140	0,115	0,120	0,020	0,069	0,062	0,111
32	0	0,012	0,005	0	0,026	0,018	0,010
32,2	0,270	0,092	0,145	0,186	0,113	0,152	0,092
33	0	0,002	0	0	0,007	0,007	0,001
33,2	0,050	0,040	0,035	0,074	0,022	0,043	0,031
34	0	0	0	0	0	0,004	0
34,2	0,010	0,007	0,005	0,020	0,011	0,025	0,005
35	0	0	0	0,005	0	0,004	0,001
TH01							
6	0,185	0,227	0,265	0,221	0,201	0,259	0,208
7	0,090	0,125	0,135	0,118	0,142	0,124	0,172
8	0,235	0,110	0,180	0,275	0,084	0,172	0,133
9	0,175	0,202	0,170	0,191	0,230	0,230	0,197
9,3	0,315	0,300	0,245	0,196	0,336	0,215	0,275
D3S1358							
14	0,005	0,112	0,075	0,025	0,145	0,066	0,105
15	0,375	0,265	0,255	0,289	0,203	0,261	0,264

Material anexo

16	0,195	0,255	0,310	0,294	0,225	0,221	0,251
17	0,215	0,195	0,210	0,191	0,188	0,199	0,208
18	0,210	0,160	0,135	0,186	0,207	0,217	0,153
19	0	0,010	0,015	0,010	0,007	0,018	0,011
D18S51							
11	0,010	0,010	0,030	0,010	0,011	0	0,009
12	0,045	0,102	0,135	0,162	0,084	0,080	0,140
13	0,100	0,132	0,135	0,123	0,106	0,058	0,130
14	0,120	0,187	0,120	0,225	0,157	0,164	0,143
15	0,230	0,130	0,175	0,137	0,164	0,150	0,147
16	0,075	0,150	0,165	0,201	0,179	0,182	0,159
17	0,325	0,107	0,085	0,069	0,131	0,208	0,112
18	0,045	0,092	0,070	0,044	0,088	0,073	0,065
19	0,045	0,045	0,035	0,015	0,037	0,011	0,042
20	0	0,027	0,020	0,005	0,029	0,040	0,024
D5S818							
9	0,080	0,042	0,055	0,050	0,051	0,036	0,031
10	0,090	0,087	0,040	0,054	0,087	0,062	0,061
11	0,195	0,312	0,250	0,233	0,348	0,236	0,340
12	0,535	0,360	0,425	0,450	0,333	0,493	0,377
13	0,100	0,182	0,215	0,203	0,163	0,167	0,171
14	0	0,010	0,010	0,010	0,007	0,007	0,010

	Romani- portuguesa	Romani- hungara	Húngara	Vasco- francesa	Francesa	Polaca
CSF1PO						
9	0	0,014	0,038	0	0,054	0,047
10	0,228	0,300	0,307	0,313	0,250	0,278
11	0,256	0,354	0,262	0,458	0,321	0,260
12	0,480	0,277	0,318	0,208	0,304	0,313
13	0,033	0,050	0,063	0	0,071	0,083
14	0,004	0	0,009	0	0	0,014
15	0	0	0	0	0	0
TPOX						
7	0	0	0,004	0	0	0,001
8	0,435	0,550	0,596	0,521	0,500	0,580
9	0,061	0,100	0,099	0,104	0,125	0,093
10	0,102	0,032	0,074	0,104	0,071	0,069
11	0,390	0,304	0,204	0,271	0,304	0,226
12	0,012	0,014	0,027	0	0	0,031
13	0	0	0	0	0	0
D16S539						
8	0,020	0,004	0,022	0	0,107	0,005
9	0,098	0,182	0,103	0,146	0,161	0,093
10	0,069	0,145	0,054	0,021	0,071	0,042
11	0,382	0,232	0,296	0,188	0,25	0,284
12	0,195	0,264	0,303	0,396	0,304	0,358
13	0,224	0,154	0,182	0,229	0,089	0,192

