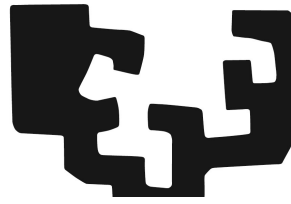


eman ta zabal zazi



# Haplotipos *MAPT* en *Homo sapiens*

Ibone Espinosa Carcabilla

2019

# Haplotipos *MAPT* en *Homo sapiens*

Memoria de Tesis Doctoral dirigida por:

Dr. José Ángel Peña García

Dr. Miguel Ángel Alfonso Sánchez

para la obtención del grado de 'Doctora en Biología' de:

Ibone Espinosa Carcabilla

Departamento de Genética, Antropología Física y  
Fisiología Animal

Leioa, 2019

Este proyecto ha sido posible gracias al Gobierno Vasco, por la concesión de una beca de investigación predoctoral y por la financiación de los proyectos "Variabilidad Genética en la Población del País Vasco" (subvención para actividades de grupos de investigación consolidados IT-833-13), "Origen de los haplotipos MAPT de *Homo sapiens*" (Programa SAIOTEK S-PE10UN54) y "Biología de las poblaciones humanas de los Pirineos" (Comunidad de Trabajo de los Pirineos CTP08-R08).





## AGRADECIMIENTOS

A mis directores de Tesis Doctoral, el Dr. **Jose Ángel** Peña García y el Dr. **Miguel Ángel** Alfonso Sánchez, por aceptarme primero como alumna interna del laboratorio y después como doctoranda, por la disponibilidad que han mostrado y porque sus críticas han sido siempre constructivas, porque nunca les ha faltado paciencia y buen humor, porque siempre han estado dispuestos a compartir variopintos chascarrillos y pintorescas anécdotas.

Al Dr. **Luis** Gómez Pérez - ¡qué ilusión poder escribir Dr. Luis Gómez Pérez! – porque en ninguno de los universos paralelos de Marvel existe un compañero de laboratorio mejor que tú (ni un lugar más frío que la fuente del hielo polar al fondo del pasillo).

A las Dras. **Susana** García Obregón y **Ana** María Pérez Miranda porque siempre han tenido una palabra de ánimo o una buena recomendación.

A l@s alumn@s intern@s y estudiantes de máster que han pasado por el laboratorio y muy especialmente a **Sainza**, que se acordó de mí al ver una oferta laboral bastante curiosa gracias a la cual se abrió una puerta cuando se cerraba otra, y a **Ainhoa**.

A los miembros del grupo BIOMICs, la Dra. **Marian** Martínez de Pancorbo, Ainara García y Amara Muro, por facilitarme el uso de las instalaciones del Instituto Lucio Escaray de Vitoria.

Como soy afortunada en extremo, me resulta absolutamente imposible mencionar a todas las personas que en mi vida personal me han apoyado, animado y ayudado a seguir adelante con un proyecto que en algunos momentos pareció inabarcable e interminable pero no puedo dejar de nombrar...

A mis padres, **Suso** y **Begoña**, y en general a mi familia. *A los que están, a los que no y a los que tienen que llegar.*

A mis amig@s. A los más antiguos, **Tati** y **Pablo**. A las doctoras con vara, **Cris** y **Mari**.

A **Gustavo**. Tú sabes que tus tupperts me han mantenido cuerda y con vida ;)

Y en general, a todos lo que habéis dedicado un momento de vuestro tiempo a ocuparos de esta tesis o de esta doctoranda: muchísimas gracias, **eskerrik asko!**



# ÍNDICE

---



## INTRODUCCIÓN

Región genómica estudiada: locus 17q21.31.

Antecedentes.

Región 17q21.31.

Estructura.

Fenotipos asociados.

Distribución geográfica.

Historia evolutiva.

Herencia de otras especies del género *Homo*.

Selección natural como principal moldeadora.

Historia demográfica de las poblaciones como principal moldeadora.

Marcadores analizados.

Indel

STRPs

SNPs

Poblaciones estudiadas

Europa. Origen del patrimonio genético europeo.

La península Ibérica.

El Norte de la Península.

Los vascos: lengua, genética y arqueología.

Vascos de Gipuzkoa.

Vascos de Navarra.

Residentes del Gran Bilbao.

El Mediterráneo.

Valencia.

Los nómadas que llegaron de Asia. El pueblo *Rrom*.

Orígenes asiáticos.

*Out of India* y expansión.

El pueblo *Rrom* en la actualidad. Perfil genético.

Gitanos Vascos.

El Continente Americano. Procesos de colonización.

Afrodscendientes del Chocó.

Una etnia amazónica: los Waorani.

Jujuy.

HIPÓTESIS DE TRABAJO

OBJETIVOS

MATERIAL Y MÉTODOS

Muestras analizadas:

Vascos de Gipuzkoa.

Vascos de Navarra.

Residentes vascos.

Gitanos residentes en el País Vasco.

Afrocolombianos del Chocó.

Jujeños de Argentina.

Etnia Waorani de Ecuador.

Marcadores y análisis molecular.

Indel, PCR y electroforesis en gel de agarosa.

Microsatélites, PCR y electroforesis en capilar.

SNPs y High Resolution Melt Analysis.

Análisis estadístico de los datos.

Frecuencias alélicas.

Frecuencias haplotípicas.

Error estándar de la media.

Análisis del equilibrio Hardy-Weinberg.

Corrección de Bonferroni.

Diversidad génica y Contenido de Información Polimórfica.

Test de Desequilibrio de Ligamiento.

AMOVA.

Clínicas.

Mapas sintéticos.

Autocorrelación espacial, índice de Moran y análisis del modelo de Aislamiento por la Distancia.

Coeficientes de Distancias Genéticas.  $F_{ST}$  de Reynolds y  $R$  de Harpending y Jenkins.

Test de Mantel.

Análisis de escalamiento multidimensional, MDS.

Método del centroide.

Diagrama de Shepard.

Análisis de bases de datos de ADN antiguo.

Origen de los datos.

Preparación de los datos.

Análisis estadísticos.

AMOVA.

Correlación parcial.

Mapa sintético.

## RESULTADOS

### Sección 1. Frecuencias alélicas y haplotípicas

1.1. Marcadores bialélicos

1.2. Marcadores STRPs

1.3. Test de equilibrio HW

1.4. Desequilibrio de ligamiento

1.5. Haplotipos extendidos

1.6. Test de neutralidad

1.7. AMOVA

### Sección 2. Variabilidad geográfica de los haplotipos SNP

2.1. Europa

2.2. África

3.3. Asia-Pacífico

2.4. América

### Sección 3. Distribución geográfica de los haplotipos SNP/STRP.

### Sección 4. Perspectiva histórica.

4.1. Características de las muestras analizadas

4.2. AMOVA.

4.3. Correlaciones parciales.

4.4. Variabilidad geográfica y temporal. Mapas sintéticos.

## DISCUSIÓN

## CONCLUSIONES

## REFERENCIAS

## MATERIAL ANEXO









# INTRODUCCIÓN

---

Para esta memoria, se ha dividido la introducción en dos grandes bloques, correspondientes a la región genómica estudiada y a las poblaciones analizadas.

En el primer bloque, se ha realizado un repaso de los antecedentes más relevantes en la investigación de la región 17q21.31 y posteriormente se han tratado su estructura, los fenotipos asociados a las distintas configuraciones y su distribución geográfica, así como las hipótesis concernientes a su historia evolutiva, siendo estos dos últimos apartados de especial importancia para la discusión de los resultados y para la obtención de conclusiones en este proyecto. Finalmente se describen los marcadores analizados durante el trabajo de laboratorio.

El segundo bloque se ha dividido a su vez en tres grandes subapartados, correspondientes a la región geográfica de la que forman parte las muestras: Europa, Asia y América. Se ha procurado mencionar los eventos más influyentes en la conformación del acervo genético de los grupos poblacionales estudiados, así como sus características más notables.

“Comprender la arquitectura y la distribución de estos haplotipos es importante, tanto para entender la genética de poblaciones y la historia como para desarrollar una comprensión de la patogénesis de las enfermedades asociadas (...) La distribución de los haplotipos MAPT es de interés por tres razones. La primera, eleva la cuestión del origen del haplotipo H2, ya que parece ser tan divergente de su otra configuración, el H1, que no puede haber recombinación entre ellos en una región de casi 2Mb; segundo, sugiere que el haplotipo MAPT puede ser empleado como marcador de ascendencia europea de una población; y tercero, deja abierta la posibilidad de que las poblaciones con frecuencias mayores de H1 tengan una mayor incidencia de las enfermedades asociadas, especialmente PSP y CBD”  
Evans et al, 2004

## 17q21.31, la región genómica estudiada.

### ANTECEDENTES

#### 1975 – 1999: Del descubrimiento de la proteína $\tau$ al descubrimiento de los haplotipos MAPT.

En 1975 Murray D Weingarten y colaboradores publicaron el hallazgo y aislamiento de “una proteína estable al calor esencial para el ensamblaje de microtúbulos” (Weingarten *et al*, 1975) a la que bautizaron con el nombre de tau ( $\tau$ ) “debido a su habilidad para inducir la formación de microtúbulos”. Era la primera proteína necesaria para la polimerización de los microtúbulos que se conocía después del descubrimiento de la tubulina. El artículo, titulado “*A protein factor essential for microtubule assembly*”, finalizaba alentando a la comunidad científica a examinar a fondo la función de tau en la célula.

No fue hasta casi diez años después cuando Chapel y colaboradores (1984), mientras investigaban las inmunoglobulinas y otras proteínas en el fluido cerebroespinal de pacientes con Alzheimer, encontraron que la proteína tau se hallaba más frecuentemente en aquellas personas que padecían la enfermedad que en controles y sugirieron que debía ampliarse la búsqueda a otros marcadores genéticos “no necesariamente confinados al cromosoma 6”, donde se ubican los genes codificantes de inmunoglobulinas pero no de tau. Durante los años siguientes se relacionaron varias alteraciones en la proteína tau - fosforilación aberrante (Grundke-Iqbal *et al* 1986), expresión intensificada (Barton *et al* 1990), disfunción de los microtúbulos (Matsuyama *et al*, 1989) - con el Alzheimer (Ashman *et al*, 1992).

Las investigaciones en este campo continuaron y durante la década de los noventa del siglo pasado el estudio de varias enfermedades neurodegenerativas, tanto esporádicas como familiares, llevó al descubrimiento de que muchos de estos desórdenes tenían una característica común y distintiva: la formación de depósitos filamentosos intracelulares compuestos por la proteína tau asociada a microtúbulos (MAPT). De esta manera, un grupo heterogéneo de demencias y trastornos motores con un amplio espectro de manifestaciones fenotípicas pasó a conocerse como **tauopatías** (Lee *et al*, 2001). Este grupo incluye, además de la enfermedad de Alzheimer (AD, Alzheimer's Disease) ya mencionada, la Parálisis Supranuclear Progresiva (Progressive Supranuclear Palsy, PSP), la Degeneración CorticoBasal (CorticoBasal Degeneration, CBD), la enfermedad de Pick (Pick's disease, PiD) y la Demencia FrontoTemporal y Parkinson ligado al cromosoma 17 (FrontoTemporal Dementia and Parkinsonism linked to chromosome 17, FTDP-17), además de otras menos (re)conocidas. La identificación de tau como el componente mayoritario de los ovillos neurofibrilares posicionó el locus *MAPT* como gen candidato para estas enfermedades neurodegenerativas (Caffrey *et al*, 2007).

Con la asociación entre la PSP y un marcador dinucleotídico del intrón 9 de MAPT (Conrad et al, 1997) y con el descubrimiento de varias mutaciones en el gen que conllevan la formación de agregados anormales de la proteína tau en la forma familiar de la FTDP (Gambetti et al, 1980; Kang et al, 1987) quedó probado que la disfunción de la proteína tau era suficiente como causa de enfermedad (Foster et al 1997, Hutton et al 1998, Poorkaj et al, 1998, Spillantini et al 1998c). Unos años antes, en 1994, ya se había conectado el cromosoma 17, lugar donde está ubicado el gen *MAPT*, con una forma hereditaria autosómica dominante de la FTDP (Kosik et al, 1986). Otros autores encontraron también esta relación y la afección, con un espectro clínico que oscilaba entre los fenotipos de demencia a los de Parkinson, pasó a denominarse Demencia Frontotemporal y Parkinsonismo ligado al cromosoma 17 (FTDP-17).

Se formaron dos líneas generales de investigación en torno a MAPT. Una línea bioquímica centrada en los componentes moleculares de las lesiones neuropatológicas que provocaban las enfermedades y otra línea de ámbito genético, que estaba enfocada al estudio de las formas familiares de las enfermedades para identificar los genes implicados en ellas, que, en la mayoría de los casos, codificaban o aumentaban la expresión de los componentes moleculares descritos por los estudios bioquímicos (Goedert et al, 2005).

### **1 9 9 9 – Actualidad: La tercera vía.**

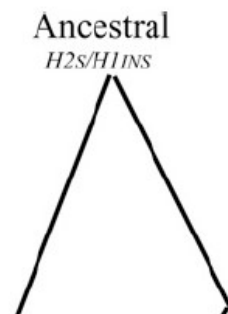
Cuando en 1999 Baker y colaboradores, analizando el comportamiento del gen *tau* en la Parálisis Supranuclear Progresiva (PSP) identificaron en 200 individuos caucásicos un completo desequilibrio de ligamiento de alrededor de 100kb de ADN y definieron los haplotipos MAPT\*H1 y MAPT\*H2 en función de un fragmento de 238 pares de bases que aparecía en H1 pero faltaba en H2, además de ampliar la diana de las investigaciones de manera que los análisis abarcaran la extensión total de los nuevos haplotipos, abrieron una tercera vía de acercamiento a la región MAPT: la genética de poblaciones.

Los primeros estudios en abordar de manera directa el origen de los haplotipos MAPT y su distribución geográfica datan de 2004. Evans y colaboradores (2004) emplearon el marcador rs72377086 (indel) para determinar las frecuencias de los haplotipos H1 y H2 en poblaciones de todo el mundo además de en algunos primates no humanos. Hallaron una asociación prácticamente completa entre el haplotipo H2 y las poblaciones de descendencia caucásica, resaltando su valor como marcador para la estima de mestizaje con grupos europeos. Debido a la constancia de la distribución del haplotipo H2 desde las islas Orcadas hasta Oriente Próximo, propusieron que el origen de este haplotipo debía ser antiguo y que podría haberse debido a un evento mutacional o a un aporte de una población humana anterior a través del mestizaje. Además, sugirieron una historia evolutiva compleja para el

locus MAPT ya que todos los primates no humanos estudiados habían mostrado la inserción de 238 pares de bases indicativa del haplotipo H1 mientras que en otros marcadores - como el de la saitoquina, analizado por Holzer y colaboradores (2004) - habían mostrado secuencias más similares a las del haplotipo H2.

La posibilidad de que el haplotipo H2 mostrara la frecuencia y la distribución que lo caracterizan debido a un evento de mestizaje con otro grupo poblacional fue desarrollada por Hardy y su equipo en un artículo titulado “*Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens*” (2005). En él argumentan que los haplotipos MAPT H1 y H2 no siguen una relación precursor - producto dado que en algunos puntos de la secuencia los chimpancés muestran similitudes con una de las configuraciones de los haplotipos mientras que en otros puntos las coincidencias son con la configuración contraria. Se basan principalmente en dos loci, el polimorfismo de 238 pares de bases descrito por Baker y colaboradores (1999), también conocido como indel rs72377086, y en el SNP A/G del gen de la saitoquina que provoca la sustitución del séptimo aminoácido de la proteína, transformando el aminoácido glutamina (Q) en arginina (R). Consideran que ambos haplotipos deben haber derivado independientemente de un precursor más lejano en el tiempo, una secuencia ancestral que presentaría los alelos R (característico del haplotipo H2) e inserción (que define el haplotipo H1), igual que la secuencia detectada en chimpancés. De ella surgirían los linajes del haplotipo H1, en el que habría mutado el gen de la saitoquina (STH7) y que sería el que presentarían los primeros *Homo sapiens*, y del haplotipo H2, en el que habría ocurrido la delección y sería propio de los *Homo neanderthalensis*.

indicates sthQ7, H2s indicates sthR7. The same diagrams c  
for the other polymorphisms in Table 1 (see also [17,18].



**Figura I.1.** Árbol de máxima parsimonia que muestra la relación entre los polimorfismos saitoquina (Q7R) e indel en el locus MAPT. En él se indica que las variantes H1 y H2 derivaron de un ancestro común. H1s indica sthQ7, H2s indica sthR7 (Hardy et al, 2005).

Según éste grupo, tanto la distribución geográfica del haplotipo H2, consistente con una intrusión en la población de *Homo sapiens* en torno al tiempo en el que se fundó la población europea (Evans et al, 2004), hace aproximadamente 30.000 años, como los resultados del análisis del *slippage*

de los marcadores microsatélite (Weber et al, 1993; Goldstein et al, 1995), que indican que el evento debió de ocurrido hace entre 6.675 y 21.700 años, señalarían la aparición del haplotipo H2 en *Homo sapiens* dentro de la ventana de tiempo en la que convivieron ambas especies de homínidos en Europa hace entre 45.000 y 18.000 años (Stringer et al, 2001).

Otra explicación a la llamativa distribución de frecuencias de los haplotipos MAPT llegó ese mismo año. El equipo de Stefansson (2005) descubrió en la región 17q21.31 un fragmento de 900 kb que aparecía invertido en el haplotipo H2 al compararlo con el H1. Empleando dos SNPs asociados a la inversión estimaron que la frecuencia del haplotipo invertido (H2) era del 20% en Europa, del 6% en África y de menos del 1% en el Este de Asia, en la línea de lo descrito previamente (Baker et al, 1999; Evans et al, 2004) y mediante el análisis de árboles de máxima parsimonia concluyeron que los haplotipos H1 y H2 habrían divergido hace aproximadamente 3 millones de años, antes incluso de la aparición de la primera especie considerada dentro del linaje *Homo*. Bajo la hipótesis de que las altas frecuencias del haplotipo H2 en Europa en comparación con otras poblaciones del mundo se debían a presiones selectivas, evaluaron el impacto de la selección natural en la distribución actual de las frecuencias de estos haplotipos empleando índices de reproducción en Islandia. Consideraron que los datos obtenidos mostraban que aquellas mujeres que portaban al menos una copia del haplotipo invertido (H2) tenían más hijos (a pesar de que la diferencia era de menos de un niño por mujer) y que el modelo más adecuado para explicar la distribución de frecuencias de los haplotipos MAPT sería el de la selección balanceadora.

Uno de los estudios más completos de esta región lo realizaron Donnelly y colaboradores (2010). Ampliaron los datos existentes hasta el momento de la distribución global de los haplotipos MAPT\*H1 y MAPT\*H2 en seres humanos empleando 21 SNPs como marcadores de la inversión en seres humanos; con esos mismos SNPs analizaron el genotipo de quince PNH, estimaron el rango de tiempo en el que pudo existir el antepasado común más reciente para el cromosoma H2 y su lugar de origen más probable usando haplotipos de STRPs, discutiendo así mismo el papel de la selección natural en la distribución geográfica actual de las frecuencias de los haplotipos MAPT.

En 2012 se publicaron dos estudios que proponían al menos ocho configuraciones estructurales alternativas para la región 17q21.31 basadas en la identificación y estudio exhaustivo de los polimorfismos en el número de copia de tres duplicaciones (Boettger et al, 2012; Steinberg et al, 2012); empleando varias técnicas y distintos marcadores bosquejaron la historia evolutiva de la región 17q21.31. Adicionalmente, Steinberg y su equipo realizaron una exploración más sistematizada de la diversidad de los haplotipos MAPT en África, ya que hasta el momento las muestras de este



continente se habían limitado a poblaciones del oeste y sur de África. Este enfoque más detallado reveló que los haplotipos invertidos (clado H2), aunque prácticamente ausentes en el oeste, aparecen mucho más frecuentemente que lo reportado hasta el momento en otros trabajos (Stefansson et al 2005; Donnelly et al, 2010). Además, ambos equipos expusieron que aún más profundas que las diferencias en la distribución geográfica de los haplotipos H1 y H2 es la estratificación hallada entre haplotipos duplicados y haplotipos sin duplicaciones, independientemente de su pertenencia a la familia H1 o a la familia H2, ya que encontraron que al menos el 50% de los europeos analizados presentaban duplicaciones mientras que en poblaciones de ascendencia no europea los haplotipos apenas presentaban duplicaciones.

Ese mismo año, y recogiendo el trabajo de secuenciación de los genomas de Neandertales y Denisovanos (Green et al, 2010; Meyer et al, 2012), Setó-Salvia y colaboradores (2012) desvelaron que las secuencias de DNA analizadas de ambas especies se correspondían con la estructura de los cromosomas H1, confirmando así los resultados de un estudio previo en Neandertales (Green et al, 2010). Cabe destacar que los datos genéticos disponibles de Denisovanos y *Homo neanderthalensis* son escasos, por lo que, como indicaban en el propio artículo, la existencia de cromosomas H2 en las poblaciones de ambos grupos no podían descartarse.

Uno de los últimos artículos en tratar la historia evolutiva de la región 17q21.31 (Alves et al, 2015) se ha centrado en el estudio de los haplotipos invertidos (H2' y H2D), analizando su distribución geográfica, diversidad, diferenciación de poblaciones, las relaciones filogenéticas entre los haplotipos y una posible desviación de la neutralidad de sus frecuencias. Este trabajo ha aportado datos sobre poblaciones previamente no analizadas y ha mostrado que la familia H2 es genéticamente más diversa y diferenciada de lo que se había observado hasta el momento.

En palabras de Steinberg (2012), el locus 17q21.31 es una de las regiones más dinámicas y complejas del genoma humano. Tal y como se ha expuesto en el apartado anterior, tras más de quince años de estudio se han descubierto un gran número de polimorfismos en este fragmento del genoma que permiten diferenciar entre los dos linajes principales de haplotipos MAPT, el H1 y el H2, o entre los subhaplotipos de ambas familias. A continuación, se ofrecen los detalles estructurales de la región 17q21.31, los fenotipos asociados a las configuraciones alternativas y las principales hipótesis sobre su origen, evolución y aparición en el *Homo sapiens*.

## REGIÓN 17Q21.31

El entorno genómico de MAPT es una región estructuralmente compleja. Una de sus principales características es una zona de 1.600 kilobases de

desequilibrio de ligamiento en la que no existe recombinación y que ha favorecido la aparición de dos linajes de haplotipos, el H1 y el H2 (Baker et al, 1999; Stefansson et al, 2005). Estas configuraciones, además de presentar particularidades “arquitectónicas” como una inversión de casi una megabase de longitud (Stefansson et al, 2005) o un sinnúmero de variaciones en el número de copias (CNV) (Steinberg et al, 2012; Boettger et al, 2012), alberga varios genes y marcadores, algunos de ellos asociados con la fertilidad femenina (Stefansson et al, 2005), la recombinación meiótica femenina (Stefansson et al, 2005; Chowdhury et al, 2009; Fledel-Alon et al, 2011) y las causas de enfermedades neurológicas. Adicionalmente, se ha propuesto que el haplotipo H2 está bajo selección positiva en europeos a pesar de su relación con el síndrome de la microdeleción 17q21.31.

### Estructura.

La región 17q21.31 se encuentran en el brazo largo del cromosoma 17 del ser humano, en el tercio más cercano al centrómero. El desequilibrio de ligamiento que afecta a casi 2 Mb (Pittman et al, 2004) es una de las características más destacables de esta zona; la falta de recombinación que ha originado este desequilibrio de ligamiento, y por tanto la aparición de los haplotipos MAPT\*H1 y MAPT\*H2, es causado principalmente por una inversión de 0,9 Mb (Stefansson et al, 2005), que parece estar flanqueada por dos bloques de duplicaciones de 203 kb (proximal) y 484 kb (distal). En humanos, los puntos de ruptura de la misma se corresponden con duplicaciones de la familia de genes LRRC37 (*leucine-rich repeat-containing 37A*), distribuidos por el cromosoma 17 y una de las pocas duplicaciones comunes en macacos, chimpancés y humanos (Zody et al, 2008).

La inversión se ha empleado como marcador de los haplotipos H1, por consenso estado directo, y H2, estado invertido, y contiene a su vez otros polimorfismos, además de varios genes involucrados en diversas funciones, entre ellos el que ha dado nombre a los haplotipos, el gen de la proteína tau asociada a microtúbulos, MAPT.



Figura I.2: Ubicación de la región 17q21.31 en el cromosoma 17 (Pennisi et al, 2008)

El gen MAPT consta de 16 exones (Andreadis et al, 1992) y se expresa principalmente en las neuronas del sistema nervioso central y periférico donde

su papel está relacionado con la formación y estabilización de microtúbulos, con la polaridad neuronal y con la transducción de señales (Shahani y Brandt, 2002). El primer marcador diagnóstico de los haplotipos H1 y H2 descubierto, el indel rs72377086, se ubica entre los exones 9 y 10 de este gen, en el intrón 9. Se trata de un pequeño fragmento de 238 pares de bases presente en el haplotipo MAPT\*H1 pero ausente en MAPT\*H2 (Baker et al, 1999).

En este mismo intrón se ha descrito además otro gen, el de la saitoquina (STH) (Conrad et al, 2002), gen dentro de gen, una situación muy poco habitual en el genoma humano. El gen STH carece de intrones y codifica para una proteína de 128 aminoácidos con un patrón de expresión similar al de la proteína tau, aunque con función desconocida hasta el momento (Conrad et al, 2002) y sin ninguna homología evidente con otras proteínas o motivos proteicos (Gao et al, 2005). Aparece tanto en seres humanos como en los primates más cercanos a *Homo sapiens* (chimpancés, bonobos y gorilas) pero no en otros primates ni en roedores (Conrad et al, 2004; Holzer et al, 2004; Gao et al, 2005).

Curiosamente, en este gen dentro de gen se ha descrito un marcador para los haplotipos MAPT, el polimorfismo Q7R; una sustitución de un nucleótido de adenina por otro de guanina (A → G) que provoca el cambio de glutamina (Q) a arginina (R) en el séptimo aminoácido de la proteína. El alelo Q aparece en el contexto H1 mientras que el R en el del haplotipo H2.

Otro de los genes presentes en esta región es el KANSL1, anteriormente KIAA1267, que codifica una proteína nuclear implicada en la modificación de la cromatina (Smith et al, 2005). La relevancia de este gen en la estructura de los haplotipos MAPT radica en que en ambos linajes, H1 y H2, han ocurrido duplicaciones parciales recurrentes que engloban la región promotora y el primer exón de KANSL1. En el contexto del haplotipo H1 aparece una duplicación de 205.000 pares de bases y en el contexto del haplotipo H2 surge una duplicación de menor tamaño, 155.000 pares de bases, a las que en la literatura se refieren como CNP205 y CNP155, respectivamente (Steinberg et al, 2012) o como  $\beta$  y  $\alpha$  (Boetger et al, 2012).

Una tercera duplicación abarca prácticamente la totalidad del gen NSF, aguas arriba de KANSL1, cuyo producto es una ATPasa esencial en la mayor parte del tráfico entre membranas en la célula (Whiteheart et al, 2001). Esta duplicación tiene un tamaño de unas 210.000 pares de bases, aparece en ambos haplotipos y se conoce como CNP210 (Steinberg et al, 2012) o como  $\gamma$  (Boetger et al, 2012).

Las tres duplicaciones descritas aparecen con distinto número de copias y conforman, al menos, ocho subhaplotipos diferentes, cinco correspondientes a la familia H1 y tres a la H2, con tamaños comprendidos entre las 1,08 Mb y 1,49 Mb (Steinberg et al, 2012; Boetger et al, 2012).

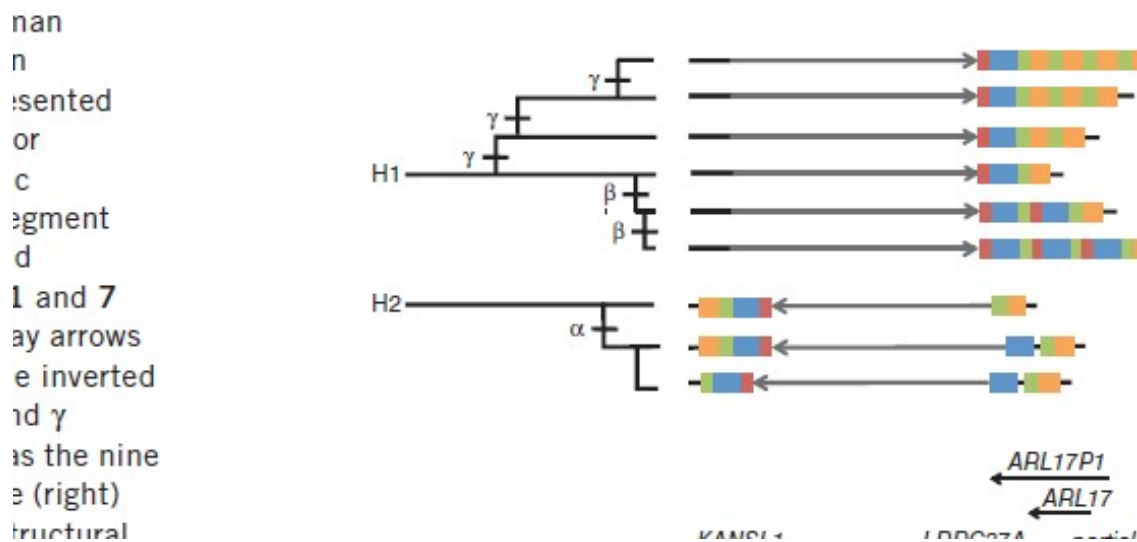
**Tabla I1.** Descripción de los subhaplotipos identificados hasta el momento. En la columna “Tamaño” se especifica el tamaño total del haplotipo y el contenido en duplicaciones entre paréntesis, tal y como se recoge en el trabajo de Steinberg y colaboradores (2012).

Boetger et al (2012)	Steinberg et al (2012)	Descripción	Tamaño
H1. $\beta$ 1. $\gamma$ 1	H1.1	Haplotipo directo (H1) y sin duplicaciones, con una sólo copia de CNP205/ $\beta$ (H1) y de CNP210/ $\gamma$	1.08 Mb (518kb)
H1. $\beta$ 1. $\gamma$ 2	H1.2	Como el anterior pero con una copia extra de CNP210/ $\gamma$ (dos copias, una duplicación)	1.29 Mb (728kb)
H1. $\beta$ 1. $\gamma$ 3	H1.3	Dos copias extra de CNP210/ $\gamma$ (tres copias, dos duplicaciones)	1.50 Mb (938kb)
H1. $\beta$ 2. $\gamma$ 1	H1D	Haplotipo directo con una duplicación (dos copias) de CNP205 / $\beta$	1.29 Mb (826kb)
H1. $\beta$ 3. $\gamma$ 1	H1D.3	Haplotipo directo con tres copias de CNP205 / $\beta$	1.49 Mb (1031kb)
H2. $\alpha$ 1. $\gamma$ 1	H2.1	Haplotipo invertido (H2) y sin duplicaciones, con una sólo copia de CNP155/ $\alpha$ (H2) y de CNP210/ $\gamma$	1.19 Mb (630kb)
H2. $\alpha$ 1. $\gamma$ 2	H2.2	Como el anterior pero con una copia extra de CNP210/ $\gamma$ (dos copias, una duplicación)	1.41 Mb (845kb)
H2. $\alpha$ 2. $\gamma$ 2	H2D	Haplotipo invertido con una duplicación (dos copias) de CNP155/ $\alpha$	1.48 Mb (996kb)

En contraste con datos previos (Steinberg et al, 2012), Alves y colaboradores (2015) han descubierto niveles apreciables de diversidad haplotípica en ambos subtipos invertidos, siendo las más altas las halladas en H2 $\gamma$  y H2D en poblaciones africanas. No observaron niveles significativos de estratificación en las poblaciones al analizar el haplotipo H2D, pero los cromosomas H2sub-saharianos son sustancialmente distintos de los del resto de grupos. Las mayores tasas de diferenciación genética aparecieron entre los grupos sub-saharianos y las poblaciones del norte de África y de Oriente Próximo.



**Figura I.3:** Estructuras alternativas de los haplotipos de la región 17q21.31. De los ocho haplotipos cinco corresponden a H1 y tres a H2, variando en tamaños desde las 1,08 Mb a las 1,49 Mb. Entre paréntesis se indica la cantidad de material duplicado. (Steinberg et al, 2012)



**Figura I.4:** Estructuras alternativas de los haplotipos de la región 17q21.31 y su frecuencia en diferentes poblaciones. En el dibujo esquemático inferior se indica qué segmento cromosómico está representado en cada color. Las flechas grises indican la orientación de la inversión. (Boettger et al, 2012)

Hasta ahora se han descrito tan solo una serie de SNPs marcadores de la inversión - capaces de diferenciar entre los haplotipos principales H1 y H2 -

(Stefansson et al, 2005; Donnelly et al, 2010; Steinberg et al, 2012) o diagnóstico de las duplicaciones, que distinguen entre los subhaplotipos H2 sin duplicaciones (H2) de aquellos que las presentan (H2D) (Steinberg et al, 2012).

Otros de los polimorfismos de interés hallados dentro de los límites de la inversión son los microsatélites MAPT, que se hallan dispersos por toda la región. Algunos de los identificados dentro de los límites de la inversión se han empleado para estimar la fecha del antepasado común más reciente para los haplotipos (Donnelly et al, 2010).

A pesar de que los distintos haplotipos se agrupan claramente dentro de las familias H1 o H2, existe una región de 30 000 pares de bases en segmento 5' del locus CRHR1 que muestra una ausencia total de diversidad entre algunos haplotipos H1 y H2 (Steinberg et al, 2012). Los descubridores de esta particularidad consideran que probablemente represente un evento de doble recombinación entre H1 y H2, ya que se trata de un fragmento demasiado grande para haber sufrido una conversión génica. Al hallarse además en todas las grandes agrupaciones por continente de *HapMap*, proponen que ésta doble recombinación ocurriera antes de la dispersión de los humanos modernos fuera de África.

### **Fenotipos asociados a las distintas configuraciones.**

La familia de haplotipos H1 ha sido asociada con varias enfermedades neurodegenerativas en las que el denominador común es la acumulación en las células nerviosas de microfibrillas de proteína tau o de inclusiones citoplasmáticas de esta misma proteína cristalizada. Englobadas bajo el término de tauopatías algunas de las más estudiadas son la demencia frontotemporal (FTD), la enfermedad de Alzheimer (AD), la enfermedad de Parkinson (PD) o la parálisis supranuclear progresiva (PSP) (Ballatore et al, 2007).

En el sistema nervioso central de un adulto coexisten seis isoformas de tau generadas mediante el *splicing* alternativo de los exones 2, 3 y 10 del gen MAPT (Goedert et al, 1989; Andreadis et al, 1992; Caffrey et al, 2007). Los exones 2 y 3 codifican pequeños insertos amino-terminales que forman parte de la proyección acídica que probablemente interactúa con la membrana plasmática (Brandt et al, 1995) y que regula el espacio entre los microtúbulos (Chen et al, 1992). La interacción entre la proteína tau y los microtúbulos es mediada por los dominios correspondientes formados por repeticiones imperfectas de 31 o 32 aminoácidos en el extremo carboxilo terminal, codificados por los exones del 9 al 12 (Goedert et al 1988, Gustke et al, 1994; Trinczek et al 1995) El *splicing* alternativo del exon 10 genera proteínas con tres (tau 3R) o cuatro (tau 4R) dominios de unión a microtúbulos. La composición de los agregados de proteína que caracterizan las tauopatías difiere en función de las mismas; por ejemplo, la enfermedad de Alzheimer

presenta ambas isoformas mientras que en la enfermedad de Pick aparece sólo la 3R y en la PSP, CBD, AGD y FTDP-17 el componente mayoritario es 4R.

Parece que la región promotora de MAPT\*H1 es más eficiente provocando la transcripción que la secuencia promotora del haplotipo MAPT\*H2 (Kwok et al, 2004), pero esto no se ha relacionado aún con ninguna otra característica apreciable.

Los cromosomas con configuración H2, por otro lado, son susceptibles a deleciones y microduplicaciones *de novo* que provocan retraso en el crecimiento y problemas de aprendizaje (Koolen et al, 2006; Sharp et al, 2006; Shaw-Smith et al, 2006). Dado que el subhaplotipo H2D es el único con duplicaciones homólogas en orientación directa flanqueando la región crítica para el desarrollo de esta enfermedad, sólo los portadores de H2D están predispuestos a la microdelección 17q21.31 mediante la recombinación homóloga no alélica (Steinberg et al, 2012).

Adicionalmente, se ha observado que los cromosomas H2D presentan mayores ratios de recombinación en mujeres (Fledel-Alon et al, 2011), lo que lleva a un mayor número de descendencia en estos individuos (Kong et al, 2004; Coop et al, 2008). Estas observaciones, descritas también por Stefansson y colaboradores (2005), llevaron a este grupo a proponer que los haplotipos H2D podían estar siendo seleccionados de manera positiva en Islandia.

### **Distribución geográfica.**

Los primeros estudios realizados en cuanto a la distribución de frecuencias de los haplotipos MAPT apuntaban a que la familia H2 era prácticamente exclusiva de Europa y Próximo Oriente - donde aparecía en una proporción máxima del 25% aproximadamente, con la excepción de Finlandia, más cercana a la media de Asia central (5%) - ya que en el resto de regiones analizadas - África, Este de Asia y nativos americanos - su frecuencia era virtualmente de cero (Evans et al, 2004) o aparecía en frecuencias mucho menores (África: 6%, este de Asia: < 1%. Stefansson et al, 2005).

Posteriormente, estudios más completos - con mayor número de individuos y una mayor representación de poblaciones por región - han desvelado frecuencias del haplotipo H2 de hasta casi el 40% en Cerdeña o entre los judíos Askenazi, acotando la zona con mayor presencia de este tipo de cromosomas a la región Mediterránea del suroeste de Asia y Europa (Donnelly et al, 2010) y reconociendo su presencia en algunas regiones de África, especialmente en el Norte y en el Este (Mozabites y Bejas ~13%) de donde se creía totalmente ausente (Donnelly et al, 2010; Steinberg et al, 2012).



Por tanto, el haplotipo mayoritario en todo el mundo es el H1 y el H2 aparece fundamentalmente en Eurasia, siempre con unas frecuencias inferiores a H1, donde además también son casi exclusivos los haplotipos con duplicaciones, independientemente de si se trata de los haplotipos de orientación “estándar” o invertida. Su versión no duplicada, por el contrario, presenta las frecuencias más bajas de todos los subhaplotipos y es principalmente africana, en concreto de las poblaciones cazadoras-recolectoras (Boettger et al, 2012; Steinberg et al, 2012); de hecho, las poblaciones africanas sub-saharianas, donde ambos subtipos del haplotipo invertido aparecen con frecuencias similares, son la excepción, ya que la mayoría de los grupos geográficos muestran un enriquecimiento en la configuración duplicada (H2D) con respecto a la invertida (H2) (Steinberg et al, 2012; Alves et al, 2015).

### Historia evolutiva.

*“The evolutionary history of the 17q21.31 inversion polymorphism is puzzling”*

*Stefansson et al, 2005*

La llamativa distribución geográfica de frecuencias de los haplotipos MAPT fue lo que propició en un primer momento que los estudios de la región 17q21.31 abordaran la problemática del origen e historia evolutiva de este locus.

“(…) Estos polimorfismos, tomados de manera conjunta, definen dos haplotipos extendidos que contienen al menos todo el gen tau (exones 1 – 13; ~100kb) y que se han mantenido ininterrumpidos por recombinación en todos los individuos de este estudio. Estos datos sugieren claramente que ambos haplotipos se establecieron temprano en la historia de la población caucásica (...) El hecho de que no exista recombinación en toda la longitud del gen tau sugiere que debe de existir una supresión de la recombinación en esta región o que la selección actúa en contra de los recombinantes” Baker et al, 1999.

Desde el momento del descubrimiento de los haplotipos MAPT, la elevada antigüedad de su origen es algo en lo que coinciden todos los investigadores, entre ellos Evans y colaboradores, uno de los primeros grupos en analizar la región 17q21.31 por su interés en genética de poblaciones. Para ellos la constancia en la distribución del haplotipo H2 desde las Islas Orcadas hasta Oriente Medio es una prueba de que H2 se originó en torno a la época de la colonización de Europa (Evans et al, 2004). Del mismo modo, plantean que el motivo de la falta de recombinación entre ambos haplotipos principales, H1 y H2, sea una mutación que evitara la recombinación o el aislamiento genético del grupo en el que surgió el haplotipo H2.

“Es difícil imaginar el origen del haplotipo H2, ya se plantee como un evento mutacional o como resultado del mestizaje con una población humana anterior: la divergencia de los dos haplotipos es extensa y sugiere una considerable separación



genética. Esto podría reflejar la existencia de una mutación que evitara la recombinación, manteniendo por tanto la integridad de cada haplotipo, o un antiguo aislamiento genético del grupo de humanos en el que surgió este haplotipo” Evans et al, 2004.

El momento de la separación de los linajes H1 y H2 ha sido fechado en hace 3 millones de años (Stefansson et al, 2005) y en hace 2 ´ 3 millones de años (Zody et al, 2008; Steinberg et al, 2012). Desde el descubrimiento de estos se ha discutido cuál de los dos haplotipos es el original, el H1 o el H2. Las últimas investigaciones apuntan a la familia H2 como el linaje ancestral, en concreto a un haplotipo muy similar al H2actual (Zody et al, 2008; Steinberg et al, 2012; Alves et al, 2015). Otros autores proponen que el haplotipo H2 sería el original en primates no humanos, pero no en humanos, en quienes la orientación original habría sido la no invertida (H1), proponiendo una historia de inversiones recurrentes de esta región en primates (Donnelly et al, 2010; Setó-Salvia et al, 2012).

Los trabajos que han analizado la composición alélica del locus MAPT en primates no humanos han arrojado resultados que en principio parecen contradictorios. Por una parte, al analizar el marcador indel rs72377086 del DNA en varios primates - incluyendo chimpancé, gorila, tití, monos nocturnos del género *Aotus*, macaco cangrejero (*Macaca fascicularis*) y mono verde africano (*Chlorocebus sabaeus*) -, todos los individuos resultaron homocigotos para el haplotipo H1 (Evans et al, 2004). Por otro lado, el estudio de varios sets de SNPs en primates no humanos determinó que todo el material genético analizado se correspondía con la configuración H2, como el trabajo de Holzer et al, (2004) con un set de 8 SNPs, incluyendo uno que determinaba la configuración del gen de la saitoquina, en tres chimpancés, un gorila, dos gibones, un macaco cangrejero y un mono verde africano y el trabajo de Donnelly et al, (2010), con 15 SNPs en gorila, gibones, chimpancé, bonobo y orangután. Estos últimos resultados concuerdan con los hallados durante el análisis del estado de la inversión en varios primates mediante la técnica de la hibridación fluorescente in situ o FISH (*Fluorescence in situ hybridization*) realizada por Zody y colaboradores (2008). El equipo encontró que este marcador es altamente polimórfico en ambas especies de chimpancé, mientras que la orientación invertida H2 es predominante en *Macaca* y en orangután. Propusieron que la región 17q21.31 es susceptible a inversiones recurrentes y que precisamente esto es lo que ha ocurrido en el linaje primate.

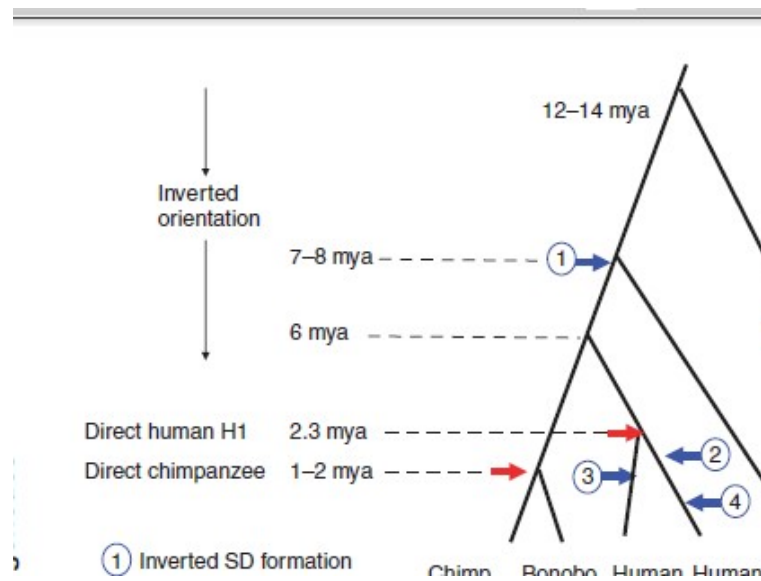


Figura I.5. (Zody et al, 2008)

“(…) La inversión de la región MAPT es, de manera similar a lo que encontramos en humanos, polimórfica en otras especies de grandes simios, y presentamos evidencias de que estas inversiones han ocurrido de forma independiente en chimpancés y en humanos. (…) Nuestros análisis apuntan a la configuración H2 como el estado ancestral más probable de esta región, con inversiones recurrentes durante la evolución primate (…) Aunque los haplotipos H1 y H2 son exclusivos de humanos, para simplificar, nos referiremos a ellos como orientación H1 y orientación H2 al describir la configuración en especies de primates no humanas. Las tres especies de macaco analizadas (*Macaca fascicularis*, *Macaca arctoides* y *Macaca mulatta*) y el orangután mostraron configuraciones FISH consistentes con la orientación H2, sugiriendo que esta orientación representa la configuración ancestral. Sorprendentemente, el examen de un único individuo de cada una de las dos especies de chimpancés estudiadas (*Pan paniscus* y *Pan troglodytes*) mostró que eran heterocigotos para la inversión. Al analizar una población con mayor número de individuos de chimpancés no emparentados encontramos que la inversión era altamente polimórfica. A diferencia de las poblaciones humanas, la configuración H2 representa el alelo mayoritario, con una frecuencia de 56%, en chimpancés. Todos los orangutanes de Sumatra resultaron homocigotos para H2 y el análisis de un único orangután de Borneo mostró que era heterocigoto, indicando que la inversión es probablemente polimórfica en todas las subespecies de este primate”

Zody et al, 2008.