Procesos microevolutivos en la población romaní residente en el País Vasco

	14	0,004	0,014	0,040	0,021	0,018	0,026
FGA							
	18	0	0,004	0,020	0,042	0	0,015
	19	0,061	0,109	0,076	0,167	0,071	0,093
	20	0,220	0,136	0,157	0,104	0,161	0,138
	21	0,077	0,132	0,195	0,250	0,125	0,179
	22	0,171	0,159	0,191	0,042	0,232	0,191
	23	0,110	0,132	0,114	0,250	0,125	0,138
	24	0,163	0,136	0,123	0,063	0,143	0,105
	25	0,024	0,082	0,056	0,083	0,089	0,078
	26	0,057	0,036	0,040	0	0,036	0,020
VWA							
	14	0,175	0,123	0,108	0,125	0,161	0,107
	15	0,073	0,082	0,114	0,125	0,071	0,101
	16	0,187	0,259	0,206	0,229	0,179	0,179
	17	0,370	0,323	0,307	0,208	0,268	0,283
	18	0,081	0,177	0,173	0,208	0,232	0,209
	19	0,077	0,023	0,072	0,063	0,089	0,106
D7S820							
	7	0,004	0,023	0,009	0,021	0,018	0
	8	0,195	0,132	0,143	0,208	0,214	0,005
	9	0,154	0,145	0,137	0,208	0,125	0,093
	10	0,126	0,191	0,294	0,292	0,304	0,042
	11	0,256	0,309	0,235	0,146	0,143	0,284
	12	0,256	0,177	0,150	0,104	0,125	0,358
	13	0,004	0,009	0,025	0,021	0,054	0,192
	14	0,004	0,014	0,004	0	0,018	0,026
D8S1179							
	8	0	0	0,011	0	0	0,007
	9	0	0,004	0,016	0,021	0,018	0,008
	10	0,020	0,100	0,076	0,125	0,125	0,064
	11	0,122	0,082	0,081	0,042	0,071	0,057
	12	0,142	0,118	0,182	0,146	0,089	0,157
	13	0,341	0,300	0,287	0,313	0,304	0,358
	14	0,215	0,236	0,240	0,146	0,304	0,212
	15	0,138	0,100	0,083	0,208	0,071	0,101
	16	0,020	0,054	0,020	0	0,018	0,033
	17	0	0,004	0,002	0	0	0,002
D13S317							
	8	0,203	0,173	0,137	0,087	0,107	0,138
	9	0,053	0,086	0,096	0,022	0,161	0,074
	10	0,110	0,050	0,049	0,087	0,071	0,045
	11	0,228	0,345	0,327	0,391	0,250	0,371
	12	0,386	0,250	0,260	0,261	0,304	0,243
	13	0,008	0,073	0,092	0,109	0,089	0,083
	14	0,012	0,018	0,038	0,043	0,018	0,043
D21S11							
	27	0	0,023	0,049	0,042	0,036	0,023
	28	0,073	0,141	0,191	0,208	0,125	0,158

Material anexo

29	0,321	0,232	0,195	0,250	0,196	0,191
30	0,150	0,218	0,217	0,229	0,321	0,219
30,2	0,049	0,050	0,049	0	0,036	0,058
31	0,024	0,032	0,063	0	0,089	0,066
31,2	0,061	0,086	0,096	0,125	0,125	0,112
32	0,008	0,004	0,011	0	0	0,016
32,2	0,248	0,186	0,076	0,125	0,054	0,092
33	0	0	0	0	0	0
33,2	0,061	0,023	0,04	0,021	0	0,050
34	0	0	0	0	0	0
34,2	0	0,004	0,007	0	0	0,005
35	0	0	0,003	0	0	0,001
TH01						
6	0,207	0,195	0,220	0,250	0,286	0,219
7	0,110	0,132	0,158	0,042	0,125	0,145
8	0,199	0,214	0,114	0,167	0,107	0,085
9	0,179	0,200	0,209	0,333	0,161	0,203
9,3	0,301	0,250	0,283	0,208	0,321	0,333
D3S1358						
14	0,037	0,054	0,087	0,188	0,071	0,142
15	0,317	0,259	0,247	0,229	0,339	0,243
16	0,293	0,332	0,244	0,250	0,250	0,255
17	0,191	0,168	0,231	0,125	0,107	0,222
18	0,163	0,182	0,161	0,208	0,214	0,119
19	0	0,004	0,020	0	0,018	0,013
D18S51						
11	0	0,004	0,016	0	0	0,016
12	0,077	0,118	0,112	0,239	0,111	0,106
13	0,069	0,045	0,150	0,065	0,148	0,114
14	0,252	0,209	0,170	0,152	0,204	0,126
15	0,236	0,141	0,119	0,174	0,130	0,157
16	0,102	0,173	0,112	0,217	0,148	0,190
17	0,142	0,118	0,117	0	0,093	0,124
18	0,049	0,050	0,083	0,022	0,093	0,083
19	0,028	0,045	0,052	0,087	0	0,035
20	0,020	0,036	0,043	0,022	0,037	0,027
D5S818						
9	0,004	0,054	0,040	0,042	0,018	0,056
10	0,085	0,095	0,078	0,042	0,036	0,071
11	0,211	0,264	0,318	0,354	0,339	0,324
12	0,423	0,432	0,370	0,313	0,375	0,386
13	0,260	0,141	0,166	0,208	0,196	0,150
14	0,016	0,014	0,020	0,042	0,036	0,009
CSF1PO						
9	0,010	0,025	0,019	0,007	0,011	0,028