Así pues, actualmente se considera que la configuración primitiva de los haplotipos MAPT era una más similar al haplotipo H2 y que la familia H1 se habría originado a partir de ella. Pero ¿cómo llegó la familia de haplotipos H1 en humanos a extenderse por todo el globo sustituyendo casi por completo la configuración original? ¿Por qué se encuentran altas frecuencias de H2 en torno a la región Mediterránea de Europa y Asia? ¿Ofrecen los haplotipos con duplicaciones algún tipo de ventaja local para ser predominantes y prácticamente exclusivos de Europa? ¿Cómo se explica que haplotipos relacionados con patologías graves como las tauopatías o el síndrome de la microdelección hayan alcanzado frecuencias tan altas?

De entre las hipótesis propuestas para explicar la distribución geográfica de frecuencias de haplotipos MAPT pueden destacarse las siguientes ideas (3):

- (1) El haplotipo H2 como herencia de otras especies de homínido. La posibilidad de que el haplotipo H2 fuera introducido por otra especie de homínido en el *pool* genético de *Homo sapiens* ha sido propuesto por varios autores (Evans et al, 2004; Hardy et al, 2005; Stefansson et al, 2005; Donnelly et al, 2010) pero la hipótesis más llamativa es la que aparece en el trabajo “*Evidence suggesting that Homo neanderthalensis contributed H2 MAPT to Homo sapiens*” (Hardy et al, 2005), fundamentalmente después de la confirmación de la hibridación de ambas especies (Prüfer et al, 2014; Sankararaman et al, 2014). Para Hardy y su equipo, la distribución del haplotipo y el análisis del *slippage* de ciertos marcadores sugieren que H2 entró a formar parte de la herencia genética de *Homo sapiens* en torno a hace 30.000 – 10.000 años pero la comparación de las secuencias de los haplotipos H1 y H2 con las de chimpancé determina que el fundador común de ambos haplotipos es mucho más antiguo. Proponen por tanto que el haplotipo H2 apareció en *Homo neanderthalensis* y entró en las poblaciones de *Homo sapiens* durante la coexistencia de ambas especies en Europa hace 45.000 – 18.000 años y que desde ese momento el haplotipo H2 ha sido favorecido por la selección debido al papel del haplotipo H1 en las tauopatías, tal y como se ha detallado anteriormente.
- (2) Selección natural como principal modeladora de la historia evolutiva de la región 17q21.31. En el mismo artículo en el que se describen por primera vez los haplotipos MAPT, se baraja la posibilidad de que la selección natural haya actuado sobre ellos debido a la relación diferencial que presentan las familias H1 y H2 con varias enfermedades neurodegenerativas.

“(…) Desde el punto de vista biológico hay tres posibilidades que podrían explicar la asociación entre tau y la PSP: (i) podrían existir diferencias cruciales entre ambos haplotipos en cuanto a la expresión de la proteína tau; (ii) podría haber diferencias en el *splicing* de ambas proteínas; o (iii) podría haber ocurrido una mutación patogénica pero no codificante en el contexto del haplotipo H1. El hecho de que los cerebros de pacientes con PSP presenten NFTs que consisten en las isoformas de Tau con cuatro repeticiones del motivo de unión a microtúbulos sugiere que podría existir una disrupción del *splicing* alternativo del exón 10 en aquellas zonas del cerebro lesionadas por la PSP. Por tanto, parecería que el mejor candidato para tener un papel en la patogénesis de la PSP es la variabilidad genética que afecta al *splicing* alternativo del exón 10”

Baker et al, 1999.

En el artículo exponen que, si la presencia o ausencia del pequeño fragmento de 238 pb afectara el proceso del *splicing* alternativo del exón 10 de MAPT, éste podría ser el factor de riesgo en cuestión y concluyen aconsejando ampliar las investigaciones al respecto.

Algunos autores sostienen que la selección natural actuó en contra del haplotipo H2 de manera temprana, lo que propició que el haplotipo H1 aumentara su frecuencia y se convirtiera en el alelo predominante en todas las poblaciones humanas (Zody et al, 2008), otros defienden que el haplotipo H2 fue favorecido por la selección natural en Europa, proceso que aún podría detectarse en la tasa de fertilidad de las mujeres islandesas (Stefansson et al, 2005).

Ambas propuestas podrían ser complementarias ya que la selección en contra de H2 de Zody y colaboradores (2008) habría ocurrido poco después de la separación de ambos haplotipos mientras que la selección positiva de H2 de Stefansson y colaboradores (2005) sería mucho más reciente y en un ambiente con presiones selectivas diferentes.

“(…) Nuestros resultados sugieren que la selección positiva es el determinante más probable de la inusual estructura haplotípica observada en poblaciones europeas en el locus 17q21.31”

Stefansson et al, 2005

Stefansson y colaboradores (2005) emplean datos de los ratios de reproducción en Islandia para sugerir que el linaje H2 está siendo seleccionado de manera positiva en europeos. Propusieron un modelo de selección balanceadora pero no fueron capaces de proponer un modelo dominante de selección (Donnelly et al, 2010). El equipo de Stefansson propuso que las altas frecuencias de la inversión en Europa se deben o bien a selección positiva o bien a selección balanceadora. Sus análisis se limitan a la población islandesa y están basados en el ratio de fertilidad de aquellas mujeres con al menos una copia del cromosoma H2 comparados con aquellas que no la tienen. Aunque la frecuencia de este cromosoma es alta en Islandia, es mucho más alta en otras regiones del sur de Europa y del suroccidente de Asia. Donnelly y colaboradores (2010) proponen estudiar estas regiones de una manera similar al trabajo realizado en Islandia para poder confirmar esta hipótesis.

Otros autores, como Boettger y colaboradores (2012), defienden que el rápido aumento de las frecuencias de los haplotipos con duplicaciones - H1D, H2D - en poblaciones humanas sugiere que las distintas duplicaciones del gen KANSL1 producen un fenotipo similar en ambos contextos (haplotipo directo H1, haplotipo invertido H2) que podría estar siendo seleccionado en humanos modernos, contradiciendo propuestas anteriores en las que la diana de la selección era la propia inversión (Stefansson et al, 2005; Zody et al, 2008; Donnelly et al, 2010). En esta misma dirección se orientan las reflexiones de Steinberg y colaboradores (2012): Si hay una ventaja selectiva sobre los

haplotipos H1D y H2D, una posibilidad es que las responsables sean las duplicaciones recurrentes que involucran tanto al promotor como al primer exón del gen KANSL1. Como se ha comentado anteriormente, KANSL1 ha sido asociado con el síndrome de la microdelección 17q21.31 (Sharp et al, 2005; Dubourg et al, 2011; Koolen et al, 2011) y se ha descubierto que se sobre expresa en el cerebro de pacientes con PD (Tobin et al, 2008), sugiriendo que una desregulación de la expresión de este gen podría tener consecuencias fenotípicas relevantes.

Steinberg y colaboradores (2012) realizan más aportaciones en torno a las posibles dianas de la selección natural en la región 17q21.31. Estudiaron 887 SNPs informativos entre H1 y H2; encontraron 9 sin sentido, 9 sinónimos y 22 mutaciones UTR en H2. Siete de las nueve mutaciones sin sentido están ubicadas en el gen IMP5, que codifica para la proteasa intermembrana 5, y cuatro de ellos alteran la estructura proteica según las predicciones de *PolyPhen*. Dos de los alelos H1 de estas cuatro mutaciones que alteran aminoácidos han sido previamente asociadas con la enfermedad de Párkinson (Simon-Sanchez et al, 2009) y por tanto los alelos sin sentido H2 podrían representar sustituciones protectoras bajo selección positiva. Curiosamente, los haplotipos H2 portan los alelos derivados para estos cuatro SNPs mientras que otros SNPs que no alteran la estructura de la proteína mantienen su estado ancestral al compararlo con chimpancés. El haplotipo H2 parece haber cambiado funcionalmente mientras que el H1 ha mantenido su estructura más similar a la de los chimpancés, aportando cierto apoyo a la hipótesis de selección sobre H2. No puede dejarse de lado el hecho de que los cambios no sinónimos son habituales durante la evolución. Las dos mutaciones sin sentido restantes están en el gen KANSL1, un gen que cuando está mutado resulta en un fenotipo similar al del síndrome de la microdelección 17q21.31 (Koolen et al, 2011).

Otra de las incógnitas en torno a la región MAPT es por qué los haplotipos H1D han alcanzado frecuencias tan altas en Europa. Dado que ciertos haplotipos H1 están asociados con desórdenes neurológicos tales como la enfermedad de Párkinson (Skipper et al, 2004; Tobin et al, 2008), la enfermedad de Alzheimer (Tuzun et al, 2005) y la parálisis supranuclear progresiva (Baker et al, 1999; Pittman et al, 2004; Hoglinger et al, 2011), Steinberg y colaboradores (2012) examinaron la asociación entre las duplicaciones y los haplotipos que predisponen a la enfermedad en varios individuos del *1000 Genomes Project*. Encontraron que la duplicación estaba menos presente que lo predicho por desequilibrio de ligamiento para el haplotipo H1c, que es un factor de riesgo reconocido para AD y PSP (Pittman et al, 2004; Myers et al, 2007). La duplicación es extremadamente rara en los haplotipos que suponen un factor de riesgo (Tobin et al, 2008; Pittman et al, 2004; Seto-Salvia et al, 2011), en cualquier caso, dado el tamaño de muestra y las bajas frecuencias de los haplotipos de riesgo para PD en la población, no pudieron aportar datos significativos. Consideraron que estas observaciones

sugieren un posible papel protector para las duplicaciones que debe de ser estudiado en profundidad. El haplotipo H2, que prácticamente siempre aparece con duplicaciones en las poblaciones europeas, parece ser protector contra muchas enfermedades (Tobin et al, 2008; Seto-Salvia et al, 2011), pero predispone a una microdelección asociada a discapacidad intelectual.

“Desde un punto de vista evolutivo, se cree que las variantes en el número de copia han surgido de manera independiente en las dos familias haplotípicas. Aunque se ha argumentado que el origen de las duplicaciones es anterior a las primeras expansiones del ser humano (el *Out of Africa*), los haplotipos enriquecidos en duplicaciones se encuentran a frecuencias mucho mayores en poblaciones actuales no africanas (Sudmant et al, 2010; Boettger et al, 2012; Steinberg et al, 2012) En consecuencia, han emergido nuevas hipótesis relativas a la historia evolutiva de la región, ubicando estas CNPs como candidatos potenciales a poseer algún tipo de ventaja selectiva. Esto es particularmente evidente en poblaciones europeas, considerando que el haplotipo asociado a la inversión que porta duplicaciones de la copia corta – H2D – está presenta a altas frecuencias (de hasta el 25%) y muestra niveles muy bajos de diversidad genética (Boettger et al, 2012; Steinberg et al, 2012)” Alves et al, 2015

Aunque en la mayoría de los estudios referentes a la región 17q21.31 se ha planteado que alguno de los *loci* presentes en ella pueda estar sujeto a selección, debido principalmente a la curiosa distribución de frecuencias de los haplotipos, no se ha obtenido ninguna prueba concluyente.

(3) Historia demográfica de las poblaciones. La identificación del proceso que condujo a la expansión de los haplotipos con duplicaciones en el ser humano se mantiene esquivo, ya que los patrones observados podrían también deberse al resultado de la compleja historia demográfica de las poblaciones sin necesidad de invocar a la selección natural (Zody et al, 2008; Donnelly et al, 2010; Steinberg et al, 2012; Alves et al, 2015). Procesos de migración, deriva genética (efecto fundador o cuellos de botella) podrían haber tenido un peso mayor que la selección natural en el proceso de modelado de las frecuencias actuales de los haplotipos.

En cualquier caso - y como en la mayoría de las ocasiones - estas ideas no son completamente excluyentes, por lo que diversos autores, a la hora de tratar de explicar la historia evolutiva de la región MAPT, han empleado una combinación de las mismas. A continuación, se presentan las principales alternativas.

Según Stefansson y colaboradores (2005) este tipo de cromosoma habría sido raro en los primeros grupos que abandonaron África hace más de 60.000 años y que dieron lugar a las poblaciones europeas y asiáticas, dado que los haplotipos H2 apenas aparecen en África ni en Asia. Las altas frecuencias encontradas en Europa junto con la extrema homogeneidad del



haplotipo H2 serían consistentes con una rápida expansión que partiera de unos pocos cromosomas fundadores y que fuera debida a la selección natural. Este episodio podría reflejar la exposición a algún tipo de presión ambiental local en Europa y Oriente Próximo.

Ante la cuestión de cómo justificar la presencia de dos linajes tan antiguos dentro del acervo genético humano ancestral, con una divergencia de secuencia indicativa de una edad de coalescencia de 3 millones de años - una fecha previa a la emergencia del *Homo sapiens* anatómicamente moderno en África (hace ~ 150.000 años) e incluso al origen del género *Homo* (hace ~ 2,5 millones de años) (Carroll SB, 2003) - el equipo de Stefansson (2005) propone dos hipótesis. Por un lado (1) que ambos linajes hayan sido mantenidos en el acervo genético humano ancestral desde el momento en el que ocurrió la inversión debido a algún tipo de selección balanceadora que previno que alguno de los dos linajes se perdiera por deriva genética (Bamshad et al, 2003; Andolfatto et al, 2001). Por otro, (2) que ambos linajes hubieran divergido en acervos genéticos aislados durante un largo periodo de tiempo y que el linaje H2 fuera introducido en el *pool* genético humano ancestral antes de la migración (o migraciones) de los humanos anatómicamente modernos fuera de África hace ~60.000 años. Este escenario implicaría mestizaje con alguna de las especies de homínidos contemporáneas como *Homo heidelbergensis* u *Homo erectus*. Dado que varios estudios aportan evidencia indirecta de que el *Homo sapiens* anatómicamente moderno probablemente se encontró con otras especies de homínidos en Asia (Brown et al, 2004; Reed et al, 2004), proponen que este tipo de encuentro también podría haberse dado en África, conduciendo a la introducción del haplotipo H2 en el acervo genético de *Homo sapiens*. De hecho, Hammer et al (2011) han encontrado evidencias de una introgresión de alrededor del 2% del genoma humano desde otra línea de humanos arcaicos en África, si bien la datan en tan sólo 35.000 años de antigüedad.

La hipótesis del mestizaje predice que existirán otras regiones, especialmente aquellas con bajos ratios de recombinación, con edades de coalescencia similares (Stefansson et al 2005).

La propuesta de Hardy y colaboradores (2005), como se ha comentado previamente, se basa en la mezcla de *Homo neanderthalensis* y *Homo sapiens* durante la época de convivencia hace entre 45.000 y 18.000 años. Los Neandertales habrían aportado el alelo H2 a las poblaciones de *Homo sapiens* de Eurasia, *donde* aumentó su frecuencia debido a la capacidad protectora de esta variante frente a ciertas enfermedades. No explican el origen primordial de los haplotipos MAPT, no detallan si consideran que la inversión se dio en Neandertales, antes de que éstos divergieran o antes de la salida de África de los grupos que dieron lugar a esta especie o si simplemente se perdió en los grupos africanos, pero se mantuvo en Neandertales que lo reintrodujeron en Europa por hibridación.

Zody y colaboradores (2008) sí plantean una hipótesis para el origen primero de los haplotipos MAPT. Tomando en cuenta los resultados de sus análisis y de trabajos anteriores, proponen que la orientación original para primates sería la invertida (haplotipos H2) y que a lo largo de la evolución de este orden habría cambiado de orientación generando haplotipos directos o H1 en numerosas ocasiones (en el linaje de los orangutanes, en el de *Homo sapiens* hace 2.3 millones de años y más recientemente en chimpancés, hace 1 o 2 millones de años).

Un alelo similar a los haplotipos H2 modernos habría sido el predominante entre las poblaciones ancestrales (*Homo heidelbergensis*) (Hardy et al, 2005) pero su frecuencia se habría reducido y prácticamente eliminado en las poblaciones ancestrales africanas de *Homo sapiens* dado que una de las consecuencias de esta inversión fue la creación de grandes bloques de duplicaciones en orientación directa que predispondrían microdelecciones y por tanto a la enfermedad asociada. Debido a una selección negativa sobre los haplotipos H2, el H1 habría aumentado su frecuencia y se habría convertido en el alelo predominante en las poblaciones humanas.

Durante la fundación de las poblaciones europeas tras el *out-of-Africa*, el alelo H2 habría resurgido como resultado de un barrido selectivo parcial o debido a un cuello de botella en la población fundadora, tal y como sugieren la falta de diversidad entre los haplotipos H2 existentes y su aparente antiguo origen, de hace entre dos y tres millones de años, como ya ha sido anteriormente postulado (Stefansson et al, 2005).

“(…) Aunque no podemos descartar que el locus 17q21.31 esté afectado por la selección, la clina que presentan las frecuencias de sus configuraciones alternativas puede ser fácilmente explicada por un evento demográfico ocurrido durante la expansión Neolítica desde el suroeste de Asia hacia Europa sin necesidad de invocar a la selección natural” Donnelly et al, 2010

Bajo el modelo neutral de Donnelly y colaboradores (2010), la inversión se habría originado inicialmente en el suroccidente asiático o en el este o centro de África. En este último supuesto, habría penetrado en Europa previo paso por el suroeste asiático. En los dos casos, el aumento en la frecuencia del haplotipo H2 habría ocurrido en el suroeste de Asia, debido a la selección natural o deriva genética, alcanzando frecuencias moderadamente altas antes de entrar en Europa, donde las frecuencias de H2 se habrían ido diluyendo con las migraciones hacia el norte.

La ausencia de la inversión en el este de Asia puede ser explicada por efectos fundadores que habrían eliminado alelos en múltiples *loci* en la expansión hacia el este de los humanos modernos desde el suroeste de Asia. Alternativamente, el origen en el suroccidente asiático podría haber ocurrido tras esta expansión (Donnelly et al, 2010).



En el supuesto del origen asiático de la inversión, la presencia del haplotipo H2 en África podría ser explicado de dos formas: (i) vía flujo genético hacia África o (ii) que los haplotipos H2 en África no presenten la inversión (Donnelly et al, 2010).

“La topología de los árboles y la falta de diversidad de los haplotipos H2 es indicativo de un cuello de botella recientes seguido de una expansión poblacional o un barrido selectivo” Steinberg et al, 2012

Steinberg y colaboradores (2012) se decantan por un origen africano de la inversión. Consideran que el haplotipo H2' es el ancestral en el género *Homo* y que surgió en el este o centro de África expandiéndose hacia el sur antes del origen de los humanos anatómicamente modernos. Hace unos 2.3 millones de años habría surgido el haplotipo H1', que habría reemplazado casi completamente al H2' en África.

Las duplicaciones en el contexto del haplotipo minoritario H2 habrían ocurrido hace en torno a 1.3 millones de años, aunque sólo habrían alcanzado las frecuencias actuales en poblaciones europeas hace relativamente poco tiempo y de manera muy rápida, dada su distribución y su escasa diversidad genética (Steinberg et al, 2012).

Las duplicaciones en el contexto del haplotipo H1, en cambio, serían mucho más recientes, de hace aproximadamente 250 000 años. Estos haplotipos no se observan en las poblaciones africanas o asiáticas, indicando que quizá se hayan perdido por deriva genética (Steinberg et al, 2012).

Actualmente, al realizar una búsqueda sobre MAPT en PubMed, los artículos orientados a la investigación biomédica siguen siendo mayoritarios, pero “la tercera vía” permitió vislumbrar el interés de esta región genómica para la antropología y la genética de poblaciones.

## **Marcadores genéticos empleados.**

De todos los polimorfismos descubiertos en la región 17q21.31 se han empleado en este estudio para caracterizar los haplotipos H1 y H2 el indel rs72377086 (Baker et al, 1999), los microsatélites de Donnelly (2010) y los SNPs de Steinberg (2012) que a continuación se describen con cierto detalle.

### **Indel rs72377086**

Al secuenciar los intrones adyacentes al exón 10 del gen MAPT, Baker y colaboradores (1999) detectaron, entre los nucleótidos -951 y -713 aguas arriba del exon 10, en el intrón 9, una delección de 238 pares de bases heredada como parte del haplotipo menos común, el H2.

Es un marcador bialélico fácil de analizar mediante técnicas relativamente sencillas, rápidas y poco costosas como PCR y electroforesis en gel de agarosa.

Es uno de los primeros marcadores diagnóstico para los haplotipos H1 y H2 en ser descubierto y uno de los más empleados en estos estudios. Los estudios consultados en los que se emplea este marcador se basan en las frecuencias de inserción / delección encontradas en las muestras analizadas.

Se ha considerado que la presencia o ausencia de este fragmento de 238 pares de bases pueda influencias el *splicing* alternativo del exon 10, afectando por tanto al riesgo de desarrollar PSP u otras tauopatías (Baker et al, 1999) pero no se han hallado evidencias hasta el momento.

Aunque se han realizado varios estudios sobre el estado ancestral de este polimorfismo y en general de los haplotipos H1 y H2, aún no se ha llegado a ningún consenso (ver apartado historia evolutiva de la región 17q21.31).

### Microsatélites.

Los microsatélites, o STR (*Short Tandem Repeat*), son fragmentos de ADN cuya estructura está basada en una secuencia nucleotídica o motivo que se repite en tándem un número determinado de veces.

En este estudio se han empleado los cuatro microsatélites ubicados dentro de los límites de la inversión diseñados por Donnelly y colaboradores (2010): MAPT07, MAPT08, MAPT09 y MAPT14.

Se trata de STRs cuyas unidades de repetición presentan dos pares de bases (microsatélites dinucleotídicos, MAPT07 y MAPT09) o cuatro pares de bases (microsatélites tetranucleotídicos, MAPT 08 y MAPT09) que se repiten entre quince y veintiséis veces (MAPT07), siete y trece (MAPT08), trece y veinticuatro (MAPT09), nueve y veintitrés (MAPT14), oscilando el tamaño total global entre las 99 pares de bases y las 223 pares de bases (Tabla Intro 3).

**Tabla I.2:** Características de los microsatélites ubicados en la inversión 17q21.31

STRP	Posición	Distancia al siguiente STRP (pb)	nº bases repetidas	nº de repeticiones	tamaño (pb)
MAPT07	41 328 075	7 979	2	15 - 26	201 - 223
MAPT08	41 320 096	16 495	4	7 - 13	187 - 211
MAPT09	41 303 601	139 891	2	13 24	107 - 129
MAPT14	41 163 710		4	9 - 23	99 - 155

## Polimorfismos de un sólo nucleótido o SNPs (Single Nucleotide Polymorphism)

Los SNPs o polimorfismos de un solo nucleótido son, como su propio nombre indica, una variación en la secuencia de ADN que afecta a una sólo base, sustituyéndola por otra. Como todo polimorfismo, debe darse al menos en un 1% de la población para ser considerado como tal y no como una mutación puntual. Constituyen hasta el 90% de las variaciones genómicas humanas y aparecen de media en una de cada 1300 bases.

En este estudio se han empleado dos grupos de SNPs; por un lado, cinco SNPs capaces de discriminar entre los haplotipos principales H1 y H2 y por tanto en principio equivalentes al marcador indel rs72377086 (Donnelly et al, 2010). Por otro, cinco SNPs que permiten diferenciar los haplotipos H1/H2del H2D (Steinberg et al, 2012).