Procesos microevolutivos en la población romaní residente en el País Vasco

10	0,289	0,253	0,296	0,217	0,242	0,288
11	0,327	0,316	0,333	0,333	0,264	0,297
12	0,278	0,342	0,278	0,319	0,393	0,316
13	0,062	0,052	0,074	0,109	0,070	0,042
14	0,018	0,009	0	0,007	0,017	0,019
15	0	0,003	0	0,007	0	0
TPOX						
7	0,003	0,003	0	0	0,006	0,009
8	0,490	0,507	0,464	0,275	0,432	0,401
9	0,137	0,112	0,089	0,174	0,108	0,085
10	0,059	0,064	0,143	0,101	0,082	0,075
11	0,296	0,282	0,304	0,435	0,33	0,382
12	0,015	0,033	0	0,014	0,037	0,033
13	0	0,001	0	0	0,006	0,005
D16S539						
8	0,015	0,029	0,018	0,022	0,075	0,108
9	0,108	0,119	0,089	0,232	0,156	0,118
10	0,070	0,054	0,054	0,116	0,073	0,123
11	0,302	0,287	0,357	0,275	0,369	0,311
12	0,314	0,298	0,304	0,181	0,198	0,217
13	0,165	0,186	0,125	0,159	0,101	0,099
14	0,026	0,027	0,054	0,014	0,020	0,024
FGA						
18	0,008	0,011	0	0	0,022	0
19	0,085	0,072	0,054	0,065	0,056	0,052
20	0,149	0,117	0,125	0,101	0,095	0,099
21	0,168	0,184	0,179	0,159	0,154	0,151
22	0,183	0,164	0,179	0,145	0,156	0,104
23	0,152	0,150	0,143	0,159	0,156	0,193
24	0,137	0,162	0,179	0,123	0,182	0,255
25	0,070	0,099	0,125	0,167	0,098	0,090
26	0,039	0,026	0,018	0,043	0,036	0,038
VWA						
14	0,106	0,095	0,071	0,094	0,106	0,104
15	0,119	0,120	0,143	0,174	0,056	0,057
16	0,237	0,210	0,161	0,297	0,307	0,250
17	0,271	0,287	0,321	0,232	0,271	0,297
18	0,173	0,200	0,232	0,174	0,159	0,189
19	0,085	0,074	0,071	0,029	0,084	0,099
D7S820						
7	0,023	0,016	0	0,022	0,020	0,024
8	0,160	0,162	0,143	0,239	0,215	0,231
9	0,157	0,130	0,125	0,072	0,087	0,104
10	0,253	0,270	0,321	0,203	0,212	0,208
11	0,240	0,217	0,232	0,283	0,218	0,222
12	0,126	0,176	0,143	0,138	0,207	0,165
13	0,039	0,029	0,036	0,022	0,042	0,042
14	0,003	0,003	0	0,022	0	0,005
D8S1179						

Material anexo

8	0,008	0,027	0,036	0	0,011	0,014
9	0,005	0,011	0,036	0,007	0,014	0
10	0,080	0,085	0,107	0,196	0,176	0,146
11	0,090	0,056	0,071	0,145	0,067	0,104
12	0,124	0,083	0,161	0,036	0,073	0,113
13	0,345	0,300	0,250	0,174	0,209	0,156
14	0,224	0,244	0,107	0,181	0,190	0,179
15	0,093	0,132	0,161	0,196	0,162	0,217
16	0,023	0,052	0,054	0,058	0,087	0,066
17	0,008	0,009	0,018	0,007	0,008	0,005
D13S317						
8	0,144	0,137	0,161	0,138	0,190	0,212
9	0,046	0,069	0,089	0,159	0,064	0,090
10	0,044	0,070	0,018	0,101	0,112	0,075
11	0,322	0,305	0,268	0,283	0,232	0,231
12	0,235	0,287	0,286	0,225	0,307	0,278
13	0,160	0,100	0,125	0,051	0,075	0,075
14	0,044	0,031	0,054	0,022	0,014	0,024
D21S11						
27	0,013	0,042	0,071	0,014	0,011	0,005
28	0,149	0,083	0,161	0,145	0,140	0,137
29	0,242	0,292	0,214	0,181	0,145	0,203
30	0,278	0,292	0,268	0,210	0,246	0,212
30,2	0,023	0	0	0,014	0,042	0,014
31	0,052	0,042	0,018	0,051	0,022	0,057
31,2	0,121	0,125	0,107	0,072	0,134	0,113
32	0,010	0,042	0,018	0,007	0,003	0
32,2	0,080	0,042	0,125	0,174	0,142	0,160
33	0,005	0	0	0	0,003	0
33,2	0,018	0,042	0,018	0,072	0,081	0,057
34	0,005	0	0	0	0,003	0
34,2	0	0	0	0,022	0,011	0,009
35	0,003	0	0	0	0,003	0
TH01						
6	0,490	0,271	0,250	0,312	0,232	0,330
7	0,137	0,155	0,161	0,109	0,187	0,179
8	0,113	0,120	0,107	0,174	0,145	0,085
9	0,276	0,201	0,321	0,232	0,282	0,264
9,3	0,209	0,231	0,161	0,167	0,145	0,137
D3S1358						
14	0,198	0,065	0,018	0,072	0,056	0,061
15	0,175	0,243	0,214	0,304	0,251	0,274
16	0,015	0,237	0,286	0,239	0,318	0,297
17	0,010	0,250	0,268	0,239	0,251	0,231
18	0,126	0,185	0,196	0,145	0,112	0,118
19	0,162	0,015	0,018	0	0,008	0,019
D18S51						
11	0,178	0,027	0	0,036	0,020	0,019
12	0,142	0,188	0,125	0,116	0,089	0,042