Figure 1 | The consensus build (Build 34: July 2003) of the tau haplotype

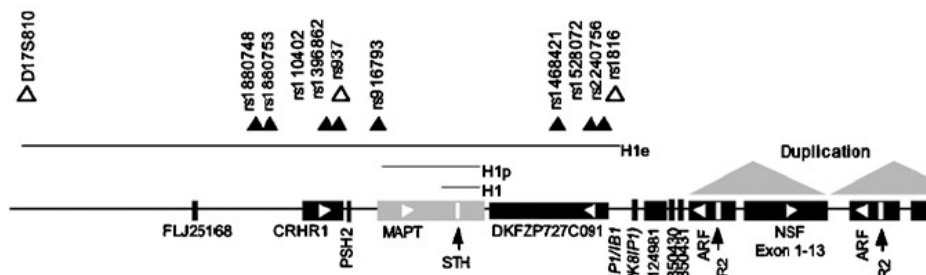


Figura I.6: Ubicación de los marcadores más relevantes (Hardy et al, 2005)

## Poblaciones estudiadas.

En el siguiente apartado se describen las poblaciones estudiadas y su contexto. Se ha procurado mencionar los eventos más importantes que pudieron influir en la configuración de la herencia genética de dichas poblaciones.

Las ocho poblaciones analizadas para esta tesis doctoral se han agrupado en tres conjuntos distintos en función de su historia común; por un lado, en el subapartado “Europa”, además de repasar dos de los fenómenos más relevantes para la conformación del acervo genético europeo –recolonización postglacial y difusión de la agricultura–, se discute la importancia de la orografía de la Península Ibérica, de la región norte y de la región mediterránea, y los contextos más concretos de las tres poblaciones vascas y de la población valenciana.

La muestra de población gitana vasca se ha tratado bajo el subapartado “El pueblo Rrom”, por tratarse de un grupo tradicionalmente nómada con unas características étnicas e historia distinta.

Por último, en el subapartado “América” se han descrito tres poblaciones muy diferentes: los afrodescendientes del Chocó, Colombia, los habitantes de la provincia argentina de Jujuy y la etnia Waorani de Ecuador, que se ha mantenido aislada hasta fecha reciente.

## EUROPA.

Europa se encuentra en tres de sus puntos cardinales limitada por extensas masas de agua; al norte el océano Ártico, al oeste el Atlántico y al sur el mar Mediterráneo, el cual, en ocasiones, ha supuesto más que una barrera un importante medio de comunicación e intercambio entre civilizaciones. Al este queda limitada por los montes Urales, el río Ural, el mar Caspio, la cordillera del Cáucaso, el mar Negro y los estrechos del Bósforo y de Dardanelos, que la separan de Asia.

Representa tan sólo el 6,8% de las tierras emergidas del planeta, abarca unos 10.180.000 kilómetros cuadrados de superficie y está habitada por algo más de 743 millones de habitantes, aproximadamente el 11% de la población total del planeta (UN DESA, 2017).

En diferentes disciplinas, como la Fisiogeografía, la Geología o la Antropología, se emplea el concepto de Eurasia como supercontinente dado que Europa y Asia forman en realidad una sola masa continental.

Los primeros humanos anatómicamente modernos (HAM) llegaron a Europa hace aproximadamente 45.000 años (Richard et al, 2000; Torroni et al, 2006) desde el sudoeste de Asia, pero, salvo por la excepción de algunas áreas meridionales del continente (denominadas refugios glaciales), éste quedó despoblado durante el máximo glacial ocurrido hace unos 25.000 años. Conforme remitía el frío y Europa volvía a ser habitable (hace ~15.000 años) se dio una nueva colonización paulatina de las tierras que habían estado cubiertas por el hielo (la llamada recolonización postglacial) y posteriormente, hace 8.000 años, con la difusión de la agricultura, dio comienzo la transición o revolución del Neolítico. Se considera que estos tres eventos son los responsables de la mayor parte de los patrones de diversidad genética hallados en Europa (Pinhasi et al, 2012) aunque existen otras fuentes de variación como la hibridación con las otras especies del género *Homo* que ya estaban presentes en Europa a la llegada de los HAM y que se presumen descendientes de alguna migración anterior fuera de África (Neandertales, Denisovanos y un tercer grupo que no está claramente identificado) (Green et

al, 2010; Reich et al, 2010; Reich et al, 2011) o como los movimientos poblacionales post-neolíticos (Hervella, 2010; Busby et al, 2015), en general con un efecto más destacable a nivel regional, por lo que se tratarán en los subapartados correspondientes. Se ha planteado también la posibilidad de una heterogeneidad genética interregional debida a una influencia diferencial de las migraciones de finales del Neolítico procedentes del oeste de Asia (Brandt et al, 2013).

### RECOLONIZACIÓN POST-GLACIAL.

Los cambios climáticos tales como las glaciaciones pueden modificar profundamente la distribución espacial y la diversidad genética de las especies (Hewitt, 2000; Arenas et al, 2012).

Se han registrado hasta cinco grandes periodos glaciales en la historia de la Tierra, siendo el último de ellos el que comenzó hace unos 100.000 años. Tuvo su apogeo durante el Último Máximo Glacial (UMG), hace entre unos 26.500 y 19.000 años. En ese momento el hielo cubría el norte de Europa, incluida gran parte de Alemania, Gran Bretaña e Irlanda (Harcourt, 2016). Aunque se trató de un evento global, las consecuencias para las diferentes poblaciones que habitaban el planeta en aquel momento fueron diferentes (Carolin et al, 2013). Las latitudes norte estaban parcialmente cubiertas de glaciares y se desarrollaron en ellas las características de los desiertos polares quedando en consecuencia totalmente despobladas (Jochim, 2011); los cinturones tropicales se tornaron excesivamente áridos, conformando desiertos y semi-desiertos, reduciendo la disponibilidad de terreno y modificando la distribución de la fauna; la población se concentró en las franjas entre ambos extremos, entre los fríos polares y la aridez radical, en las regiones denominadas **refugios glaciales** (Mirazón, 2016) siguiendo las migraciones de otros animales que a su vez buscaban las zonas donde hubiera agua dulce líquida disponible y vegetación de la que alimentarse.

La mayoría de las poblaciones colapsaron siguiendo patrones regionales propios; donde existe evidencia de esta época, las pruebas sugieren que todas las comunidades humanas se vieron afectadas, siendo un momento de pérdida significativa de la diversidad y de la citada modificación de la distribución espacial de las poblaciones (Hewitt, 2000; Arenas et al, 2012; Mirazón, 2016).

Los principales refugios reconocidos gracias a datos geológicos, genéticos y ecológicos, en particular de la distribución de especies animales y vegetales clave, son la región Franco-Cantábrica, la de los Balcanes y el Cáucaso y la del sur de Italia (Jochim M, 1987; Hewitt GM, 1996; Taberlet P et al, 1998; Gamble C et al, 2004) en Europa y algunas regiones del sur y sureste de Asia (Banks et al, 2008; Stewart y Stringer, 2012). También los patrones espaciales establecidos a partir de los restos materiales de ciertas culturas indican que éstas llegaron a colonizar regiones a 52º de latitud durante el Gravetiense (periodo cultural europeo entre hace 30.000 y 20.000 años, previo al último máximo glacial), que posteriormente se replegaron coincidiendo con el UMG y volvieron sobre sus pasos cuando mejoró el clima, con algunas regiones que muestran ocupación continua en las latitudes meridionales (Verpoorte A, 2009). Estas áreas habrían servido como un reservorio de la variación genética para determinados linajes, que se habrían expandido subsecuentemente por Europa y Asia, en su caso, en sincronía con la mejora de las condiciones climáticas, dejando trazas detectables en la distribución genética actual (Oppenheimer, 2006; De Fanti et al, 2015). La diferenciación genética entre las poblaciones de estos refugios se podría haber acentuado durante el periodo de aislamiento (Alves et al, 2015) favorecida también por la mencionada disminución del número de integrantes de las diferentes poblaciones y la consiguiente reducción de diversidad genética por efectos de la deriva (Barbujani et al., 1998; Torroni et al., 1998, 2001; Cardoso, 2008).

Varios autores han relacionado la difusión de linajes característicos del Paleolítico con la recolonización europea tras el UMG (Richards et al, 2000; Torroni et al, 2001; Pereira et al, 2005; Pala M et al, 2012).

## DIFUSIÓN DÉMICA DEL NEOLÍTICO.

Las transiciones del Neolítico ocurrieron en distintos momentos en distintas partes del mundo; la que transformó Europa comenzó en Oriente Próximo hace unos 10.000 años y se extendió por el continente en unos pocos milenios. El estilo de vida de los humanos viró de comunidades cazadoras-recolectoras a sociedades basadas en la domesticación de plantas y animales, lo que condujo al desarrollo de asentamientos permanentes, comercio y finalmente a la aparición de grandes civilizaciones. Esta transición cultural tuvo consecuencias considerables en la variabilidad genética humana al estimular el crecimiento del tamaño de las poblaciones y al desencadenar numerosas expansiones y mezclas, así como la adaptación a determinadas dietas o enfermedades.

El debate dedicado a entender el origen de la agricultura en Europa, particularmente centrado en descubrir si se extendió desde su lugar de origen mediante (1) la difusión démica (movimiento de agricultores e intercambio genético sustancial con los grupos cazadores-recolectores) o (2) por la aculturación de los cazadores-recolectores indígenas (movimiento de ideas sin intercambio genético) ha durado décadas y ha proporcionado numerosos artículos y estudios tratando de aportar luz a la discusión.

El primero de ellos fue portada en la revista *Science* en 1978; el trabajo de Cavalli-Sforza y colaboradores basado en el análisis de técnicas multivariantes para revelar clinas a partir de diez marcadores apoyaba el modelo de la difusión démica (Haber et al, 2016). Trabajos posteriores analizaron una gran variedad de loci, incluyendo ADNmt y marcadores del cromosoma Y. Algunos de los autores consideraron que sus resultados apoyaban la aculturación (Zvelebil M y Zvelebil KV, 1988; Currat M y Excoffier L, 2005), es decir, que la entrada de la agricultura y ganadería en Europa no habría supuesto un flujo de genes significativo sino solo de conocimientos y aprendizaje de nuevas técnicas entre grupos humanos vecinos. Otros investigadores defendieron que los datos obtenidos respaldaban la difusión démica (Ammerman AJ y Cavalli-Sforza LL, 1984; Dupanloup I et al., 2004), es



decir, que la expansión de la agricultura por Europa habría supuesto un importante flujo génico entre los agricultores-ganaderos de Oriente Próximo y los grupos de cazadores-recolectores locales. Un tercer grupo ha propuesto un modelo que aunaba ambas hipótesis (Balaesque et al, 2010).

Estudios recientes revelan sin embargo que el poblamiento de Europa ha sido mucho más complejo y que la transición Neolítica ni siquiera fue el evento que más influyó en el panorama genético actual (Haber et al, 2016).

## LA PENÍNSULA IBÉRICA.

“La primera parte de ella (Europa) es, como decíamos, el occidente; es decir, **Iberia**; ésta, en su mayor extensión, es poco habitable, pues casi toda se halla cubierta de montes, bosques y llanuras de suelo pobre y desigualmente regado” Estrabón\*, Geografía, libro III (s.l).

\*Nunca estuvo en la Península Ibérica.

La Península Ibérica, de unos 583.256 kilómetros cuadrados, está situada en el extremo sudoeste de Europa, separada del continente por la cordillera de los Pirineos. Su principal característica topográfica es que el centro de la península es una meseta, denominada Meseta Central, de unos 650 metros de altitud media que ocupa la mayor parte de la misma. Salvo el litoral bañado por el mar Mediterráneo, al sur y al este, la costa de la Península Ibérica es mayoritariamente rocosa, con abundancia de acantilados.

Su orografía ha determinado que en ocasiones el mar y los océanos que la rodean hayan actuado de barrera y en otras de vía de intercambios entre culturas; los Pirineos han sufrido también esta dicotomía, a veces barrera infranqueable, a veces exponiendo pequeños cañones y valles como camino de entrada para diferentes grupos indoeuropeos como los celtas o de otras culturas como los gitanos.

Gran parte de la responsabilidad de esa dicotomía han sido las transformaciones en el relieve de la Península que han provocado el clima y las condiciones hidrológicas a lo largo de la historia del planeta y muy especialmente durante el Pleistoceno / Paleolítico, que se han mostrado de gran importancia para la evolución humana y su dispersión (Drake et al, 2008; Hoffmann et al, 2016).

La principal característica climática de Pleistoceno / Paleolítico fue la sucesión de periodos glaciares con sus correspondientes periodos

interglaciares. Durante las fases frías se habrían formado glaciares en los Pirineos y parte de los sistemas montañosos del norte y la meseta habría presentado un clima especialmente frío, como atestigua la presencia en estos lugares de restos de animales como el mamut o el reno. Las fluctuaciones en la temperatura habrían provocado también cambios en el nivel del mar. Por su ubicación, además, habría estado afectada por los periodos húmedos africanos y en general los ríos habrían sido más caudalosos que en la actualidad, propiciando la aparición de terrazas fluviales cerca de donde se han encontrado gran número de restos de asentamientos humanos (Fullolla i Pericot JM y Nadal J, 2005).

En la Península Ibérica existen indicios de asentamientos de especies arcaicas del género *Homo* de hace más de un millón de años en la Sima del Elefante, ubicada en la Sierra de Atapuerca, Burgos (Bermúdez de Castro et al, 2011). Los yacimientos de esta zona, Patrimonio de la Humanidad desde el 30 de noviembre del 2000 y que llevan excavándose desde 1978, han dado algunas de las mayores sorpresas en cuanto a evolución humana se refiere en el entorno europeo: además de restos de uno de los pobladores más antiguos de Europa (el más antiguo de Europa Occidental), clasificado de momento de manera general como *Homo sp.*, el descubrimiento de una nueva especie de homínido que habitó la zona hace 850.000 años, el *Homo antecessor*; campamentos de *Homo heidelbergensis* con una antigüedad mínima de 500.000 años y del que se han podido recuperar restos de todas las partes esqueléticas, de *Homo neanderthalensis* y por supuesto de *Homo sapiens*.

Indicios relevantes de la ocupación del norte de la península se han encontrado también en Asturias, Cantabria y País Vasco, en las cuevas de Tito Bustillo, Altamira y Santimamiñe respectivamente.

Otra de las zonas de las Península Ibérica que ha sido poblada desde antiguo y con cierta continuidad es la costa del Mediterráneo, donde, además de las últimas evidencias halladas en Europa de presencia Neandertal, se encuentra la puerta de entrada del Neolítico en España.

Las sociedades neolíticas dieron lugar poco a poco y con el descubrimiento de los metales a las culturas y asentamientos de tribus autóctonas y procedentes de otros lugares como íberos, celtas, fenicios, griegos y cartagineses

La península formó parte del Imperio Romano, fue cuna de los Reinos Visigodos después, durante siete siglos estuvo conquistada por los musulmanes, cuyo reino, ubicado en el sur de la península se llamó Al-Andalus y mantuvo guerras con los reinos cristianos del norte durante la llamada Reconquista que culminó en el siglo XV con la toma de Granada.

Las muestras analizadas correspondientes a la Península Ibérica incluyen varias poblaciones con distintas características “étnicas” de la región nororiental y oriental. De la Comunidad Autónoma del País Vasco, vascos de Gipuzkoa, residentes vascos y gitanos afincados en el Gran Bilbao; de la Comunidad Foral de Navarra otra muestra de vascos autóctonos; y de la Comunidad de Valencia, una población de la propia provincia de Valencia.

## EL NORTE DE LA PENÍNSULA.

El norte de la Península Ibérica es una estrecha franja de tierra que discurre paralela al mar Cantábrico, en dirección este – oeste, y que queda limitada al sur por la Cordillera Cantábrica, que con unos cien kilómetros de anchura media y una continuidad de casi quinientos, supone una barrera natural a tener en consideración.

La Cordillera Cantábrica abarca desde el Macizo Galaico-Leonés, en la provincia gallega de Lugo y en la castellano-leonesa de León, hasta la llamada depresión vasca, en Euskadi y Navarra, aunque desde el punto de vista geológico sigue ininterrumpida hasta las estribaciones del Pirineo. Se formó después que éste durante la orogenia Alpina, que, habiendo comenzado varios millones de años antes, culminó hacia la mitad de la era Cenozoica, hace unos 45 millones de años. Cuando la placa ibérica chocó con la placa euroasiática, las fuerzas que actuaban desde el borde de las mismas se

transmitieron hacia el interior, creando también el resto del relieve de la Península tal y como se conoce actualmente.

En esta región, como se ha adelantado, se han encontrado evidencias de poblamiento temprano y continuado hasta la actualidad. Debido a su orografía ha sido en general una zona poco transitada e incluso aislada en ocasiones. Se corresponde además con la mitad occidental del ya mencionado refugio franco-cantábrico.

En este contexto, y no sin cierta polémica sobre su origen y procedencia, surgieron los vascos; y a este contexto también pertenecen por tanto las muestras de población vasca de Gipuzkoa y Navarra estudiadas.

## Vascos.

Hasta hace poco, muchos de los habitantes de lo que hoy en día es el País Vasco, en el sentido más amplio del término, han estado distribuidos en diferentes valles, con poca comunicación y flujo génico entre ellos, quizá con la excepción de la provincia de Álava (Cardoso et al, 2011). Los vascos han permanecido por tanto relativamente aislados por motivos orográficos, pero también culturales, de los cuales el factor de mayor importancia sería la enorme diferencia entre el euskera, lengua vernácula de los vascos, y las lenguas romance vecinas, que imposibilitaría la comunicación con otras comunidades y por tanto la panmixia (García-Obregón et al, 2007). El proceso de ruptura del aislamiento comenzó a mediados del siglo XIX, con la primera revolución industrial y la llegada de emigrantes de otras comunidades.

La condición de aislamiento apoyada por algunos estudios genéticos junto con la constancia arqueológica de que esta región ha estado habitada de forma continuada desde el Paleolítico ha conducido a la idea de que los vascos son un pueblo de origen muy antiguo. No obstante, el componente genético de la población vasca actual es el resultado de múltiples migraciones, tanto prehistóricas como históricas.

La posibilidad de poder explorar el genoma de grupos humanos poco expuestos a los efectos homogenizadores del flujo génico y el mestizaje, y que por tanto podrían haber conservado un cierto número de similitudes genéticas con grupos preneolíticos, es lo que definitivamente ha incentivado el desarrollo de numerosas investigaciones dirigidas a la caracterización genética de la población autóctona del área vasca desde mediados del siglo pasado. Estos estudios genéticos se han llevado a cabo utilizando tanto marcadores clásicos (grupos sanguíneos, polimorfismos de enzimas eritrocitarias, etc.), como marcadores de ADN (Pérez-Miranda et al, 2005). Los análisis de los polimorfismos en poblaciones humanas aisladas son relevantes para lograr algunos retos científicos como aquellos relacionados con el mapeo de enfermedades y la filogeografía humana (Alfonso-Sánchez et al, 2008).



Figura I.7: Mapa del País Vasco en sentido amplio.

“El origen de este pueblo es el de su lengua”

Julio Caro Baroja (1914 - 1995)

Aunque existen términos para diferenciar a aquel que aprendió *euskera* como lengua materna (*euskaldunzahar*, euskaldun viejo) de aquel que lo aprendió como segunda lengua (*euskaldunberri*, euskaldun nuevo), la manera más habitual de designar a una persona vasca en *euskera* es, simplemente, *euskaldun*, “el que posee el *euskera*”. Por encima de cualquier otro rasgo, los vascos están caracterizados por el *euskera*, una lengua no Indo-Europea considerada un lenguaje aislado al no pertenecer a ninguna de las familias lingüísticas existentes hoy en día (Campell, 1998; Kurlansky, 1999; Comrie et al, 2003).

La población vasca actual está situada a ambos lados de la franja occidental de los Pirineos y el *euskera* es hablado por casi el 27% de los habitantes\* de estas zonas, algo más de 700.000 personas (Eusko Jaurjaritza, 2013; \*el estudio tiene en cuenta sólo a los mayores de 16 años). Éste área está dividida en siete provincias, cuatro pertenecientes a España, agrupadas en la Comunidad Autónoma Vasca y la Comunidad Foral de Navarra, y tres a Francia, comúnmente reconocidas bajo el nombre de *Iparralde*, “el lado norte”.

Existen profundas diferencias en cuanto al uso del *euskera* en el País Vasco, entendido éste como las siete provincias mencionadas, y dentro de cada una de las regiones que lo componen (Fig. I.1).

La zona con mayor proporción de vascoparlantes es la Comunidad Autónoma del País Vasco, dentro de la cual, la provincia de Gipuzkoa es la que presenta no sólo el mayor porcentaje de euskaldunes de la comunidad autónoma sino que posee el mayor número absoluto de hablantes de *euskera* de todo el territorio, 300.000 personas. Aproximadamente un cuarto de la población vizcaína es bilingüe activa, lo que supone 254.000 personas, mientras que en Álava sólo un 17% de la población emplea el *euskera* como vía principal de comunicación (46.000 bilingües).

Los datos de la Comunidad Foral de Navarra, dividida en tres territorios en función del uso lingüístico, como se aprecia en la figura I.1, muestran que el mayor número y porcentaje de personas bilingües aparece en la región

vascófona (59,3%, 30.000 personas), seguida de la región mixta, donde 28.500 personas son bilingües, lo que supone sólo un 9,8% de la población. En la región no vascófona un 2,3% de la población es bilingüe (4.500 personas).

En Iparralde también existen grandes diferencias en la capacidad lingüística de sus habitantes entre unas regiones y otras. Algo más de la mitad de la población de la Baja Navarra (Nafarroa Behera) y de Zuberoa es bilingüe, un 52,1%, lo que supone unas 17.000 personas. En el interior de Lapurdi, en cambio, tan sólo un cuarto de la población es euskaldún (23,9%, 25.500 personas). En la zona metropolitana de la costa de Lapurdi, BAM (Baiona, Angelu, Miarritze) el porcentaje de bilingües no llega al 10% (8,7%, 8.500 personas).