Procesos microevolutivos en la población romaní residente en el País Vasco

13	0,103	0,184	0,107	0,152	0,179	0,137
14	0,111	0,143	0,304	0,283	0,246	0,321
15	0,077	0,114	0,107	0,167	0,131	0,160
16	0,041	0,119	0,071	0,072	0,117	0,151
17	0,021	0,094	0,143	0,065	0,064	0,104
18	0,028	0,067	0,089	0,036	0,064	0,009
19	0,085	0,020	0,018	0,022	0,047	0,024
20	0,320	0,027	0	0,029	0,022	0,009
D5S818						
9	0,366	0,041	0,036	0,022	0,036	0,009
10	0,180	0,061	0,071	0,159	0,109	0,118
11	0,010	0,351	0,464	0,290	0,349	0,335
12	0,366	0,359	0,232	0,377	0,344	0,368
13	0,180	0,177	0,179	0,145	0,142	0,156
14	0,010	0,007	0,018	0,007	0,008	0,009
Rajastán Pastún Sindhi Saraiki Balochi Iraní						
CSF1PO						
9	0,032	0,019	0,015	0,010	0,041	0,017
10	0,213	0,215	0,291	0,308	0,257	0,248
11	0,304	0,364	0,252	0,293	0,339	0,352
12	0,357	0,294	0,374	0,313	0,312	0,304
13	0,072	0,079	0,053	0,063	0,028	0,065
14	0,017	0,014	0,01	0,005	0,009	0,013
15	0,003	0,005	0	0,010	0	0
TPOX						
7	0,002	0	0	0	0	0
8	0,377	0,449	0,437	0,380	0,417	0,488
9	0,129	0,122	0,175	0,106	0,124	0,113
10	0,070	0,112	0,053	0,115	0,069	0,117
11	0,376	0,280	0,311	0,370	0,353	0,233
12	0,045	0,033	0,019	0,024	0,018	0,046
13	0,002	0,005	0	0,005	0	0
D16S539						
8	0,076	0,028	0,073	0,067	0,097	0,051
9	0,166	0,154	0,160	0,168	0,051	0,110
10	0,097	0,159	0,117	0,058	0,055	0,093
11	0,296	0,336	0,325	0,322	0,327	0,411
12	0,204	0,220	0,233	0,226	0,341	0,191
13	0,130	0,089	0,087	0,144	0,065	0,123
14	0,028	0,014	0,005	0,014	0,060	0,017
FGA						
18	0,002	0,023	0,005	0	0,005	0,009
19	0,044	0,014	0,044	0,063	0,037	0,044
20	0,120	0,117	0,097	0,125	0,078	0,113
21	0,133	0,159	0,117	0,154	0,106	0,139
22	0,143	0,187	0,083	0,144	0,151	0,196

Material anexo

23	0,184	0,182	0,180	0,192	0,174	0,122
24	0,186	0,150	0,204	0,082	0,261	0,139
25	0,107	0,103	0,180	0,130	0,133	0,126
26	0,041	0,037	0,034	0,048	0,041	0,044
VWA						
14	0,109	0,093	0,073	0,125	0,087	0,164
15	0,074	0,093	0,073	0,120	0,096	0,084
16	0,240	0,271	0,248	0,207	0,216	0,188
17	0,308	0,280	0,296	0,308	0,307	0,280
18	0,179	0,154	0,218	0,159	0,197	0,188
19	0,067	0,079	0,083	0,048	0,069	0,084
D7S820						
7	0,032	0,089	0,049	0,019	0,041	0,052
8	0,217	0,168	0,228	0,178	0,183	0,160
9	0,083	0,093	0,058	0,072	0,064	0,086
10	0,221	0,206	0,233	0,231	0,225	0,220
11	0,241	0,192	0,233	0,317	0,298	0,276
12	0,158	0,220	0,175	0,163	0,170	0,168
13	0,045	0,028	0,019	0,010	0,014	0,035
14	0,003	0,005	0,005	0	0,005	0,004
D8S1179						
8	0,012	0,023	0,044	0,005	0,014	0,004
9	0,005	0,009	0	0,014	0,005	0,012
10	0,170	0,122	0,199	0,144	0,119	0,111
11	0,080	0,061	0,083	0,053	0,046	0,083
12	0,095	0,098	0,092	0,087	0,073	0,103
13	0,150	0,257	0,160	0,236	0,211	0,278
14	0,200	0,159	0,189	0,236	0,257	0,214
15	0,187	0,187	0,180	0,159	0,188	0,123
16	0,081	0,070	0,044	0,067	0,060	0,071
17	0,016	0,009	0,010	0	0,023	0
D13S317						
8	0,192	0,166	0,150	0,120	0,097	0,136
9	0,084	0,066	0,078	0,082	0,051	0,080
10	0,092	0,076	0,073	0,101	0,055	0,088
11	0,274	0,327	0,282	0,303	0,327	0,332
12	0,243	0,251	0,296	0,260	0,341	0,276
13	0,080	0,090	0,083	0,091	0,065	0,068
14	0,023	0,024	0,039	0,029	0,060	0,012
D21S11						
27	0,010	0,019	0	0	0,009	0,012
28	0,131	0,173	0,165	0,154	0,229	0,172
29	0,185	0,164	0,175	0,197	0,220	0,244
30	0,191	0,173	0,160	0,168	0,170	0,212
30,2	0,030	0,028	0,024	0,010	0,028	0,028
31	0,054	0,042	0,053	0,048	0,037	0,044
31,2	0,124	0,140	0,146	0,130	0,151	0,112
32	0,009	0,009	0,005	0	0	0,008
32,2	0,191	0,154	0,170	0,178	0,092	0,124