**Tabla I.3:** Capacidad lingüística en función del territorio en valores absolutos. (EJ, 2011)

**1. taula. Hizkuntza-gaitasuna lurraldearen arabera. Euska**

	Euskal Herria	EAE	N
Guztira	2.649.000	1.873.000	5
Elebidunak	714.000	600.000	€



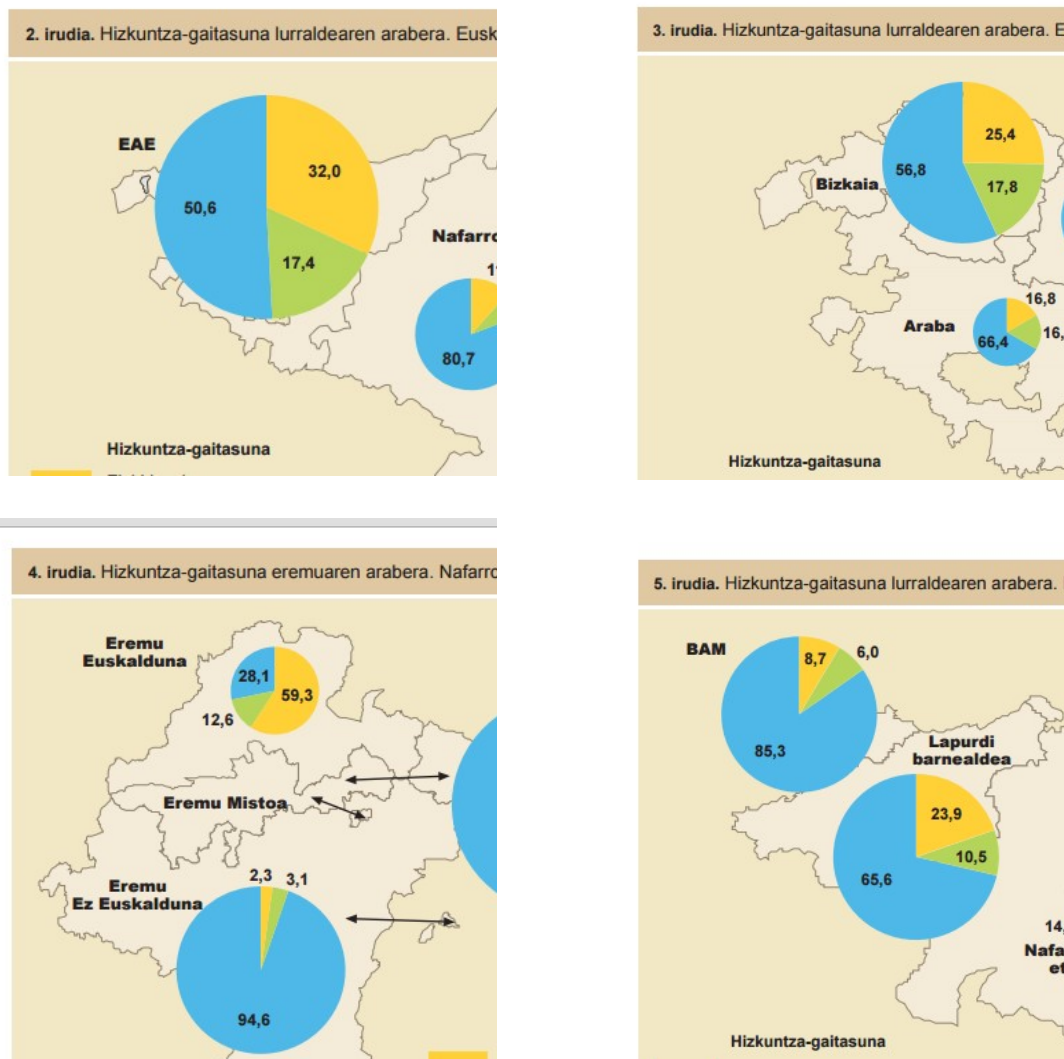


Fig. I.8. Capacidad lingüística en función del territorio. En amarillo bilingües, en verde bilingües pasivos (entienden el euskera, pero no lo hablan), en azul monolingües (castellano o francés). (EJ, 2011)

Los datos toponímicos sugieren que el *euskera*, y por tanto los vascos, ocuparon un territorio más amplio que el actual, extendiéndose desde el río Garona, al norte, hasta el río Ebro, en el sur, en época histórica (Calafell *et al*, 1994 a, b; Bertrandpetit *et al*, 1991). Algunos estudios lingüísticos apuntan incluso a que la lengua de la que derivaría el *euskera* actual pudo ser la lengua original de los europeos prehistóricos (Vennemann, 2003), siendo un relictos de la antigua diversidad lingüística europea previa a la llegada de la agricultura, con raíces en el Paleolítico (Cavalli-Sforza, 1988).

Sin embargo, la propia historia lingüística de Europa es confusa. El resto de lenguas existentes en el occidente europeo pertenecen a la familia Indo-Europea (Gorrochategui y Lakarra, 2013) y su origen es en sí mismo un tema controvertido (Mallory 2013), estando el debate polarizado entre quienes defienden la hipótesis de estepa, la cual propone que el Indo Europeo fue introducido desde el Este durante la Edad del Bronce (~4.500 años antes del presente) (Chang et al, 2015), o la hipótesis de Anatolia, que plantea que la dispersión del lenguaje Indoeuropeo ocurrió unos miles de años antes, durante el Neolítico, desde Anatolia (Bouckaert et al, 2012; Gray, 2003).

Pero lengua no es lo mismo que genética y establecer una correlación clara y directa entre ambas es complicado. Se conoce la existencia de familias vascas autóctonas que actualmente no conocen ni emplean el *euskera*. Parte de la población vasco hablante, además, no tiene orígenes familiares vascos, ha aprendido *euskera* como segunda lengua y la usa como forma principal de comunicación (Cardoso et al, 2013; EJ 2011). Los cambios demográficos de los últimos dos siglos han afectado también al uso del euskera, que redujo su presencia en los núcleos urbanos e industriales que acogieron a una gran cantidad de inmigrantes de otras provincias no vascófonas, manteniéndose en las áreas rurales, donde los flujos migratorios masivos de la segunda mitad del siglo XIX no tuvieron tanta repercusión. Esta tendencia se intensificó durante los cuarenta años de dictadura (1936 - 1975) que promovieron la persecución de las lenguas autóctonas en favor de un país monolingüe donde sólo debía tener cabida el castellano. El *euskera* se mantuvo principalmente en el ámbito doméstico de las zonas rurales mientras que fue marginado del mundo de la cultura y de la educación (Michelena et al, 1977).

A pesar de la complejidad de la historia del uso del euskera y de que numerosos estudios apuntan a que la diferenciación lingüística no siempre va ligada a una diferenciación genética subyacente (Cavalli-Sforza et al. 1994; Sajantila y Pääbo 1995; Sajantila et al. 1995), otras investigaciones basadas en marcadores clásicos han defendido la evolución paralela de lenguaje y genes en el caso de los vascos (Mourant 1947; Aguirre et al. 1991; Bertranpetit and

Cavalli-Sforza 1991; Calafell and Bertranpetit 1994; Manzano et al. 1996a,b). Es más, la continuidad genética desde la última glaciación hasta el presente ha sido propuesta como un vehículo probable de la preservación del euskera (Behar et al, 2012). Aunque existen voces contrarias a la hipótesis de que los vascos sean un relicto de los europeos Mesolíticos y por tanto, de la citada continuidad genética desde la última glaciación – y que se exponen en los siguientes párrafos – parece obvio que el papel que ha jugado el idioma en la conformación del patrimonio genético de los vascos no debe ser subestimado, y más teniendo en cuenta que el lenguaje ha demostrado ser una barrera efectiva para el flujo génico y la panmixia (Barbujani et al, 1990; 1997) como algunos autores han propuesto para el caso del *euskera* (García-Obregón et al, 2007; Alfonso-Sánchez et al, 2008). Según estos estudios, el *euskera* habría contribuido al aislamiento genético de los vascos.

La cuestión de la coevolución de genes y lenguas es considerada por tanto fundamental pero raramente se ha estudiado empleando métodos filogenéticos formales (Duda *et al*, 2016). Aunque la evolución de la genética y la lingüística pueda a menudo estar correlacionada, la asunción directa de este proceso coevolutivo puede llevar a interpretaciones erróneas (Pakendorf *et al*, 2014) ya que el proceso evolutivo que da forma a la diversidad genética no es directamente análogo a aquel que da forma a la diversidad lingüística (Creanza *et al*, 2015) y por tanto, los datos genéticos y lingüísticos pueden implicar diferentes escenarios históricos (Steele *et al*, 2010).

En cuanto a su **identidad genética**, los vascos también son uno de los grupos étnicos europeos que ha recibido más atención en el último siglo. Esta atención se ha traducido en una cantidad casi inabarcable de bibliografía presentando resultados muy diversos e incluso contradictorios, pudiendo agruparse principalmente en tres líneas: (1) origen espaciotemporal de la población vasca y continuidad de la misma población desde el Paleolítico, (2) nivel de homogeneidad de la población (3) en qué medida se trata de una población única y singular genéticamente alejada del resto de las poblaciones

de Europa y el motivo de dicha diferenciación en caso de resultar significativa, considerando, entre las principales razones un aislamiento pre o post neolítico.

No hay acuerdo en cuanto a la singularidad de los vascos cuando son comparados con sus poblaciones vecinas. Un gran número de estudios genéticos, principalmente centrados en los polimorfismos clásicos y antígenos HLA (Calafell and Bertranpetit 1994; Cavalli-Sforza et al, 1994; Lucotte and Hazout 1996; Comas, Calafell, et al, 1998; Comas, Mateu, et al, 1998; Bauduer et al, 2005), pero también basados en evidencias de ADNmt (García et al, 2011), cromosoma Y (Chikhi et al, 2002) y marcadores autosómicos (Perez-Miranda et al, 2005; Li et al, 2008; Rodríguez-Ezpeleta et al, 2010), argumentan a favor de un estado genético *outlier* distintivo y de un fuerte aislamiento y flujo genético limitado entre vascos y grupos circundantes (García et al, 2011).

Por el contrario, otros estudios han mostrado que los vascos encajan perfectamente dentro del acervo genético de Europa (Bertranpetit et al, 1995; Alonso et al, 2005; Alzualde et al, 2005; Adams et al, 2008; Garagnani et al, 2009; Laayouni et al, 2010). Desde la perspectiva del DNA mitocondrial, por ejemplo, se observa que todos los haplogrupos principales presentes en vascos contemporáneos aparecen también en el acervo materno general de Europa (Bertranpetit et al, 1995; Salas et al, 1998; Richards et al, 1996; Richards et al, 2000; Alfonso-Sánchez et al, 2008; García et al, 2011; Prieto et al, 2011). Aceptando la premisa de que la población vasca está relacionada con otras europeas, algunos autores, tomando como referencia datos *genome-wide*, detectan que los componentes genéticos característicos europeos se observan en mayor proporción en los vascos (Behar et al, 2010; Rodríguez-Ezpeleta et al, 2010).

Gargani y colaboradores (2009) analizan 123 SNPs en una región de una megabase de longitud en el cromosoma 22 en vascos ibéricos y franceses, así como en otras poblaciones de la Península Ibérica y del norte de África, y concluyen que los vascos no muestran las propiedades genéticas que se esperarían en una población aislada.

Quienes defienden la singularidad genética de los vascos la han explicado haciendo alusión a su aislamiento lingüístico y geográfico, identificándolos como candidatos potenciales a ser una población del Paleolítico con poca mezcla con las poblaciones Neolíticas que llegaron de Oriente Próximo (Chikhi et al, 2002; Bauduer et al, 2005).

Otras investigaciones, en cambio, defienden que las características genéticas únicas de los vascos aparecen como resultado de un menor flujo génico posterior a la llegada de los primeros agricultores neolíticos que el que recibieron sus poblaciones vecinas y como consecuencia de un mayor efecto de la deriva genética (Adams et al, 2008).

La similaridad genética entre grupos vascos sugerida por los antígenos HLA (Comas, Calafell, et al. 1998) y SNPs autosómicos (Rodríguez- Ezpeleta et al. 2010) ha sido puesta en duda por el grado de heterogeneidad genética hallada dentro de la población vasca mediante análisis de marcadores clásicos y autosómicos (Manzano et al. 2002; Iriondo et al. 2003; Perez-Miranda et al. 2005; Alfonso-Sanchez et al. 2008). Varios estudios defienden un origen común de los vascos con evidencias de microdiferenciación genética local de los subgrupos geográficos debidas éstas últimas a procesos de deriva genética asociada a efectos fundadores locales y / o cuellos de botella (Cardoso et al, 2011 y referencias en el interior). Esta visión se veía apoyada por las evidencias de mayores distancias genéticas de grupos vascos entre sí que las halladas entre vascos y grupos íberos (García et al, 2011).

Martínez-Cruz y colaboradores (2012) consideran que estos resultados divergentes se deben a que la mayoría de los estudios realizados sufren considerables limitaciones. Exponen como ejemplo que el uso de marcadores clásicos susceptibles a la selección natural puede reflejar la acción de factores ambientales en vez de los procesos demográficos de interés para determinar el origen de los vascos. Argumenta también que en algunos de esos estudios los vascos no son analizados en un contexto poblacional adecuado o que están definidos de manera simplista como una única población (Li et al, 2008;

Rodríguez-Ezpeleta et al, 2010), limitando por tanto la obtención de conclusiones claras sobre su relación genética con las poblaciones vecinas.

En su estudio analizan muestras de todas las poblaciones actuales donde se habla en *euskera*, donde se ha empleado este idioma hasta el siglo pasado, poblaciones francesas donde se hablaba gascón en tiempos históricos y poblaciones adyacentes al territorio vasco donde se sospecha que se habló *euskera* hasta la Edad Media. Incluyen además poblaciones vecinas españolas y francesas tomadas de la bibliografía para comparar sus resultados de los análisis de alta resolución de marcadores uniparentales (cromosoma Y y ADNmt) en casi 900 individuos de dieciocho áreas geográficas diferentes.

Observan que la singularidad genética de los vascohablantes es compartida por los grupos que habitan la región geográfica correspondiente con el llamado refugio franco-cantábrico y que el acervo genético de todas estas poblaciones es en general similar al de otras poblaciones del occidente de Europa, variando sólo las frecuencias de los marcadores analizados.

Las poblaciones vasco-hablantes sí que muestran en cambio una menor diversidad de haplogrupos de cromosoma Y y de DNA mitocondrial que las poblaciones vecinas y son, junto con franceses, los únicos grupos en los que se halla el haplogrupo mitocondrial U8a, descrito como herencia del periodo Paleolítico Superior (González et al, 2006).

En cuanto a la diversidad dentro de la población vasca detectan por un lado cierta heterogeneidad genética con respecto a las muestras no vascas en los linajes de cromosoma Y pero no de DNA mitocondrial, además de cierto nivel de estructuración genética intrínseco en esta población que atribuyen a diferentes factores culturales, geográficos e históricos.

Martínez-Cruz y colaboradores (2012) correlacionan los resultados de su trabajo con las particularidades genéticas de la población vasca y la distribución geográfica de su diversidad genética con la afiliación a las distintas tribus pre-romanas. A partir de los datos obtenidos de los análisis del cromosoma Y construyen una filogenia según la cual hace 4.500 años una

población común ancestral comenzó a separarse, siendo los primeros en “diferenciarse” las tribus aquitanas y beronas, pueblos pre-romanos que habitaban, respectivamente, la región francesa delimitada por los Pirineos, el Océano Atlántico y el río Garona, y que hoy en día se conoce como Aquitania y el territorio alrededor de la depresión del Ebro en la actual provincia de La Rioja, España, y las zonas adyacentes de las provincias de Burgos, Álava y Navarra. Se sucedieron varias divisiones hasta conformar las seis tribus descritas por los historiadores romanos a su llegada al noroeste de la Península Ibérica hace alrededor de 2.000 años (Autrigones, Caristios, Várdulos, Vascones, Berones y Aquitanos). La formación de las tribus durante la Edad de Bronce y su consolidación en la Edad del Hierro es coherente con hallazgos arqueológicos y antropológicos previos (Almagro Gorbea 2005). Este tiempo puede haber sido infravalorado ya que las mezclas poblacionales, aunque sean modestas, han demostrado que pueden resultar en una reducción del tiempo estimado de divergencia (Haber et al. 2010).

Los autores de este trabajo consideran verosímil que la formación de las tribus haya sido debida al crecimiento demográfico y económico del Calcolítico, entre los 6.000 y 4.000 años antes del presente, que estuvo asociado a un aumento de la complejidad social (Almagro Gorbea 2005) y que, por tanto, parte de las diferencias o singularidades halladas en los pueblos vascos se deban a la deriva genética causada por los altos niveles de endogamia presentes en las tribus (Martínez-Cruz et al, 2012). Otros autores, en cambio, consideran poco probable que la estructura intrínseca mencionada haya resultado de la deriva durante la expansión demográfica del Neolítico y sugieren que ésta puede datar del Mesolítico (Iriondo et al. 2003). Se ha sugerido también que antiguos patrones organizativos en torno a los ríos principales y sus tributarios podrían causar la estructuración genética y que su persistencia podría estar relacionada con pequeñas unidades poblacionales (Iriondo et al, 2003). Esto implicaría que tanto las estructuras genéticas como dialectales hayan sido provocadas por los mismos factores ambientales que aún afectan a estos grupos.

Adicionalmente, Martínez-Cruz y colaboradores (2012) informan de la existencia de barreras al flujo génico masculino en tiempos históricos que consideran consistente con los territorios de las tribus pre-romanas.

De esta investigación concluyen que la similitud genética hallada en los genomas uniparentales de vascos y algunas de las poblaciones no vascas analizadas es resultado de una menor influencia externa y un bajo flujo genético de migraciones recientes respecto a las poblaciones vecinas. Consideran que este hecho puede convertir a la población vasca – y a otras en situaciones similares, como los bretones de Francia - en buenos candidatos a representar el perfil genético de las poblaciones europeas más antiguas.

En cuanto al origen de los vascos, son muchos los investigadores que abogan por la existencia de una continuidad genética entre esta población y los europeos del Paleolítico – Mesolítico, proponiendo a la población vasca como los últimos representantes de los cazadores-recolectores que habitaron Europa durante esta época (Behar et al, 2012; Alonso et al, 1998; Cavalli-Sforza et al, 1994; Calafell et al, 1994 a y b; Bertranpetit et al, 1991; Ammerman et al, 1984).

Otros, en cambio, proponen un origen Neolítico. Belarásque y colaboradores (2010), sugieren que la mayor parte de la variedad actual del cromosoma Y proviene de los inmigrantes neolíticos de Próximo Oriente. En esta misma línea, Calderón y colaboradores (1998) sitúan el origen de la población vasca en la colonización de éste área por una pequeña población neolítica proveniente del territorio norte del Cáucaso que habría introducido la agricultura en la zona. Este grupo habría experimentado un rápido crecimiento demográfico de manera temprana y no habría hibridado con los cazadores recolectores que habitaban el área.

Uno de los artículos más recientes al respecto (Gunther et al, 2015) se opone a este último punto. A partir del análisis de los haplogrupos de ADNmt y de cromosoma Y, y de *genome wide sequence data* de restos de ocho individuos de los periodos Calcolítico (hace 5.500 años) y de la Edad de Bronce (hace 3.500 años) encontrados en El Portalón de Cueva Mayor de la



Sierra de Altamira, han concluido que los vascos actuales son descendientes de una población resultado de la mezcla entre los cazadores recolectores que vivían en la zona y un pequeño grupo inmigrante de los primeros agricultores que podrían haber portado también un lenguaje no indo-europeo. Esta población mestiza habría dado lugar a todos los grupos íberos posteriores, incluidos los vascos, que, a diferencia del resto de grupos, habrían recibido un menor aporte de flujo génico en migraciones posteriores a la península, del Imperio Romano y de los reinos moriscos que gobernaron la mayoría de la península entre los siglos VII – XIII.

La discusión y las hipótesis divergentes se extienden, de manera más general, al papel desempeñado por el refugio Franco Cantábrico en la recolonización postglacial de Europa (Alzualde et al, 2005; Rodríguez-Ezpeleta et al, 2010).

Además de los datos lingüísticos y genéticos existe otro campo que ha dado importantes evidencias de la interesante historia del pueblo vasco: la **arqueología**.

Las primeras evidencias de asentamientos humanos en el País Vasco se remontan al Paleolítico Inferior (Arrizabalaga, 2005). El resto más antiguo es un húmero femenino hallado hace más de cincuenta años en Lezetxiki, Arrasate (Gipuzkoa) (Basabe, 1966) del que recientemente se ha podido estimar la edad en 164.000 años (C De-la-Rua, 2106). Además del yacimiento de esta cueva, el campamento al aire libre de Irikaitz (Zestoa, Gipuzkoa) excavado bajo la dirección de Arrizabalaga, las terrazas del embalse de Urrunaga (Legutiano) (Sáenz de Buruaga, Fernández Eraso, Urigoitia), así como las del río Arga, cerca de Pamplona (Navarra) (García-Gazolaz) han aportado interesantes vestigios sobre esta época.

Evidencias del Paleolítico Medio y de la transición al Paleolítico Superior, así como del periodo que abarca desde este último a la Edad de Bronce, son más abundantes y no dejan lugar a duda de la continuidad del poblamiento en esta zona (Mariezkurrena, 1990). Algunos yacimientos representativos del Paleolítico Medio en País Vasco son la cueva de Arrillos en Murua, Álava

(Saéñz de Buruaga), la de Axlor en Dima, Bizkaia (Barandiarán, JE González, JJ Ibáñez), el abrigo rocoso de Olha en Kanbo, Lapurdi (Passemarde) o la cueva de Amalda en Zestoa, Gipuzkoa (Altuna), en la que se encontró un importante asentamiento del periodo Musteriense en su nivel VII.

La transición al Paleolítico Superior, al igual que en otros lugares de Europa, parece haberse dado de manera gradual desde el punto de vista cultural, con la notable excepción de la aparición del arte rupestre (Arrizabalaga, 2005, 2009), cuya eclosión hace aproximadamente 40 000 años ha dificultado el esclarecimiento de cuestiones tales como la transición de neandertales a humanos modernos o el solapamiento en el tiempo y convivencia de ambas especies (García-Obregón, 2013).

Yacimientos como Isturitz (Donamartiri, Baja Navarra), Gatzarria (Zühara, Zuberoa), Bolinkoba (Abadiño, Vizcaya), Urtiaga (Deba, Guipúzcoa) o Aitzbitarte (Errenteria, Guipúzcoa) dan cuenta de este periodo, como también lo hicieron hasta su destrucción por la actividad de una cantera la primera y por las obras de la variante de Arrasate la segunda, los de Koskobilu (Olazti, Navarra) y la cueva de Labeko Koba (Arrasate, Guipúzcoa). En cuanto a arte gráfico, las muestras más importantes de País Vasco son la cueva de Ekain (Deba, Guipúzcoa) y la de Santimamiñe (Bizkaia). En la cueva de Arlanpe, Lemoa (Bizkaia) se localizó en 2011 un gran bloque de caliza que presentaba en dos de sus caras grabados femeninos esquemáticos conocidos como la Dama de Arlanpe, asociados a ocupaciones magdalenenses de hace unos 17 500 años (Ríos-Garaizar et al, 2015). Es el primer ejemplo de este tipo que se encuentra en la Península Ibérica y se encuentra entre las más antiguas del tipo Gönnersdorf-Lalinde, lo cual supone un refuerzo para la hipótesis de que esta clase de símbolos se originó en la región franco-cantábrica y acompañó a las poblaciones que se extendieron por el Norte de Europa hacia el final de la última glaciación y apoyando la idea de una unidad cultural de esta región durante el Magdalenense (Ríos-Garaizar et al, 2015).

Como se ha expuesto anteriormente, diversos estudios apuntan a que la población vasca se ha visto menos afectada por procesos de flujo génico que

otras poblaciones de la península, conservando en mayor medida su diversidad genética inicial (Bertranpetit y Cavalli-Sforza, 1991; Peña et al., 2006). Sin embargo, una serie de sucesos históricos documentados en el territorio del País Vasco podrían haber contribuido a cierta dilución de ese aislamiento, como, por ejemplo, la llegada y el asentamiento de los romanos a partir del siglo II a.C. o las invasiones visigodas y francas (García-Obregón, 2013). Aunque la llegada de los musulmanes a la Península Ibérica fue sin duda un evento de importantes consecuencias, parece que su influencia se limitó al territorio situado al sur de Pamplona (Collins, 1989).

Más recientemente, la primera revolución industrial en el País Vasco a mediados del siglo XIX atrajo a emigrantes de las comunidades vecinas, fundamentalmente de la Cornisa Cantábrica, Castilla y León y La Rioja, incrementando de manera gradual el tamaño de la población e iniciando el proceso de ruptura del mencionado aislamiento (Alfonso-Sánchez et al, 2001). La segunda revolución industrial, acaecida en el siglo XX, amplificó los efectos de la primera de manera que actualmente, un porcentaje considerable de la población residente en el País Vasco tiene ancestros con orígenes en otras regiones españolas (García-Obregón, 2013). Es más, algunos autores consideran que, a efectos genéticos, sólo un cuarto de la población asentada en el área del País Vasco puede ser considerada autóctona (Calderón *et al*, 1993). Por tanto, los criterios empleados durante el proceso de muestreo para la selección de individuos autóctonos deben ser especialmente rigurosos (García-Obregón, 2013).

Las áreas en las que se ha realizado el muestreo, la provincia de Gipuzkoa y el extremo norte de la provincia de Navarra, son las dos regiones del territorio vasco tradicional que han conservado mejor las características culturales y genéticas expuestas, por el hecho de encontrarse casi completamente rodeadas de otras poblaciones vascas, lo cual ha repercutido también en una mayor preservación del *euskera* (Alfonso-Sánchez et al, 2005).

## POBLACIONES Y ÁREAS CONCRETAS ESTUDIADAS.

### Vascos de Gipuzkoa.

La Comunidad Autónoma del País Vasco, también llamada Euskadi, supone aproximadamente un 1% de la superficie peninsular. Situada en el extremo oriental de la Cornisa Cantábrica, consta de tres provincias, Álava, Bizkaia y Gipuzkoa.

Algunos estudios señalan a Gipuzkoa como la provincia vasca con mayor peculiaridad genética, mientras que los niveles más altos de mezcla poblacional se encuentran en Álava, lo que concuerda con el mapa lingüístico del euskera (Calderón et al., 1998; Alfonso-Sanchez et al, 2005).

Gipuzkoa es, con 1980 km<sup>2</sup>, la provincia más pequeña de la Comunidad Autónoma Vasca y de todo el estado. Se ubica en la unión de la Cordillera Cantábrica (al oeste) y de los Pirineos (al este), por lo que presenta una orografía muy accidentada. Posee además 86 km de costa que alterna arenales con montes costeros. Sus acantilados componen una de las formaciones geológicas más importantes y reconocidas del mundo, el *flysch*, capas creadas por acumulación de sedimentos debajo del mar hace millones de años, que, tras el levantamiento de los Pirineos y la erosión marina, hoy son capas de rocas casi verticales que dejan al descubierto 50 millones de años de historia.

Gipuzkoa está totalmente rodeada por territorio tradicionalmente vasco – Bizkaia al oeste, Álava al sudoeste, Navarra al sudeste e Iparralde al este – y quizá por este motivo es la zona con mayor prevalencia del *euskera*: 300 000 bilingües (49.9% de la población gipuzkoana) que emplean principalmente este idioma como lengua para comunicarse y en total alrededor de 400 000 personas entienden el idioma.

La región del Goierri, de donde proviene la muestra analizada, está ubicada en el sudeste de la provincia. El término Goierri significa “pueblo alto” y hace referencia a su localización en las zonas altas de los valles del río Oria y del Urumea, relativamente aislados hasta mediados del siglo XIX, en

especial el valle del Urumea, que es más angosto y profundo. El Goierri gipuzkoano está enclavado entre los Parques Naturales de Aralar y de Aizkorri-Aratz, a alrededor de 40 km de la costa y de la capital de provincia, Donostia. Tanto en la Sierra de Aralar como en Aizkorri se han encontrado túmulos y menhires de varios miles de años de antigüedad.

La comarca está formada por 18 municipios, en su mayoría pueblos rurales, aunque en la actualidad aproximadamente el 50% del empleo se concentra en el sector industrial y la agricultura y la ganadería son consideradas actividades económicas secundarias. Dentro del sector servicios, que ocupa al 46% de la población, el turismo ha incrementado su presencia en los últimos años.

Algunos estudios señalan Gipuzkoa como la provincia vasca con mayor peculiaridad genética (Calderón et al., 1998). En el trabajo de Martínez-Cruz y colaboradores (2012), aunque en general los vascos muestran una mayor diversidad de haplogrupos de cromosoma Y que los encontrados en estudios anteriores, la muestra de Guipúzcoa que analizan presenta niveles sustancialmente más bajos de diversidad que el resto de las poblaciones estudiadas por este equipo.

## **Vascos de Navarra.**

La Comunidad Foral de Navarra, con una extensión de 10.390 km<sup>2</sup> y una población de 638.390 habitantes en 2013, se encuentra en el noreste de España y limita al norte con el departamento francés de Pirineos Atlánticos, al oeste con Euskadi, al suroeste con la Rioja y al este y sureste con la Comunidad Autónoma de Aragón.

A pesar de ocupar tan sólo aproximadamente del 2% de España, se sucede de norte a sur una gran diversidad geográfica, orográfica y cultural; desde la región occidental de Pirineos, que supone su frontera norte y donde se ubica la Selva de Irati, uno de los mayores y mejor conservados hayedos de Europa a pesar de su explotación forestal desde los siglos XVI – XVII, hasta

el paisaje semi-desértico de las Bardenas Reales, al sureste, que delimita con la Comunidad Autónoma de Aragón, donde posee un exclave, Petilla de Aragón.

Por su relieve y geografía suele dividirse en tres zonas de norte a sur: La montaña de Navarra, accidentada y abrupta, con un clima húmedo y de precipitaciones frecuentes; Navarra Media, zona de transición con sierras en su parte norte y amplias llanuras al sur; La Ribera, una zona más seca y de extensas llanuras.

La ciudad más poblada de Navarra es su capital, Pamplona, ubicada en la zona centro de la comunidad y habitada por 195.650 personas, acogiendo su área metropolitana a alrededor de 335.000 habitantes (INE, 2016).

2. El vascuence tendrá también carácter de lengua oficial en las zonas vascoparlantes de Navarra.

Una ley foral determinará dichas zonas, regulará el uso oficial del vascuence y, en el marco de la legislación general del Estado, ordenará la enseñanza de esta lengua.

Ley Foral del vascuence 18/1986, de 15 de diciembre. BON Nº 154 de 17 de Diciembre de 1986.

El norte de Navarra posee en común con la Comunidad Autónoma del País Vasco e Iparralde, el País Vasco Francés, el euskera, un idioma de origen pre-indoeuropeo con varios dialectos. Son por tanto lenguas propias de Navarra tanto el castellano como el euskera de acuerdo con la Ley foral del Vascuence de 1986 (LEY FORAL 18/86, de 15 de diciembre de 1986). Según esta ley, la Comunidad Foral de Navarra está lingüísticamente dividida en tres zonas:

1. Una zona vascófona, en la frontera con Gipuzkoa, provincia con la que comparte el mayor porcentaje de prevalencia del euskera como principal forma de comunicación. En esta región vive algo menos de una décima

parte de la población total de la comunidad y alrededor de un 73% de la población comprende el idioma, siendo casi el 60% totalmente bilingüe (30 000 personas).

2. Una zona mixta donde vive la mayor parte de la población navarra (aproximadamente un 54%) y predomina el castellano. Alrededor de 28 500 personas de esta zona son bilingües (9.8%) y otro tanto son bilingües pasivos (entienden el idioma, pero no lo hablan).
3. Una zona no vascófona, que ocupa algo más de dos tercios de la superficie de la comunidad, donde la población que comprende el euskera apenas sobrepasa el 5%, siendo el 2,3% (4 500 personas) el porcentaje de habitantes que emplean el euskera como vía de comunicación principal.

Actualmente el castellano es oficial en todo el territorio de la comunidad. El euskera tiene carácter de lengua cooficial sólo en la zona vascófona aunque la Ley del Vascuence incluye una serie de disposiciones para garantizar el uso del euskera en la zona mixta. En la zona no vascófona, por el contrario, las entidades públicas no tienen obligación de atender en este idioma.

Según Cardoso y colaboradores (2011), a pesar de ser una fracción importante de la comunidad vasca ibérica, Navarra ha sido poco estudiada desde el punto de vista genético hasta la última década, cuando varios estudios se han centrado en la diversidad genética de este grupo (Pérez-Miranda et al., 2004, 2005; Calderón et al., 2006; García-Obregón et al., 2007). El propio artículo de estos autores cubre el vacío existente en esta población en cuanto a la variabilidad de los linajes maternos.

Su composición genética en cuanto al ADN mitocondrial reveló un origen común con el resto de los grupos vascos de la Península Ibérica con ciertos signos de microdiferenciación local respecto de los mismos (Cardoso et al, 2011).

Los primeros pobladores de Navarra habitaron la región desde el Paleolítico Inferior, como demuestran los hallazgos de Coscobillo, Urbasa, Estella, Lezáun, Lumbier y Viana. El Neolítico transforma a los cazadores-recolectores en pastores y agricultores y la huella de la Edad de Bronce en la región se encuentran en forma de dólmenes y talleres de sílex, desde Viana, Cirauqui y Artajona, a las sierras de Urbasa y Aralar, hasta alcanzar las cumbres pirenaicas. La Edad del Hierro aporta a los primitivos habitantes vascones nuevas técnicas y concepciones de vida traídas por los celtas y celtiberos del centro de Europa.

## Residentes

En cuanto a la población residente en el País Vasco, se ha considerado una pequeña muestra representativa de esta. Se trata de una comunidad en la que ha existido un mestizaje reciente entre población autóctona e inmigrantes provenientes del resto de España, fundamentalmente durante la segunda mitad del siglo XX.

La primera revolución industrial en el País Vasco, hacia mediados del siglo XIX, también atrajo a emigrantes de comunidades vecinas, principalmente de la cornisa Cantábrica, Comunidad Autónoma de Castilla León y de la Rioja (Alfonso-Sánchez et al, 2001). Algunos autores creen que a efectos genéticos tan solo alrededor del 25% de la población asentada en este territorio podría ser realmente considerada autóctona (Calderón et al, 1993).



## LA REGIÓN MEDITERRÁNEA.

El actual mar Mediterráneo baña las costas del norte de África, de Europa meridional y de Asia occidental y se comunica con el océano Atlántico a través del estrecho de Gibraltar, con el mar Negro a través de los estrechos del Bósforo y de los Dardanelos y con el mar Rojo mediante una vía artificial de navegación denominada canal de Suez.

En su interior existen numerosas islas a poca distancia de los territorios continentales; esto, unido a su poca profundidad y escasas corrientes, ha facilitado la navegación y el contacto entre los pueblos que han surgido en sus riberas, donde han florecido y desaparecido numerosas civilizaciones, muchas de las cuales han tenido una gran influencia militar, económica y cultural (García-Obregón, 2013).

La cuenca del Mediterráneo se considera cuna de la cultura occidental y las relaciones entre el sinnúmero de civilizaciones que prosperaron en ella, conocidas desde la prehistoria, se intensificaron a partir del siglo VIII a.C., cuando navegantes fenicios, y griegos después, surcaron este mar en busca de metales y otras materias primas (MPV, 2014).

La región oeste del Mediterráneo en concreto, donde se ubica la Península Ibérica, ha sido escenario de sucesivos conflictos de origen económico, religioso y político (García-Obregón, 2013). Como resultado de ello, la península en general y en particular la zona mediterránea de la misma, ha sido objeto de numerosas influencias externas con consecuencias demográficas y genéticas más o menos importantes (Adams et al, 2008; Hernández et al, 2015). Por ejemplo, la elevada frecuencia en la población actual de Valencia de la mutación G542X, que produce fibrosis quística, ha sido explicada por algunos autores aduciendo que esta variante del gen se introdujo en la Península a través del mar Mediterráneo, hace entre 2.500 y 3.000 años, posiblemente por los fenicios (Casals *et al.*, 1993).

En épocas más tempranas, la región mediterránea peninsular se ha destacado por el descubrimiento de asentamientos tardíos de Neandertales y

por la entrada temprana del neolítico (Barciela et al, 2013), mientras que ya en tiempos históricos la anexión del territorio al Imperio Romano primero (siglo III a.C.) y los 700 años de dominación musulmana después (siglos VIII al XII), han marcado la cultura y la genética de este pueblo.

Por tanto, la franja mediterránea de la Península Ibérica se ha caracterizado históricamente por su dinamismo demográfico y los continuos movimientos migratorios que, tras la llegada del Neolítico por vía marítima a la costa valenciana, se sucedieron en buena parte de las zonas costeras del sur peninsular. Los intensos flujos migratorios que habrían tenido lugar en esta región a lo largo de la historia podrían haber promovido procesos de mestizaje entre las diversas poblaciones ribereñas que tradicionalmente han utilizado el Mediterráneo como vía de comunicación. De esta forma, los recurrentes desplazamientos humanos en la región habrían quedado reflejados en una alta hibridación del genoma y, consecuentemente, una notable variabilidad genética de las poblaciones mediterráneas contemporáneas (García-Obregón, 2013).

POBLACIONES ESTUDIADAS.

## **La Comunidad Valenciana.**

La Comunidad Autónoma de Valencia, de forma estrecha y alargada, está situada en la costa Mediterránea, en la zona este de España. Integrada, de norte a sur, por las provincias de Castellón, Valencia y Alicante, limita al norte con Cataluña y Aragón, al oeste con Castilla-La Mancha y Aragón y al sur con la Región de Murcia. Ocupa un área total de 23.305 km<sup>2</sup>, con una línea de costa de 474 km. Su densidad poblacional es de 220,57 habitantes por kilómetro cuadrado (INE 2012).

Las montañas ocupan la mayor parte de la Comunidad Valenciana, quedando en algunos puntos muy cercanas a la costa y dejando una estrecha franja de litoral que alterna acantilados con zonas húmedas, albuferas y arenales.

La provincia de Valencia es la mayor de las tres que conforman la comunidad autónoma y aproximadamente el 70% de su población vive en el área metropolitana de su capital, Valencia. En su estatuto de autonomía la lengua valenciana (*valencià* o *llengua valenciana* en valenciano) está considerada lengua propia aunque algunos lingüistas la consideran una variante dialectal del catalán (Lewis, 2009).

Seleccionada como población representativa de la cuenca mediterránea por su posición estratégica en la región levantina desde tiempos remotos (García-Obregón, 2013), los estudios genéticos de los habitantes actuales de Valencia pueden aportar datos sobre el poblamiento de la cuenca mediterránea, la influencia de las migraciones de los pueblos ubicados en sus orillas en el patrimonio genético de la Península Ibérica y sobre el papel del Mediterráneo como barrera al flujo génico entre los Europa y África (García-Obregón, 2013). A pesar de este interés, los trabajos genéticos-evolutivos llevados a cabo en Valencia son relativamente escasos. Tal y como se recoge en la tesis doctoral de García-Obregón (2013), se han publicado datos de microsatélites autosómicos (Aler *et al.*, 2001a; Tomas *et al.*, 2001a), microsatélites del cromosoma Y (Aler *et al.*, 2001b; Rodríguez *et al.*, 2009), microsatélites del cromosoma X (Aler *et al.*, 2007) y ADNmt (Picornell *et al.*, 2005).

## PRIMEROS POBLADORES.

“Las evidencias más antiguas corresponden a homínidos del linaje neandertal, según los primeros niveles de ocupación de la Cova del Bolomor datados hace más de 350.000 años. Un hábitat de grupos de cazadores-recolectores cuyas características denotan un triunfo adaptativo y, posiblemente, el primer poblamiento humano consolidado de estas tierras”

Cuartero y colaboradores, 2013

Hace más de un millón de años, un homínido africano descendiente de alguna de las ramas evolutivas que surgen de *Homo habilis*, comenzó la colonización de Europa y derivó en un linaje homínido característico de este

continente, el *Homo neanderthalensis*, donde evolucionó hasta bien avanzado el Pleistoceno superior.

Las tierras que actualmente ocupa la Comunidad Valenciana fueron uno de los últimos refugios de los Neandertales, cuyo hábitat parece que se redujo a zonas europeas aisladas y periféricas, hasta su extinción hace en torno a 25 000 años. Desde hacía unos 15 000 años antes, *Homo sapiens* ya ocupada prácticamente todo el continente europeo. El primer borrador del genoma de *Homo neanderthalensis*, conocido en 2010, confirma la presencia de genes neandertales en poblaciones euroasiáticas, contribución que se debió producir por un cruce entre especies, cuando el ser humano moderno llegó a Oriente Medio tras su salida de África, hace unos 200.000 años (Green et al, 2010). Durante la ocupación Neandertal floreció la llamada cultura Musteriense, bien ejemplificada en más de cincuenta yacimientos por todo el territorio valenciano (Barciela et al, 2013).

El conocimiento acerca de la evolución del Pleistoceno - el período de escala temporal geológica que comienza hace unos 2,6 millones de años, caracterizado por la aparición del *Homo sapiens* y por la alternancia de épocas glaciales y otras más templadas - en el territorio de la Comunidad Valenciana, ha estado muy condicionado por el estudio de terrazas fluviales y depósitos marinos discontinuos que han proporcionado una información desigual y por las estratigrafías presentes en la mayoría de las cuevas del territorio, que se limitan principalmente al Pleistoceno superior (hasta los 50 000 años de antigüedad).

Afortunadamente existen excepciones. Las primeras evidencias de asentamientos humanos en esta zona están fechadas en el Paleolítico Inferior, hace más de 350 000 años, y pertenecen al yacimiento de la Cova de Bolomor (Tabernes de Valldigna, Valencia) que presenta, como excepción a lo expuesto en el párrafo anterior, una cronoestratigrafía continua de unos 350 000 años (Barciela et al, 2013).

Los restos óseos encontrados en esta cueva han sido clasificados dentro de la especie *Homo neanderthalensis*, salvo la pieza más antigua (HCB

05), de rasgos más arcaicos y más similar por tanto a otros restos hallados en yacimientos como Atapuerca (Burgos), Krapina (Croacia) o Mauer (Alemania), vinculados a la especie *Homo heidelbergensis*. La Cova de Bolomor fue el hábitat de grupos de cazadores-recolectores y, posiblemente, el primer poblamiento humano consolidado de estas tierras (Barciela et al, 2013)..

A partir del Paleolítico Medio se registra un mayor número de restos fósiles pertenecientes al *Homo neanderthalensis*, coincidiendo con la época en la que estos humanos prehistóricos vivieron en Europa (entre 250.000 y 40.000/30.000 años antes del presente). Además de los hallados en la Cova de Bolomor, se han encontrado evidencias de neandertales con una antigüedad estimada entre 150 000 y 35 000 años antes del presente en la región valenciana de Xàtiva, en el yacimiento de Cova Negra. Se han hallado también restos de un adulto joven en el abrigo del Salt, Alcoi, donde han aparecido, así mismo, vestigios de la fauna y las industrias propias de este periodo. (Pla-Ballester, 1985; Fernández-Peris *et al.*, 2008).

Del Paleolítico Superior (hace 35.000 años) se han encontrado distintos útiles localizados en Cova del Parpalló (Gandía), Cova de les Malladetes (Barx), Cova de les Cendres (Moraira- Teulada), la Cueva del Volcán del Faro (Cullera) y la Ratlla del Bubo (Crevillent) (García-Obregón, 2014).

El *Homo neanderthalensis* desapareció de Europa y Asia hace unos 40 000 años (Higham et al, 2014) pero varias evidencias sugieren que algunas poblaciones sobrevivieron en la periferia de estos continentes hasta hace unos 25 000 años (Finlayson et al, 2006; Higham et al, 2014). La primera fecha parece coincidir con la llegada del *Homo sapiens* a Eurasia. En el territorio valenciano es en este momento cuando se observa un mayor número de yacimientos con restos humanos, entre los que destacan la Cova del Parpalló (Gandia), la Cova de les Malladetes (Barx) y la Cova del Barranc Blanc (Ròtova) (Barciela et al, 2013).

El mosaico de poblaciones en Europa durante la transición del Paleolítico Medio al Superior sugiere que hubo tiempo suficiente para la transmisión de comportamientos culturales y simbólicos, así como de intercambio genético, entre ambos grupos (Higham et al, 2014).

“El modo de vida, basado en la caza y la recolección, se mantuvo durante el Epipaleolítico, con las consecuentes adaptaciones a los cambios en el paisaje y la fauna que tienen lugar en este momento. Evidencias de estos últimos cazadores se encuentran en yacimientos como la Cova de les Malladetes (Barx), la Cueva de la Cocina (Dos Aguas), la Covacha de Llatas (Andilla) y El Collao (Oliva). Durante el Epipaleolítico (10.000 años a.C.) se desarrollaron las últimas sociedades cazadoras, cuyo utillaje lítico se caracterizó por el microlitismo y las formas geométricas, destinado a armar las puntas de sus flechas. Ejemplos de estos materiales se han encontrado en las cuevas de la Cocina (Dos Aguas) y la Covacha de Llatas (Andilla), con una antigüedad estimada entre 7.000 y 5.000 años a.C. (Pla-Ballester, 1985).”  
(Barciela et al, 2013).

### LA LLEGADA DE LA AGRICULTURA Y LA GANADERÍA.

En el territorio valenciano, hace unos 12.000 años, se inicia el periodo conocido como Epipaleolítico-Mesolítico. Es la etapa de los últimos cazadores y recolectores, que entran en contacto con las primeras sociedades productoras del Neolítico y que acaban incorporándose al nuevo modo de vida.

El inicio del Neolítico, marcado por la llegada de los primeros colonos agricultores y ganaderos tras una rápida expansión desde el este por la cuenca mediterránea, se sitúa a mediados del VI milenio a.C, en torno a hace 5 500 años (Zilhao, 2001). Se encuentran yacimientos con niveles de corral como la Cova de l'Or (Beniarrés) y la Cova de les Cendres (Moraira-Teulada) que muestran que la ganadería toma un peso especial en emplazamientos situados en cuevas habitacionales por debajo de los 600 metros de altitud (Barciela et al, 2013). A lo largo de los cursos fluviales, en cambio, se multiplican los asentamientos con grandes viviendas de forma rectangular con un extremo en forma de ábside con un espacio interior subdividido para albergar distintas actividades (horno, molienda, etc.) que se van expandiendo, dando paso a finales del Neolítico a las grandes aldeas de los inicios de la metalurgia como la Ereta del Pedregal (Navarrés) o Les Jovades (Cocentaina). También se encuentran enterramientos acompañados de vasos cerámicos y elementos de adorno, como los de la Cova de la Sarsa (Bocairent), que junto con la estructura explicada anteriormente dan idea de una organización social más compleja (Barciela et al, 2013).

Consolidado el Neolítico, el arte, al principio basado en sencillas figuras en las piezas de cerámica o en algunos abrigos rocosos, adopta un estilo más figurativo, donde se muestra a los personajes en escenas de caza, recolección, danzas o enfrentamientos. Este arte rupestre del Arco Mediterráneo de la península Ibérica, una de las manifestaciones más genuinas de nuestra Prehistoria, fue declarado en 1998 Patrimonio de la Humanidad por la UNESCO (MPH, 2014).