Procesos microevolutivos en la población romaní residente en el País Vasco

33	0,001	0	0	0	0	0,004
33,2	0,064	0,070	0,083	0,101	0,064	0,024
34	0	0	0	0	0	0
34,2	0,006	0,009	0,015	0,010	0	0,012
35	0	0	0	0		0,004
TH01						
6	0,272	0,280	0,296	0,207	0,197	0,278
7	0,184	0,215	0,209	0,226	0,211	0,222
8	0,134	0,150	0,194	0,135	0,202	0,147
9	0,254	0,206	0,189	0,250	0,239	0,202
9,3	0,144	0,150	0,097	0,173	0,142	0,143
D3S1358						
14	0,040	0,047	0,029	0,048	0,037	0,044
15	0,289	0,304	0,218	0,274	0,408	0,254
16	0,341	0,276	0,296	0,288	0,239	0,282
17	0,227	0,196	0,257	0,226	0,220	0,226
18	0,096	0,168	0,155	0,159	0,083	0,191
19	0,004	0,009	0,034	0	0,014	0,004
D18S51						
11	0,027	0,014	0,024	0,034	0,032	0,031
12	0,090	0,098	0,083	0,091	0,078	0,152
13	0,139	0,122	0,112	0,096	0,124	0,121
14	0,258	0,215	0,180	0,264	0,206	0,228
15	0,166	0,168	0,175	0,188	0,193	0,134
16	0,110	0,150	0,112	0,144	0,161	0,094
17	0,088	0,093	0,131	0,072	0,069	0,116
18	0,046	0,051	0,029	0,029	0,055	0,054
19	0,032	0,047	0,092	0,038	0,073	0,022
20	0,028	0,033	0,029	0,024	0	0,009
D5S818						
9	0,023	0,042	0,073	0,019	0,060	0,060
10	0,120	0,182	0,092	0,115	0,073	0,076
11	0,347	0,280	0,364	0,409	0,367	0,272
12	0,337	0,322	0,335	0,288	0,307	0,412
13	0,152	0,154	0,131	0,149	0,161	0,144
14	0,017	0,014	0,005	0,014	0,014	0,028

Tabla 8: Frecuencias de haplogrupos del cromosoma Y para poblaciones romaníes, sus correspondientes poblaciones control europeas, población europea general y poblaciones del Sur de Asia.

Haplogrupo	Búlgara	Romaní-búlgara	Griega	Romaní-griega	Húngara	Romaní-húngara
A	0	0	0	0	0	0
B	0	0	0	0	0	0
C	0	0	0	0	0	0,005
D	0	0	0	0	0	0
E	0,216	0,095	0,290	0,103	0,070	0,079
F	0	0	0	0	0	0
G	0,010	0,004	0,050	0	0,053	0,035
H	0,010	0,457	0	0,672	0,053	0,342
I	0,340	0,228	0,140	0,103	0,262	0,129
J	0,093	0,177	0,170	0,103	0,102	0,163
K	0	0	0	0	0	0
L	0	0	0,010	0	0	0
M	0	0	0	0	0	0
N	0,010	0	0	0	0,005	0,005
O	0	0	0	0	0	0
P	0	0	0	0	0	0
Q	0	0	0	0	0	0
R	0,299	0,039	0,310	0,017	0,449	0,233
T	0,021	0	0,030	0	0	0

Haplogrupo	Rumana	Romaní-rumana	Eslovaca	Romaní-eslovaca	Española	Romaní-española
A	0	0	0	0	0	0
B	0	0	0	0	0	0
C	0	0	0	0	0	0
D	0	0	0	0	0	0
E	0,174	0,011	0,043	0,056	0,067	0
F	0	0	0	0	0	0
G	0,021	0	0,050	0	0,020	0
H	0	0,500	0,006	0,542	0	0,200
I	0,292	0,114	0,211	0,065	0,103	0,175
J	0,146	0,318	0,037	0,234	0,079	0,450
K	0	0	0	0	0	0
L	0	0	0	0	0	0
M	0	0	0	0	0	0
N	0,007	0	0,031	0	0,008	0
O	0	0	0	0	0	0
P	0	0	0	0	0	0
Q	0	0	0	0	0	0
R	0,347	0,057	0,615	0,103	0,700	0,175
T	0,007	0	0,006	0	0,024	0

Procesos microevolutivos en la población romaní residente en el País Vasco

Haplogrupo	Ucraniana	Romaní-ucraniana	Francesa	Holandesa	Alemana	Albana
A	0	0	0	0	0	0
B	0	0	0	0	0	0
C	0	0	0	0	0	0
D	0	0	0	0	0	0,020
E	0,095	0,143	0,087	0,037	0,062	0,216
F	0	0	0	0	0	0
G	0,024	0	0	0	0	0,020
H	0	0,286	0	0	0	0
I	0,286	0,238	0,174	0,222	0,375	0,196
J	0,071	0,238	0,173	0	0	0,275
K	0	0	0	0	0	0
L	0	0	0	0	0	0
M	0	0	0	0	0	0
N	0,024	0,048	0	0	0	0
O	0	0	0	0	0	0
P	0	0	0,043	0	0	0
Q	0	0	0	0	0	0
R	0,500	0,048	0,522	0,741	0,562	0,274
T	0	0	0	0	0	0

Haplogrupo	Macedonia	Polaca	Afgana	Iraní	Iraquí	Jordana
A	0	0	0	0	0	0,007
B	0	0	0,014	0,020	0	0
C	0	0	0,148	0,007	0	0
D	0	0	0	0	0	0
E	0,150	0,036	0,025	0,053	0,120	0,262
F	0	0	0	0,013	0,030	0,007
G	0	0	0,039	0,133	0,022	0,041
H	0	0	0,044	0,020	0	0
I	0,200	0,236	0,015	0	0,007	0,034
J	0,200	0	0,142	0,347	0,562	0,435
K	0	0	0	0,027	0,086	0,007
L	0	0	0,118	0,053	0	0
M	0	0	0,005	0	0	0
N	0	0	0,015	0,020	0	0
O	0	0	0,034	0,007	0	0
P	0	0	0	0	0	0,013
Q	0	0	0,069	0,040	0	0
R	0,450	0,728	0,324	0,260	0,173	0,192
T	0	0	0,010	0	0	0

Haplogrupo	Turca	Libanesa	Siria	Dhodia	Dubla	Konkana
A	0,004	0	0	0	0	0
B	0,002	0	0	0	0	0
C	0,013	0,010	0	0,032	0,143	0,125
D	0	0	0	0	0	0

Material anexo

E	0,111	0,298	0,117	0	0	0
F	0	0,067	0,018	0,095	0,024	0,083
G	0,109	0,029	0,027	0	0	0
H	0,006	0	0,018	0,333	0,476	0,292
I	0,054	0,019	0,009	0	0	0
J	0,335	0,461	0,549	0,079	0,119	0
K	0,025	0,010	0,063	0,079	0	0
L	0,042	0,019	0	0,048	0,048	0
M	0	0	0	0	0	0
N	0,037	0	0	0	0	0
O	0,002	0	0	0	0	0
P	0	0	0,009	0	0	0
Q	0,019	0	0	0	0	0
R	0,241	0,087	0,189	0,333	0,190	0,500
T	0	0	0	0	0	0

Haplogrupo	Vasava	Gamit	ValviChaudhari	Nana Chaudhari	Mota Chaudhari	PavagadhiChaudhari
A	0	0	0	0	0	0
B	0	0	0	0	0	0
C	0,083	0,056	0,219	0,040	0,074	0
D	0	0	0	0	0	0
E	0	0	0	0	0	0
F	0	0	0,031	0,040	0	0,034
G	0	0	0	0	0	0
H	0,625	0,611	0,156	0,600	0,667	0,069
I	0	0	0	0	0	0
J	0,042	0,167	0,250	0,040	0	0,207
K	0	0	0	0,040	0,037	0,034
L	0	0	0,031	0,040	0,074	0
M	0	0	0	0	0	0
N	0	0	0	0	0	0
O	0	0	0	0	0	0
P	0	0	0	0	0	0
Q	0,042	0,056	0	0,120	0,074	0,034
R	0,208	0,111	0,313	0,080	0,074	0,621
T	0	0	0	0	0	0

Haplogrupo	Punjabí	Gujarati	Chenchu	Koya
A	0	0	0	0
B	0	0	0	0
C	0,030	0,172	0,049	0
D	0	0	0	0
E	0	0	0	0
F	0	0	0	0,268
G	0	0	0	0
H	0,045	0,138	0,366	0,707
I	0	0	0	0

Procesos microevolutivos en la población romaní residente en el País Vasco

J	0,212	0,207	0,073	0
K	0	0,034	0	0
L	0,121	0,103	0,146	0
M	0	0	0	0
N	0	0	0	0
O	0	0	0	0
P	0	0,034	0	0
Q	0	0	0	0
R	0,591	0,310	0,366	0,024
T	0	0	0	0

Tabla 9: Frecuencias de Haplogrupos de ADNmt para poblaciones romaníes, sus correspondientes poblaciones huésped y poblaciones del Sur de Asia.

Haplogrupo	Búlgara	Romaní-búlgara	Griega	Romaní-griega	Húngara	Romaní-húngara	Rumana
C	0	0	0	0	0,018	0,009	0
D	0	0	0,010	0,019	0	0	0,012
F1B	0	0	0	0	0	0	0
H*	0,478	0,350	0,378	0,154	0,345	0,339	0,395
HV	0,044	0,040	0,071	0	0,055	0,028	0,093
I	0,033	0,036	0,031	0,058	0,018	0,055	0,012
J	0,033	0,090	0,102	0,058	0,036	0,101	0,105
K	0,022	0,018	0,061	0	0,109	0,018	0,023
L2,6	0	0	0	0	0	0	0
L2a	0	0	0	0	0	0,009	0
M*	0,011	0,238	0,010	0,442	0	0,211	0,012
N*	0	0	0,010	0	0	0,009	0,012
N1	0	0,009	0,020	0,019	0	0	0,023
R*	0,011	0,004	0	0	0,018	0	0
R0a	0,022	0	0,020	0	0	0	0
R10	0	0	0	0	0	0	0
T	0,100	0,018	0,031	0,058	0,164	0,064	0,047
U	0,022	0,027	0,010	0	0	0,009	0,012
U1	0,011	0	0,010	0	0	0	0,012
U2	0	0,004	0	0	0	0	0,023
U3	0	0,081	0,031	0,038	0,018	0,018	0
U4	0,022	0,013	0,051	0	0,036	0,018	0,012
U5	0,100	0,022	0,071	0	0,091	0,073	0,128
U6	0	0	0,010	0	0	0	0
U7	0	0	0	0	0	0	0,012
V	0,044	0	0,051	0	0,018	0,009	0,023
W	0,011	0	0,020	0	0,073	0	0,047
X	0,033	0,049	0	0,154	0	0,028	0