#### INFLUJO DE LAS CULTURAS MEDITERRÁNEAS.

Entre los siglos VII y VI a.C. se instalan asentamientos fenicios y griegos en la costa mediterránea peninsular. Se considera que este hecho provoca el enriquecimiento de la emergente sociedad local debido al intercambio de productos, conocimientos y costumbres (interacción cultural y social) que a partir del siglo VI se conocerá como cultura ibérica (MPH, 2014). A partir de ese momento otros pueblos de origen mediterráneo (cartagineses, romanos, árabes y bereberes) colonizaron territorios del este y sur de la Península Ibérica (García Obregón, 2011).

“Iberos es el término que los geógrafos e historiadores griegos y romanos utilizan para denominar a los diversos pueblos que ocupan, entre los siglos VI y I a.C., una amplia franja del litoral mediterráneo, entre el río Hérault en Francia y el Guadalquivir en Andalucía. Los iberos que habitan las tierras valencianas de norte a sur son los ilerconvones, edetanos y contestanos.”

Museo de la prehistoria de Valencia, 2014



**Figura I.8.** Mapa: Principales yacimientos ibéricos en tierras Valenciana (MPH, 2014)

Los íberos fueron por tanto una parte fundamental del poblamiento prerromano de la Comunidad Valenciana y mantuvieron intercambios activos, gracias a su cercanía al mar, con comerciantes fenicios, griegos y cartagineses principalmente.

El litoral valenciano quedó anexionado al Imperio Romano tras su derrota en la Segunda Guerra Púnica (año 202 a.C.) durante setecientos años, periodo en el que los íberos adquirieron el latín como lengua propia y participaron de la organización política, económica y social del Imperio (García-Obregón, 2013). La ciudad de Valencia, bajo el nombre *Valentia Edetanorum*, fue fundada como colonia romana en el año 138 a.C.

Varios siglos después, en el año 711 los musulmanes ocuparon la ciudad, aportando su lengua, religión y costumbres, como la implantación de sistemas de riego y la introducción de nuevos cultivos.



“Aparte de algunas voces árabes o toponimia, no parece decisiva la influencia de mozárabes o árabes en la configuración básica de la lengua en los territorios valencianos.

Las evidencias históricas demuestran que las comarcas de la costa, hoy en día consideradas valencianohablantes, fueron repobladas por catalanes y las del interior, consideradas como castellanoahablantes, fueron repobladas por aragoneses aragonohablantes, cuya lengua acabaría siendo absorbida posteriormente por el castellano (Cabanes y Ferrer, 1980).”

El siguiente gran impacto demográfico y cultural se debió por una parte al dominio musulmán de este territorio durante los siglos VIII al XIII, que en el año 711 ocuparon la ciudad de Valencia y aportaron su lengua, religión y costumbres, además de implantar sistemas de riego e introducir nuevos cultivos, y por otra a la posterior formación del Reino de Valencia a partir de los reinos taifas reconquistados entre 1232 y 1245 por el rey de Aragón Jaime I el Conquistador. Las repoblaciones cristianas desde Cataluña y Aragón dieron lugar al surgimiento del valenciano como idioma, que comenzó a emplearse por los colonos procedentes de los condados occidentales catalanes que se asentaron principalmente en la región costera de la actual Comunidad Valenciana, mientras que el territorio interior fue mayoritariamente repoblado por aragoneses (Cabanes y Ferrer, 1980).

De este modo, la población asentada en territorio valenciano ha transitado por diferentes períodos caracterizados por el protagonismo del flujo génico, cuya notable intensidad ha debido dejar una importante huella en su patrimonio genético (García-Obregón, 2014)

## Los nómadas que llegaron de Asia. El pueblo *Rrom*.

“Los más de diez millones de gitanos que viven en Europa constituyen la mayor minoría étnica del continente. Sin embargo, la historia del pueblo gitano (...) es una de las más desconocidas debido a la inexistencia de una tradición oral o escrita sobre su origen y dispersión”

Isabel Mendizabal y David Comas.

Investigación y Ciencia, noviembre de 2013.

El pueblo rrom - también denominado roma o romaní y más habitualmente en castellano, gitano - es un conglomerado de poblaciones fundadoras de raíces asiáticas embebido en una población caucásica genéticamente diferente.

Las ciencias sociales y los estudios de lingüística comparada apuntaron a un origen asiático para todas las poblaciones de gitanos dispersas por el mundo. Esta propuesta fue posteriormente apoyada por estudios de genética de poblaciones basados en SNPs, haplotipos de microsatélites del cromosoma Y y en haplotipos de ADN mitocondrial (Gresham et al, 2001; Kalaydjieva et al, 2001a; Morar et al, 2004; Almos et al, 2008).

Los romanís son una diáspora sin ningún documento escrito fiable, ni histórico ni genealógico, un pueblo ágrafo de tradición oral. Con un estilo de vida principalmente nómada y prácticas sociales endogámicas, las poblaciones geográficamente dispersas de Roma han sido históricamente marginadas y perseguidas. (Fraser, 1992)

A continuación, se presentan los orígenes del pueblo gitano y las migraciones y eventos principales que configuraron su situación genética y geográfica actual, con especial énfasis en los gitanos ibéricos, una de cuyas subpoblaciones, los gitanos vascos, ha sido analizada para este trabajo.

### **Orígenes asiáticos.**

Las evidencias más tempranas del origen indio del pueblo gitano fueron aportadas desde el ámbito de la lingüística, apareciendo las primeras

disertaciones al respecto a finales del siglo XVIII en un trabajo titulado “*Zigeuner in Ungaren*”, del etnógrafo Samuel ab Hortis Augustini (1729 - 1792). Hoy en día existen varios trabajos que apoyan esta hipótesis (Fraser et al, 1998; Gresham et al, 2001, Morar et al, 2004; Malyarchuk et al, 2006) y se considera que el romaní surgió en el noroeste del subcontinente indio, derivado de lenguas ampliamente extendidas en la región relacionadas con el sánscrito (Matras et al, 2006).

Además de las similitudes lingüísticas entre el romaní y las lenguas indoeuropeas de la India, varios rasgos antropológicos – como el nomadismo o la estructura social – concuerdan con un origen indio del pueblo rrom aunque la confirmación definitiva ha llegado de mano de la genética.

En los últimos años, varios estudios han tratado de revelar la historia de los gitanos europeos, convergiendo en una serie de conclusiones generales, como (1) el origen Indio de este pueblo y (2) que, a pesar de la gran diversidad genética que muestran los diferentes grupos de rrom debido a su rápida expansión por Europa, las subpoblaciones gitanas son más similares entre ellas que con las poblaciones locales con las que comparten espacio (Gresham et al, 2001; Morar et al, 2004; Kalaydjieva et al, 2001; Gusmão et al, 2008). Estos trabajos han abordado la investigación del pueblo rrom no sólo desde el punto de vista del cromosoma Y o del DNA mitocondrial (Gresham et al, 2001; Kalaydjieva et al, 2001; Gusmão et al, 2008), que debido a su naturaleza haploide están más inclinados a mostrar el efecto de la deriva, sino también empleando marcadores autosómicos y *genome-wide data* (Gusmão et al, 2010; Mendizabal et al, 2012). La región geográfica concreta de donde proceden con más probabilidad los *Rrom* es la zona del Punjab, en el noroeste de la India, según las evidencias genéticas disponibles (Kalaydjieva et al, 2001; Mendizabal et al, 2011), lo que concuerda con las pruebas lingüísticas expuestas en este mismo texto con anterioridad (Marushiakova et al, 1997).

Otra de las conclusiones más significativas de los estudios genéticos realizados en torno a los gitanos es que todas las poblaciones Rrom,

independientemente del país de residencia, descienden de un único grupo, probablemente pequeño y de individuos relacionados. Sobre la identidad o proveniencia de este pequeño grupo de proto-gitanos existen varias hipótesis que varían desde su pertenencia a los estratos más bajos del sistema de castas de la India a una sociedad compuesta por guerreros y seguidores de campamentos (Fraser, 1992; Kalaydjieva et al., 2005).

### ***Out of India.* Migraciones. Dispersión por el mundo.**

“La dispersión de los Rroma por Europa representa uno de los más notables movimientos migratorios de personas en tiempos históricos recientes”  
Mendizabal et al, 2011.

Entre la realidad y la leyenda está la etimología de la palabra empleada en castellano para referirse a este pueblo. Se cuenta que cuando llegaron a la Península Ibérica procedentes del este, los rrom fueron confundidos - o se dejaron confundir - con nobles egipcios en peregrinación a Santiago de Compostela. Del término egipcio habría derivado no sólo gitano sino también las voces inglesa, francesa y vasca, respectivamente, *gypsy*, *gitan* e *ijito* (Hancock, 1993). Actualmente, la séptima acepción de la palabra gitano según la Real Academia de la Lengua Española, aunque en desuso, es egipcio.

Al parecer, uno de los primeros documentos que recoge la entrada de gitanos en la Península Ibérica, data de principios del siglo XV. Se trata de un salvoconducto expedido por el rey Juan II de Aragón en el cual son identificados como condes del Egipto Menor (Singhal, 1982).

La evidencia combinada de datos lingüísticos, antropológicos y genéticos sugiere que hace entre 1000 y 1500 años el pueblo gitano migró desde el Punjab y viajó a través de Asia, a lo largo de la antigua Persia, hoy Irán, Armenia y Turquía.

El grueso de esta primera ola migratoria se movió hacia Grecia y las partes de habla eslava de los Balcanes (Fraser et al, 1992), parece que huyendo de los turcos. Está mayoritariamente aceptado que los rrom se asentaron en esta región entre los siglos XI y XII, donde permanecieron

durante 200 años antes de comenzar de nuevo su dispersión por Europa (Fraser et al, 1992; Marushiakova et al, 1997).

La dispersión por el continente europeo fue muy rápida, ya que una diáspora temprana apareció en el oeste de Europa entre los siglos XIV y XV y no tardó en alcanzar también el extremo norte del continente. De hecho, documentos históricos sitúan a los gitanos en Cataluña en 1425 y para el final del siglo XV se encontraban ya en toda la península. La ruta más importante de entrada a la misma parece haber sido la trans-Pirenaica (Callén et al, 2005). Los datos arrojados por estudios del cromosoma Y en gitanos ibéricos refutan la importancia de una posible segunda vía de entrada a la Península desde el norte de África (Gusmão et al, 2008)

Posteriormente se sucedieron tres grandes olas migratorias mucho más recientes hacia el occidente de Europa. Primero, la dispersión que ocurrió durante la última parte del siglo XIX tras la abolición de la esclavitud del pueblo Rrom en el antiguo reino rumano/ imperio Hasburgo en 1841 (Liégeois, 1994; Fraser, 1992; Hancock, 1987). Después, la salida de Yugoslavia durante las décadas de los 60 y 70 del siglo XX; La tercera gran diáspora se ha dado durante los últimos años del siglo XX, debido a los cambios políticos y económicos en el Este de Europa, principalmente la inestabilidad socio-económica en el área de los Balcanes durante los años 90 del siglo XX y la ampliación de la Comunidad Europea (Kalaydjieva et al, 2005; Corsi et al, 2008; Mendizabal et al, 2011). Ésta última ola migratoria aún continúa.

Los Rrom alcanzaron América, y de manera menos significativa otras partes del mundo, de forma paralela a los europeos. Aunque los datos son puramente estimativos, debido a la persecución que ha sufrido y sigue sufriendo este colectivo y que obligó a personas gitanas a renunciar o a ocultar su identidad étnica, está mayoritariamente aceptado que llegaron a América en uno de los viajes de Colón, que durante los siglos XIX y XX buscaron fortuna como el resto de los europeos en el mal llamado “Nuevo Continente” (Gamboa-Martínez et al, 2005).

## El pueblo rrom en la actualidad. Perfil genético. Gitanos en la Península Ibérica.

"Motho manqe, Rrom! ea, kaja amari phuv, amare plaja, amare lená, amare umàla Thaj amare vesa?	"Cuéntame, gitano, ¿dónde está nuestra tierra, de nuestras montañas, nuestros ríos, nuestros campos y nuestros bosques?
Kaj amaro them? Kaj amare limòra? -And-e lava tane, amare chibaquere"	¿Dónde está nuestra patria? ¿Dónde nuestros sepulcros? Están en las palabras, dentro de nuestra lengua"

ESLAM DRUDAK

No se sabe con exactitud la cantidad de personas que componen el pueblo rrom. Por una parte, en varios países los censos discriminatorios en función de la etnia están prohibidos, en otros los propios gitanos no reconocen su identidad étnica por miedo a la represión o la discriminación; por otra parte, el carácter itinerante de muchas comunidades gitanas, especialmente en Europa central, impide que se realicen recuentos fiables; estos factores provocan que los datos que se recogen sólo permitan estimaciones a nivel global.

Actualmente, las estimaciones de población gitana en Europa varían de cuatro a diez millones, con las mayores densidades concentradas en el centro y el sudeste del continente (Mendizabal et al, 2011), aunque es ésta última cifra la que la Comisión Europea considera correcta. En la sociedad gitana tradicional, la unidad fundamental es el grupo y los grupos forman metagrupos. Se considera que viven en una sociedad de estructura cerrada, ya que se detectan tasas relativamente altas de consanguineidad y bajas tasas de mezcla con otras poblaciones (Assal et al, 1991). Parece que hayan existido distintos eventos del tipo "cuello de botella", tanto cuando dejaron India como cuando se dispersaron por Europa. Se observa una alta diversidad intragrupal (Gresham et al, 2001)

Resulta importante tener en consideración la reflexión de Gusmão y colaboradores (2008) con respecto a la identidad rrom hoy en día, fundamentalmente en regiones como España y Portugal, donde las tasas de

mestizaje son más altas, indicando que, a través de la adopción de un estilo de vida sociocultural, muchas personas se han “convertido en gitanos” y viceversa. Consideran que el término gitano, más que reflejar una entidad biológica o lingüística debería ser percibida como un constructo social en un ambiente en continuo cambio.

### **Perfil genético del pueblo gitano.**

El pueblo gitano constituye un aislado genético tradicionalmente clasificado como Caucásicos indoeuropeos (Charlesworth, 2005) habitualmente descrito como una población fundadora fragmentada compuesta de múltiples grupos; múltiples grupos aislados dentro de un aislado genético mayor (Jobling et al., 2004). Esta afirmación se explica por la combinación de un efecto fundador primario que afectó a toda la población gitana hace entre 32 y 40 generaciones y varios efectos fundadores secundarios, hace entre 16 y 25 generaciones, que dieron lugar a sub-grupos de aislados divergentes (Kalaydjieva et al., 2005; Morar et al., 2004). Por tanto, a pesar de compartir un origen común, las comunidades rrom son altamente heterogéneas debido a la deriva genética y a diferentes niveles de mestizaje con las poblaciones autóctonas de los lugares donde se asientan (Gusmão et al., 2008) y ningún grupo de gitanos puede ser representativo de otros o de la generalidad, lo que convierte al grupo en la unidad básica para las investigaciones sociales y genéticas (Fraser, 1998)

El estudio del pueblo Rrom, siendo un aislado genético transnacional, tiene un gran interés para la genética de poblaciones y en consecuencia para las disciplinas que pudieran beneficiarse de su caracterización, como es el caso de la epidemiología y la organización de la salud pública en todas aquellas regiones con presencia de esta minoría.

Durante la primera década del siglo XXI se identificaron varias enfermedades de herencia Mendeliana en la población gitana, que presenta un espectro muy particular de enfermedades genéticas, con algunas enfermedades prevalentes en el resto de la población prácticamente ausentes y otras específicas de los rrom, (Kalaydjieva et al, 2001b) pero se han llevado muy

pocos estudios acerca de las enfermedades neurodegenerativas en este grupo, como pueden ser la tauopatías. De hecho, y hasta el momento, la única población gitana analizada para los haplotipos MAPT es la residente en Hungría, donde se observó que la población gitana muestra frecuencias de la configuración H2 mayores que las esperadas para otros grupos asiáticos, donde prácticamente es inexistente (Evans et al, 2004; Almos et al, 2008)

El análisis de los marcadores uniparentales en gitanos europeos muestra una mezcla de linajes genéticos típicamente europeos y linajes que se encuentran exclusivamente en el subcontinente indio. Ello confirma el origen indio de los gitanos y sugiere una reciente mezcla con europeos no gitanos. Además, tanto el cromosoma Y como el DNA mitocondrial muestran una reducida diversidad genética, lo que sugiere que los romanís tienen un origen común en un grupo reducido de ancestros.

Los estudios publicados enfocados en la población gitana española apoyan la hipótesis de un origen común de este grupo y el resto de rrom europeos. En términos de distancia genética, los gitanos españoles se agrupan con los rrom checos y los grupos del norte de la India en vez de con las poblaciones caucásicas vecinas no gitanas (Ramal et al, 2000). Por otro lado, también cumple con la citada característica de ser una población fundadora en sí misma, al albergar una serie de diferencias específicas, como, por ejemplo, una mutación que causa anemia de Fanconi, originada hace menos de 600 años y que no está presente en otras comunidades rrom, como la húngara, alemana o irlandesa (Callén et al, 2005).

#### **Poblaciones estudiadas en este contexto: Gitanos vascos.**

Al contrario que la mayoría de los rrom, los gitanos de la Península Ibérica no hablan romaní. Tras su entrada en torno al siglo XV fueron perdiendo progresivamente su idioma original y actualmente hablan caló, que consiste básicamente en palabras de la lengua romaní insertas en el



entramado gramatical del castellano, mientras que los gitanos (gitanos portugueses), hablan principalmente portugués (Fraser, 1998).

Existen pocas referencias sobre ello, pero alrededor de mil gitanos (*Ethnologue*) residentes en algunas de las regiones españolas y francesas vascófonas - en concreto en Lapurdi, Zuberoa, Navarra, Gipuzkoa y Bizkaia - hablan *erromintxela*, una mezcla derivada del vocabulario del romanó Kalderash y la gramática vasca. A mediados de los años 90 del siglo pasado hubo un intento de investigación de este fenómeno sociolingüístico liderado por la asociación KALE DOR KAYIKO, Euskaltzaindia y el Departamento de Euskera de la Universidad del País Vasco, pero el trabajo se vio interrumpido por falta de financiación y apenas se encuentran un puñado de artículos en distintos medios.

*“Ez hizkuntza ez dialekto, Erromintxela pagodolekto bat da; ijitoen lexikoa euskararekin nahastean sortu zen hizkera. Ijitoen kontrako aurreiritzi guztiak desarmatzeko aukera paregabea”.*

“Ni lengua ni dialecto, el erromintxela es un pagodolecto; una forma de hablar que surgió al mezclar el léxico de los gitanos y el euskera. Es una oportunidad inmejorable para desarmar todos los prejuicios contra los gitanos”

Josune Muñoz, filóloga especializada en el fenómeno erromintxela para un artículo en el periódico Deia en 2012 con motivo de la feria de Durango. Entrevista realizada por Ana Araluzea.

En las primeras décadas del siglo XX, familias rrom provenientes principalmente de Castilla y Andalucía, pero también de otras comunidades autónomas de España, emigraron al País Vasco, siguiendo la misma tendencia que el resto de la población durante los años de industrialización.

Se estima que actualmente viven unos 13 000 gitanos en la comunidad autónoma vasca, con la mayor concentración en Bizkaia, unas 8 000 personas. De ellas la mitad vive en Bilbao y la otra mitad en la región del Gran Bilbao, representando menos de un 1% de la población del Gran Bilbao.

La región del Gran Bilbao es una comarca vizcaína perteneciente a la Comunidad Autónoma del País Vasco. Comprende, además del propio municipio de Bilbao, otros 21 municipios repartidos a ambos lados de la Ría del Nervión hasta su desembocadura.

Más del 75% de la población de Bizkaia (857.016 habitantes) vive en esta región (EUSTAT, 2018). Alrededor del 49% de los habitantes nacieron en el municipio, alrededor del 19% en otros pueblos de Bizkaia, el 2% en las otras dos provincias de la Comunidad Autónoma del País Vasco, un 21% en el resto de España (principalmente en la Comunidad Autónoma de Castilla y León) y algo más del 9% en otros países (EUSTAT, 2018).

La muestra analizada en este trabajo consta de individuos predominantemente nacidos en la provincia de Bizkaia, con unas características demográficas y reproductoras distintas de las de población no gitana del Gran Bilbao; por un lado, las familias Rom tienen mayor número de hijos y a edades más tempranas. Sus tasas de paro, nivel educativo y las características de sus viviendas indican que la población gitana estudiada tiene un bajo estatus socioeconómico, aunque parece que en los individuos más jóvenes esta situación se está revirtiendo. Son, en todo caso, gitanos que emplean el castellano y el caló, no el erromintxela.

## América.

América fue el último continente en ser poblado por el ser humano (Goebel et al, 2008). Durante la época de su poblamiento, el planeta se encontraba saliendo del último periodo glacial (hace 110 000 – 12 000 años). De hecho, el último máximo glacial (*last glacial maximum*, LGM) se extendió entre hace 23 000 y 19 000 años (Mix et al, 2001). Se cree que las placas de hielo de los glaciares de Norteamérica (Cordillera y Laurentino) cubrían la totalidad de Canadá y el Norte de EE. UU., conformando una barrera de entrada a América.

Con el final del LGM, las placas de hielo Cordillera y Laurentino habrían comenzado a retroceder, abriendo un paso libre de hielo al continente, el llamado **corredor interior**, situado en el valle del río Mackenzie al oeste de Canadá (Mandryk et al, 2004; Dyke, 2004) y que habría conectado Beringia con Norteamérica.

Existen numerosas evidencias de que este “puente de tierra” resultó fundamental para la migración desde el continente euroasiático al americano de un gran número de especies vegetales y animales. Actualmente parece descartada la posibilidad de que especies anteriores del género *Homo* hubieran podido acceder a América (Fagan, 2004; Goebel et al, 2008; Fernández-Gómez y Velasco-Ortiz, 2011).

Pedersen y colaboradores (2016) sostienen que el poblamiento del continente no pudo darse a través de Beringia y del citado corredor interior, ya que, según las conclusiones de su estudio a partir del análisis de polen y restos de mamíferos, este se habría abierto hace unos 12.500 años, más tarde que las primeras ocupaciones de Norteamérica documentadas. Aunque, migraciones posteriores sí pudieran haber usado esta ruta norte-sur, el corredor quedaría descartado para los primeros pobladores de América.

Alternativamente se ha propuesto que los primeros pobladores humanos del continente podrían haber llegado mediante el **corredor de la costa del Pacífico**, con el uso de pequeñas embarcaciones y navegando cerca de la

costa (Achilli et al, 2008; Fagundes et al, 2008; Goebel, 2008; Bodner et al, 2012). Esta hipótesis abre la posibilidad de que el poblamiento del continente sucediera al menos 1.000 años antes de lo previsto en caso de emplear el corredor interior (Mandryk et al, 2004; Dyke, 2004) e incluso que el corredor de la costa Pacífico fuera un acceso viable hace entre 17.000 y 15.000 años (Dixon, 2001).