Material anexo

Haplogrupo	Romaní- rumana	Eslovaca	Romaní- eslovaca	Española	Romaní- española	Ucraniana
C	0	0	0	0	0,008	0
D	0	0,006	0	0	0	0
F1B	0	0	0	0	0	0
H*	0,329	0,408	0,311	0,481	0,126	0,375
HV	0	0,064	0,007	0,025	0,008	0,025
I	0,055	0	0,319	0,013	0	0
J	0,027	0,127	0,052	0,089	0,118	0,050
K	0,041	0,070	0	0,038	0,024	0,075
L2,6	0	0	0	0,013	0	0
L2a	0	0	0	0	0	0
M*	0,397	0	0,081	0,025	0,220	0
N*	0,068	0	0,007	0	0	0
N1	0	0,006	0	0	0	0
R*	0	0,006	0	0	0,008	0
R0a	0	0	0	0	0	0
R10	0	0	0	0	0	0
T	0,014	0,057	0,104	0,114	0,016	0,125
U	0	0,032	0,015	0,063	0,031	0,025
U1	0	0	0	0	0	0
U2	0	0,019	0,044	0,013	0	0
U3	0,014	0,013	0	0	0,370	0,05
U4	0	0,025	0,007	0,038	0	0,100
U5	0,014	0,083	0,052	0,063	0,024	0,075
U6	0	0	0	0,013	0,024	0
U7	0	0	0	0	0	0
V	0	0,038	0	0,013	0,024	0,025
W	0,014	0,019	0	0	0	0,050
X	0,027	0,025	0	0	0	0,025

Haplogrupo	Romaní- ucraniana	Austriaca	Francesa	Alemana	Italia-Toscana	Vasca
C	0	0	0	0	0	0
D	0	0	0	0	0	0
F1B	0	0	0	0	0	0
H*	0,059	0,465	0,425	0,499	0,391	0,596
HV	0	0,020	0,063	0,033	0,102	0,109
I	0,118	0	0	0	0	0
J	0,765	0,101	0,051	0,072	0,056	0,026
K	0	0,071	0,105	0,075	0,068	0,038
L2,6	0	0	0	0	0	0
L2a	0	0	0	0	0	0
M*	0	0,020	0,018	0,003	0,009	0
N*	0	0,010	0,015	0,030	0,025	0
N1	0	0,020	0,024	0,024	0,019	0
R*	0	0	0	0	0	0

R0a	0	0,010	0,009	0	0,025	0
R10	0	0	0	0	0	0
T	0	0,101	0,111	0,093	0,112	0,058
U	0	0,030	0,009	0	0	0,006
U1	0	0	0	0,006	0,006	0
U2	0	0,010	0,021	0,009	0,022	0
U3	0	0,010	0,006	0,012	0,028	0
U4	0	0,040	0,021	0,024	0,019	0
U5	0,059	0,081	0,087	0,084	0,04	0,122
U6	0	0	0,006	0	0,006	0
U7	0	0	0,003	0,003	0,025	0,006
V	0	0	0	0	0	0
W	0	0	0	0	0	0
X	0	0,010	0,012	0,009	0,025	0,013

Haplogrupo	Polaca	Rajput	Punjabí	Pastún	Baluchi	Brahui
C	0	0	0	0	0	0
D	0	0	0	0	0	0
F1B	0	0	0,02	0	0	0
H*	0,417	0	0,073	0,083	0,205	0,263
HV	0,052	0	0,013	0,028	0,103	0,053
I	0	0	0,007	0,028	0	0
J	0,094	0	0,033	0,083	0,077	0,079
K	0,039	0	0,007	0,056	0,026	0
L2,6	0	0	0	0	0	0
L2a	0	0	0	0	0	0
M*	0,022	0,714	0,360	0,333	0,333	0,264
N*	0,042	0	0	0	0	0
N1	0,028	0	0,020	0	0,052	0,026
R*	0	0,114	0,093	0,056	0	0
R0a	0	0	0	0	0	0
R10	0	0,086	0,020	0	0	0
T	0,098	0	0,027	0	0	0,053
U	0,006	0	0	0	0	0
U1	0	0	0,013	0	0	0
U2	0,007	0,057	0,100	0,083	0,026	0,053
U3	0,050	0	0	0	0	0
U4	0,046	0	0	0	0,026	0
U5	0,072	0	0,027	0,056	0,026	0
U6	0	0	0	0	0	0
U7	0,002	0,029	0,087	0,083	0,026	0,105
V	0	0	0	0	0	0
W	0	0	0,053	0,083	0	0
X	0,020	0	0	0,028	0	0

Material anexo

Haplogrupo	Gujarati	Hunza Burusho	Hazara	Pakistaní-Sur	Makrani	Parsi
C	0	0	0	0	0	0
D	0	0	0	0	0	0
F1B	0	0	0	0	0	0
H*	0,059	0,068	0,130	0,120	0,030	0,023
HV	0	0	0,043	0,040	0,061	0,114
I	0	0	0	0	0	0
J	0,029	0,068	0	0,010	0,061	0
K	0	0,045	0	0	0	0
L2,6	0	0	0	0	0	0
L2a	0	0	0	0	0	0
M*	0,441	0,250	0,303	0,490	0,091	0,545
N*	0,088	0,023	0	0,010	0	0
N1	0	0,045	0	0,030	0,030	0
R*	0	0	0	0	0	0
R0a	0	0	0	0,020	0	0
R10	0	0	0	0	0	0
T	0	0,046	0	0,010	0	0,113
U	0	0	0	0,010	0,091	0
U1	0	0,068	0,043	0	0,030	0,045
U2	0,176	0,114	0	0,110	0,090	0
U3	0	0,023	0	0	0	0
U4	0	0,045	0,087	0	0	0,136
U5	0	0,023	0,087	0	0	0
U6	0	0	0	0	0	0
U7	0,088	0,068	0,043	0,050	0,030	0,023
V	0	0	0	0	0	0
W	0	0	0	0	0	0
X	0	0,023	0	0,010	0	0

Haplogrupo	Pathan	Sindhi	Siria	Turca	Gilaki	Iraní
C	0	0	0	0	0	0
D	0	0	0	0	0	0
F1B	0	0	0	0	0	0
H*	0,045	0,087	0,229	0,321	0,135	0,174
HV	0,023	0,043	0,046	0,062	0,243	0,064
I	0	0	0	0	0	0
J	0,068	0	0,116	0,094	0,162	0,138
K	0,023	0	0,052	0,056	0,027	0,071
L2,6	0	0	0	0	0	0
L2a	0	0	0	0	0	0
M*	0,295	0,304	0,018	0,038	0,054	0,050
N*	0,045	0,174	0,040	0,029	0	0,041
N1	0,046	0,087	0,052	0,038	0,027	0,046
R*	0	0	0	0	0	0
R0a	0,068	0	0,030	0,009	0	0,014

Procesos microevolutivos en la población romaní residente en el País Vasco

R10	0	0	0	0	0	0
T	0,090	0	0,116	0,094	0,162	0,080
U	0,023	0	0,006	0,018	0,027	0,011
U1	0	0	0,034	0,038	0,027	0,030
U2	0,159	0,173	0,006	0,015	0	0,009
U3	0	0	0,085	0,053	0,027	0,025
U4	0	0	0,009	0,012	0	0,011
U5	0,023	0,043	0,012	0,026	0	0,037
U6	0	0	0,012	0	0	0,002
U7	0	0,087	0,012	0,009	0,108	0,094
V	0	0	0	0	0	0
W	0	0	0	0	0	0
X	0	0	0,015	0,038	0	0,032

Haplogrupo	Iraqí	Jordana	Mazandaria	Palestina	Persia	Libanesa
C	0	0	0	0	0	0
D	0	0	0	0	0	0
F1B	0	0	0	0	0	0
H*	0,233	0,300	0,143	0,282	0,143	0,267
HV	0,129	0,040	0,095	0,017	0,191	0,051
I	0	0	0	0	0	0
J	0,129	0,061	0,238	0,094	0,167	0,045
K	0,052	0,034	0	0,077	0,071	0,057
L2,6	0	0	0	0	0	0
L2a	0	0	0	0	0	0
M*	0,009	0,034	0	0,034	0,048	0,045
N*	0	0,012	0,095	0,026	0,024	0,023
N1	0,017	0,030	0,143	0,026	0,048	0,091
R*	0	0	0	0	0	0
R0a	0,043	0,018	0	0,026	0	0,023
R10	0	0	0	0	0	0
T	0,078	0,091	0,143	0,128	0,095	0,102
U	0,017	0,010	0	0,009	0	0,006
U1	0,009	0,051	0	0,009	0	0,051
U2	0,026	0,004	0	0,009	0,048	0
U3	0,060	0,107	0	0,009	0,048	0,034
U4	0,017	0,004	0	0,017	0,024	0,006
U5	0,009	0,004	0	0,009	0,024	0,045
U6	0,009	0,004	0	0,009	0	0,023
U7	0,026	0,016	0	0,026	0,024	0
V	0	0	0	0	0	0
W	0	0	0	0	0	0
X	0,009	0,018	0	0,034	0	0,057

Tabla 10: Número de habitantes pertenecientes a la etnia romaní en el censo de cada país huésped, % de la población romaní frente al total, grado de mestizaje en función de los marcadores del ADNmt y cromosoma Y obtenidos en este estudio. (Fuentes de los datos poblacionales: Romaní-búlgara: National Statistical Institute. Republic of Bulgaria-census, 2011, Romaní-griega: European Commission. Greece-census, 2011, Romaní-húngara: Central Statistical Office (CSO) Hungary-census, 2011, Romaní-rumana: National Statistical Institute. Republic of Romania-census, 2011, Romaní-eslovaca: Statistical Office of the Slovak Republic-census, 2011, Romaní-española: European Commission. Spain-census, 2011, Romaní-ucraniana: State Statistics Committee of Ukraine-census, 2001.)

	Romaní-búlgara	Romaní-griega	Romaní-húngara	Romaní-rumana	Romaní-eslovaca	Romaní-española	Romaní-ucraniana
Nº Habitantes	325343	265000	315583	621573	105738	725000	47587
% Población total	4,9	2,47	3,18	3,3	2	1,57	0,1
%Mestizaje ADNmt	56,3	10,7	63,6	37	73,5	29,9	61,5
%Mestizaje Crom Y	30,2	0	34,7	9,8	0	28,2	55,9