Las evidencias más antiguas aceptadas mayoritariamente de presencia humana en América datan de hace unos 14.000 años (Dillehay, 1997). Las pruebas citadas implicarían que los primeros pobladores del continente americano no habrían sido los Clovis, cultura que tradicionalmente se ha considerado la primera de América (Waters et al, 2007).

La difusión desde el norte de América al resto del continente debió de ser rápida, tal y como se puede inferir de yacimientos y restos fósiles (Dillehay, 1997; Bodner *et al.* 2012; Roewer *et al.* 2013; Callaway, 2015).

Adicionalmente, existe cierta controversia en torno al número de grandes migraciones que se sucedieron desde Eurasia a América. Entre los numerosos trabajos publicados sobre este tema, destaca el de Reich et al (2012), que después de analizar 364.470 SNPs en 57 poblaciones americanas y 17 siberianas estiman que aproximadamente la mitad del patrimonio genético de las poblaciones eskimo aleutianas provendría de una segunda oleada (de la que provendría la cultura Thule) y que los Na Dene habrían recibido un 10% de sus genes de una tercera oleada desde Asia. En todo caso, confirman una ruta migratoria de Norte a Sur por la costa para la principal migración.

Más sorprendentes son algunos trabajos (Skoglund et al, 2015; Raghavan et al, 2015) que proponen una oleada secundaria proveniente del este de Asia, en la que vendrían individuos con genomas de origen melanesio y austronesio.

Algunos autores han propuesto contactos posteriores al cierre de la entrada a través de Beringia, migraciones menores y retromigraciones, entre poblaciones de ambos continentes para explicar la presencia de algunos haplogrupos minoritarios en poblaciones de Norteamérica pero a efectos generales se considera que una vez finalizada la glaciación, las poblaciones

americanas nativas se mantuvieron asiladas del mundo durante más de 10.000 años (Kitchen et al, 2008), durante los cuales se produjo una notable diversificación de las poblaciones en los diferentes entornos (estepario, selvático, insular, de altura, etc.) a la vez que aparecían numerosas culturas.

Así, salvo la comunicación ininterrumpida entre esquimales y paleoesquimales de Alaska y Siberia y una breve colonización vikinga de la costa de Canadá y Groenlandia, no se conocen pruebas que respalden un contacto transoceánico entre la América precolombina y el resto del mundo.

“El 12 de octubre de 1492, con la llegada de Colón a tierras de América, se inició una de las transformaciones culturales más grandes en la historia de la humanidad. La mezcla de las culturas amerindias, residentes siglos atrás en el suelo americano, con la recién llegada cultura hispánica, (ya de por sí una fusión de moros, gitanos, y pueblos íberos), y posteriormente, con la introducción de los africanos (traídos como esclavos para trabajar en las plantaciones de caña de azúcar y en la extracción de minerales), dio como resultado la rica diversidad étnica existente en toda Latinoamérica”

Colombia: Una nación multicultural. DANE, 2006.

El siguiente evento que supuso una gran transformación en el patrimonio genético actual americano fue la llegada de los colonos europeos y más tarde el tráfico de personas esclavizadas procedentes de África tras el llamado “descubrimiento” de América por Cristóbal Colón en 1492, si bien la llegada de la mayor parte de los inmigrantes ocurrió a partir del siglo XVI, especialmente durante su segunda mitad.

Motivados por la exploración de nuevos territorios y el descubrimiento de nuevas fuentes de materias primas con las que compensar los gastos militares debidos a las numerosas guerras y atender la demanda del mercado de telas, especias y otros productos, España y Portugal comenzaron la época de las colonias en el siglo XV, seguidas por Inglaterra, Francia y Países Bajos, principalmente a partir del siglo XVII.

Se estima que entre un 90 y un 95 por ciento de la población total amerindia murió en los primeros 130 años de colonización europea (Mann, 2006) por efecto de las enfermedades infecciosas, las guerras y los trabajos

forzados en agricultura y minería, por ejemplo. Los registros arqueológicos e históricos corroboran el gran impacto que tuvieron las prácticas citadas en los nativos americanos (Dobyns, 1993; Crosby, 2003 y 2004; Livi-Bacci, 2006).

Parece ser que fue la drástica reducción de la población originaria lo que provocó que las potencias europeas recurrieran al “reclutamiento” de africanos para reemplazar de manera forzada la mano de obra indígena. La primera región que actuó como fuente de personas esclavizadas fue Senegambia, en el Oeste de África, pero pronto el mercado no pudo ser satisfecho simplemente por las capturas en esta área por lo que durante los siguientes siglos se exploraron regiones costeras más al sur, incluyendo Congo, Angola y Mozambique. Los africanos cautivos eran enviados principalmente a Brasil y al Caribe para trabajar en plantaciones de azúcar, café, algodón, en minas, etc (Thomas, 1998; Salas et al, 2005).

El principal cuerpo de conocimiento del tráfico Atlántico de esclavos lo componen estudios realizados por historiadores y antropólogos culturales. Las investigaciones basadas en evidencias genéticas son escasas y relativamente recientes, a pesar de que ésta es una disciplina con una larga trayectoria aportando evidencias para esclarecer episodios históricos y prehistóricos (Salas et al, 2005).

Se sabe bastante poco sobre el impacto de la llegada de los africanos en la configuración genética de América. Una de las razones principales para ello es la falta de datos de las poblaciones africanas de origen (Salas et al, 2005). Las aportaciones principales provienen estudios con marcadores autosómicos (Sans, 2000) y de DNA mitocondrial (Salas et al, 2004).

No se conoce el número exacto de personas que se transportaron en calidad de esclavos a América, pero las estimaciones para el periodo de tiempo entre 1600 y 1850 varían entre los cinco millones (Curtin, 1969; Thomas, 1997) y los once y medio (Lovejoy, 1982).

Aunque la Corona Española reconoció como súbditos a los nativos de las tierras conquistadas y abolió la esclavitud de estos mediante las Leyes de Burgos (1512), los sistemas bajo los que eran tratados seguían constituyendo regímenes semi-esclavistas (Sánchez-Domingo, 2012).

Además de los cambios demográficos, este segundo evento masivo de colonización del continente produjo una serie de cambios que pueden detectarse en la configuración genética de los americanos actuales, encontrándose poblaciones con distinto grado de mestizaje entre amerindios, africanos y europeos debido al **flujo génico** entre los grupos étnicos implicados, fundamentalmente en las regiones controladas por España y Portugal (Carvajal-Carmona et al, 2000; Salzano y Bortolini, 2002; Bedoya et al, 2006; Salzano y Sans, 2014). En las colonias inglesas el flujo génico entre grupos fue mínimo ya que practicaron deportaciones masivas, aislando a los nativos en reservas, principalmente debido a que la mayor parte de inmigrantes británicos fueron calvinistas extremos o puritanos y a que llegaban, en mucha mayor proporción que los españoles, familias enteras. A grandes rasgos por tanto encontraríamos una baja tasa de mestizaje en el norte de América y casi exclusivo entre europeos y africanos mientras que en centro y sudamérica y en la parte meridional de norteamérica se caracteriza por una mezcla mucho más marcada entre todos los grupos humanos, especialmente indígenas y europeos (Sánchez-Albornoz, 1974; Bercovith, 1975; Olson, 1980; Degler, 1984, 1986; Jones, 1996; Salzano y Sans, 2014).

Adicionalmente, los análisis genéticos realizados a diferentes poblaciones americanas han revelado un curioso fenómeno ocurrido durante esta segunda migración masiva a América; en los marcadores del cromosoma Y se ha detectado una marcada influencia europea mientras que los haplogrupos amerindios son los predominantes en los estudios del DNA mitocondrial (Merriwether et al. 1997; Sans et al. 1997; Dipierri et al. 1998; Bortolini et al. 1999; Mesa et al. 2000; Alves-Silva et al. 2000; Green et al. 2000; Carvajal-

Carmona et al. 2003; Rojas et al. 2010; Marcheco-Teruel et al. 2014). Este **mestizaje diferencial dependiente del sexo** es debido a la asimetría demográfica de los inmigrantes europeos, principalmente entre portugueses y españoles, que eran fundamentalmente varones (Boyd-Bowman, 1976). El patrón predominante de las relaciones era de hombres europeos con mujeres nativas americanas y en menor medida africanas (Gómez-Pérez, 2015).

Además de los motivos religiosos o sociales que favorecieron un mayor o menor grado de mezcla entre las poblaciones implicadas, hubo otros factores que la modelaron. Las características de la región, incluida su riqueza en recursos, y la capacidad de los nuevos colonos para adaptarse a las mismas marcaron la densidad de los asentamientos y la tasa de mestizaje.

“Las regiones montañosas, los desiertos, las selvas o las latitudes extremas fueron regiones con poco impacto migratorio, permaneciendo en algunos casos prácticamente intactas hasta una época reciente; y es precisamente en estos lugares donde se ha observado que las poblaciones nativas generalmente conservan un alto grado de componente amerindio en su genoma. En cambio, regiones más accesibles como las llanuras, zonas costeras o riberas de los ríos experimentaron una mayor influencia colonizadora, resultando en la actualidad en poblaciones muy mestizadas, incluso con presencia mayoritaria europea”

L. GÓMEZ-PÉREZ, 2015

Un buen ejemplo de ello son las tres poblaciones sudamericanas estudiadas en este trabajo y que se describen a continuación, junto con el área que habitan.

## **Sudamérica.**

Sudamérica tiene una extensión total de 17.819.100 km<sup>2</sup>, lo que supone casi la mitad del continente americano y el 12% de las tierras emergidas. Alberga el 6% de la población del planeta, alrededor de 390 millones de habitantes, y algunas de sus ciudades están situadas entre las más pobladas



del mundo. Incluye una docena de países que se suelen agrupar en las regiones de Caribe Sudamericano, cuenca del Amazonas y Cono Sur. Algunos países pueden ser clasificados dentro de más de un grupo, como por ejemplo Colombia que forma parte de los países del Caribe Sudamericano y de la región de la cuenca del Amazonas.

Los océanos Atlántico y Pacífico delimitan su superficie por el este y el oeste respectivamente. Al sur queda el océano Ártico y al norte el mar Caribe y el istmo de Panamá por el que se conecta, a través de Centro América, con el norte del continente americano. Por su tercio norte cruza la línea del ecuador, de manera que la mayor parte del subcontinente queda en el hemisferio sur.

Topográficamente se distinguen tres grandes zonas. El escudo continental, formado por los macizos de Brasilia, Guayanés y Patagónico, las tierras bajas del interior, constituidas por los llanos del Orinoco, la llanura del Amazonas y la Chacopampeana o del Plata, y la Cordillera de los Andes, que es la más larga de la Tierra (7240 km) y recorre América del Sur paralela al océano Pacífico, desde el Atlántico hasta casi hundirse de nuevo en el Caribe. En la zona central se ensancha formando el altiplano que abarca parte de Argentina, Bolivia, Chile y Perú. Al sur de Colombia, en el llamado Nudo de los Pastos, se divide en la cordillera Occidental, que se extiende en dirección norte siguiendo la línea de la costa Pacífica y en la cordillera Central, de la que más al norte surge otro ramal, la cordillera Oriental que se adentra en Venezuela. A esta región se la conoce también como Andes Tropicales.

De manera general, las poblaciones indígenas sudamericanas se caracterizan por altos niveles de heterogeneidad entre las poblaciones, lo que resulta coherente teniendo en cuenta la diversidad lingüística y cultural de los grupos de amerindios del sur. Se ha argumentado que esta heterogeneidad es la consecuencia del contraste entre regiones de los patrones de deriva genética y del flujo génico debido a la presencia de barreras geográficas y al amplio espectro de biomas generado por la compleja orografía, entre otros factores (Malhi et al, 2002; Schurr y Sherry, 2004; Gómez-Pérez et al, 2013)

## Afrodescendientes del Chocó colombiano.

Chocó hace referencia a una región biogeográfica de 187.400 km<sup>2</sup> situada al norte de Sudamérica que abarca una estrecha franja desde la región sur de Panamá y una pequeña parte del caribe colombiano y que se extiende por la costa Pacífica de Colombia y Ecuador hasta el extremo noroeste de Perú.

Es un mosaico de planicies fluviales y marinas, llanuras aluviales, valles angostos y montañas con picos de hasta 4000 m de altitud en la zona colombiana que llegan a los 5000 m en Ecuador.

Es una de las regiones más biodiversas del planeta que ostenta también el título - junto con Cherrapunji, al noreste de la India - de ser la de mayor índice de pluviosidad, más de 9000 mm de precipitación anual (IDEAM), cifra que supera por mucho la de algunas de las localidades de la franja norte de España, típicamente conocidas en el ámbito estatal por su clima lluvioso, como las ciudades gallegas de Vigo (1909 mm), Santiago de Compostela (1886 mm) y Pontevedra (1691 mm) o las vascas Hondarribia (1738 mm), Donostia (1565 mm) o Bilbao (1195 mm) (Aemet).

Chocó es además uno de los 32 departamentos de Colombia, integrado totalmente dentro de la región biogeográfica del mismo nombre y de la región Pacífico del país, donde se desarrollaron las culturas indígenas Tumaco-La Tolita, y tierra principal de la cultura afrocolombiana. Limítrofe con Panamá, es el único departamento colombiano con costas en dos océanos.

Comprende un área de 44.530 km<sup>2</sup> cubierta de bosques húmedos y con escasa infraestructura de acceso y servicios públicos. Alberga varios parques nacionales. La selva ecuatorial del Darién, compartida con Panamá, es la frontera natural entre ambos países y ha ejercido históricamente como barrera, ya que no han existido vías terrestres de comunicación que atravesasen la zona.

Tiene una de las menores densidades poblacionales del país, 7,27 habitantes por kilómetro cuadrado, de los cuales el 82% son afrodescendientes, lo que supone la densidad más alta de afrocolombianos del

país, y casi el 13% indígenas de las etnias Emberá y Wounaan, que viven predominantemente en las zonas rurales, mientras que el 5% restante lo componen blancos y mestizos que habitan preferentemente las cabeceras municipales (DANE). Es, después de Brasil, la región americana con mayor población de ascendencia africana, lo cual es un dato llamativo teniendo en cuenta que los afrodescendientes en Colombia se encuentran en general en una proporción minoritaria de alrededor del 10% o menos, salvo en alguna de las regiones Atlántica y Pacífica, como la excepción que nos ocupa.

El origen de la porción mayoritaria de la población chocoana son los esclavos llevados desde África en torno al siglo XVII, fundamentalmente desde Mauritania, Mozambique, región del Golfo de Guinea, Sao Tomé, Congo y Angola, aunque como en la mayoría de América, el lugar preciso de proveniencia de estas personas no se conoce (Salas et al, 2005).

Por tanto, la composición de la población del Chocó, donde originariamente habitaban etnias indígenas, cambió dramáticamente con la llegada de las personas esclavizadas provenientes de África. Pero el grado de mestizaje con nativos americanos y europeos en esta región es prácticamente desconocido, salvo por un puñado de artículos como el de Salas y colaboradores (2005), que documentaron un componente genético minoritario de amerindios en su muestra de afrodescendientes del Chocó y una gran diversidad del componente africano, esencialmente idéntico a su fuente más probable en el Oeste de África.

En líneas generales, el mestizaje en afroamericanos muestra unas características básicas similares con peculiaridades propias de la región de residencia, lo que les convierte en una comunidad heterogénea difícil de categorizar (Bortolini et al. 1992, 1995, 1997, 1998, 1999; Castro de Guerra et al. 1996; Da Silva et al. 1999; Martinez-Labarga et al. 1999). Salvo excepciones (Sans et al. 2002), no se han encontrado asimetrías sexuales tan marcadas como en la población mestiza de indígenas sudamericanos y europeos.

Durante la época colonial los migrantes forzados africanos ingresaban de manera legal al país por Cartagena de Indias, desde donde se distribuían

no sólo a otros puntos del Nuevo Reino de Granada (la actual Colombia) sino a regiones de centro y sur América. De contrabando, la mercancía humana llegaba bien por el litoral Pacífico a Buenaventura, Charambirá y Gorgona, bien por el Atlántico a las costas de Riohacha, Santa Marta, Tolú y el Darién. Colombia fue por tanto uno de los países más influidos y relevantes por el “descubrimiento” de América; por su ubicación geográfica se la consideró la puerta de entrada a América del Sur.

Antes del año 1550 los asentamientos de población africana en territorio colombiano se limitaban a pequeños grupos en el litoral Caribe. A partir de ese momento los registros escritos indican que la mano de obra para la explotación minera era mayoritariamente africana, en sustitución de los indígenas que venían realizando esas labores, debido a (1) su descenso poblacional y a (2) las disposiciones de la Corona en cuanto a su protección. Los esclavos fueron también incorporados a otros trabajos como la agricultura, la ganadería, la artesanía y el servicio doméstico. Fueron también objeto de operaciones de inversión, alquileres de fuerza de trabajo, permutas, trueques, etc. (Escobar, 2005).

Desde la llegada de las primeras personas esclavizadas y durante los siglos que duró esta práctica uno de los problemas principales que tuvo que afrontar la sociedad esclavista fue la huida de esclavos y su paulatina organización en cabildos y en palenques.

“En 1530 (...) los negros fugitivos incendiaron Santa Marta; hacia 1533, un buen número de esclavos traídos por el fundador de Cartagena huyó a las zonas montañosas de la provincia; en 1556 se produjo una importante rebelión de esclavos en Popayán; en 1598 se presentó una sublevación de esclavos en las minas de Zaragoza, matando a dueños y fortificándose en palenques; a finales del siglo, el gobernador de Cartagena proponía fórmulas para financiar la persecución y búsqueda, a través de cuadrilleros y la Santa Hermandad, de los “negros cimarrones que con la ocasión de los muchos, montes y aspereza de montañas crecen cada día” y para entonces ya se habían organizado los célebres palenque de la Matuna y San Basilio” (Palacios Preciado, 1989).

La palabra cimarrón, que se empleaba tradicionalmente para nombrar todo aquello que era silvestre o salvaje, comenzó a usarse primero para referirse al ganado doméstico que, escapada a las montañas en la isla de La Española, la actual República Dominicana, después para denominar a los esclavos indígenas huidos y finalmente, a punto de comenzar la década de 1540, este término ya aludía principalmente a los fugitivos africanos. La voz inglesa *maroon* y la francesa *marron* provienen de esta acepción de cimarrón.

Durante los siglos XVII y especialmente el XVIII, numerosas rebeliones y huidas de esclavos dieron lugar a multitud de palenques y comunidades de africanos y afrodescendientes fugitivos. Los palenques eran comunidades agrícolas ubicadas en lugares de difícil acceso y relativamente fortificadas, protegidas por empalizadas y fosos, vigiladas y defendidas por gente equipada con arcos, flechas, armas de fuego... Surgieron pues de las pequeñas bandas de cimarrones que se reunían en las montañas y que, tomando conciencia de grupo, se convirtieron en “apalencados” y conformaron los ya citados poblados autónomos. Albergaron gentes de diversos orígenes y su organización política, militar y social solía basarse en una combinación de las costumbres africanas y el modelo colonial.

“En los reportes del gobernador de Cartagena, Gerónimo de Suazo, se menciona que cuando los soldados españoles lograban hallarlos para destruirlos [los palenques], se encontraban con aldeas bien dispuestas en las que predominaban cultivos de maíz, gallinas y yuca” (Palacios Preciado, 1989).

Cabildos, audiencias y gobernadores establecieron penas severas para impedir y combatir la fuga y el cimarronaje. Los dueños de esclavos exigían la aplicación de castigos severos, pero tratar de exterminar los palenques y acabar con la búsqueda de libertad de las personas esclavizadas era tremendamente costoso y el precio del esclavo cimarrón era considerablemente inferior.

Este tipo de asentamientos existieron por todo el país, la mayoría localizados en la parte norte y caribeña de Colombia. Desarrollaron culturas y

lenguas propias. Entre las lenguas registradas en el país, el palenquero, la lengua criolla de una de estas poblaciones, San Basilio de Palenque, considerado el primer pueblo libre de América, fue declarada obra maestra del patrimonio oral e inmaterial de la humanidad por la UNESCO.

La abolición de la esclavitud en Colombia fue un proceso largo con muchos altibajos e inconsistencias. La discusión sobre la abolición se inició y desarrollo durante la Guerra de la Independencia por lo que se convirtió, más que en un tema de Derechos Humanos en una consigna política para ganar adeptos, prometiendo por ejemplo la libertad a aquellos que se unieran a los respectivos ejércitos.

La primera vez que se legisló en Colombia sobre la libertad de las personas en esclavitud fue en la Constitución de Cartagena de 1812. Debido a la reconquista española de 1815 esta ley no llegó a concretarse, aunque el comandante del ejército español, Pablo Morillo, otorgó la libertad a aquellos esclavizados que colaboraran con su causa.

Una vez lograda la independencia, en el Congreso de Cúcuta de 1821 se debatió uno de los aspectos más polémicos sobre la abolición de la esclavitud, la Ley de Libertad de Vientres que liberaba a los hijos nacidos de esclavas a partir del día de la publicación de la Ley, teniendo éstos que trabajar para los amos como compensación a los gastos en los que éstos incurrirían ya que estaban obligados a educarlos y mantenerlos hasta los dieciocho años.

**Artículo 1º.** Serán libres los hijos de las esclavas que nazcan desde el día de la publicación de esta ley en las capitales de provincia, y como tales se inscribirán sus nombres en los registros cívicos de las municipalidades y en los libros parroquiales.

**Artículo 2º.** Los dueños de esclavas tendrán la obligación precisa de educar, vestir y alimentar a los hijos de éstas, que nazcan desde el día de la publicación de la ley; pero ellos, en recompensa, deberán indemnizar a los amos de sus madres los gastos impendidos en su crianza con sus obras y servicios, que les prestarán hasta la edad diez y ocho años cumplidos.

**Artículo 3º.** Si antes de cumplir la edad señalada quisieren los padres, los parientes u otros extraños sacar al niño o joven, hijo de esclava, del poder del amo de su madre, pagarán a éste

lo que regule justo por los alimentos que le ha suministrado, lo que se verificará por un avenimiento particular o por el prudente arbitrio del juez.

**Artículo 4º.** Cuando llegue el caso de que por haber cumplido los diez y ocho años salgan los jóvenes del poder de los amos de sus madres, será una obligación de éstos el informar a la Junta de que se hablará después, sobre la conducta y procedimientos de los expresados jóvenes, a fin de que promueva con el Gobierno el que se les destine a oficios y profesiones útiles.

[Archivo general de la Nación.](#)

Le Ley se aprobó cuando se protegió el derecho a la propiedad privada de los esclavistas, expresado en la consigna “ser generosos con los esclavos sin dejar de serlo con los amos”. Lejos de una abolición completa de la esclavitud prometida durante el proceso de independencia, esta Ley aplazaba la libertad al año 1839 y sólo para un reducido número de personas.

El retraso en la aprobación de la citada Ley permitió la reapertura del tráfico de personas en esclavitud unos años después y provocó la frustración de la población, por lo que durante el siglo XIX el cimarronaje alcanzó su máxima intensidad.

El incremento de esta actividad, la presión del imperio inglés para acabar con la esclavitud, el ascenso del liberalismo y la entrada en escena de la Generación del 48, así llamada por la influencia recibida de la Revolución Francesa y que consideraban la Constitución de Cúcuta como un producto inacabado, sentaron las bases para que la abolición apareciera de nuevo como consigna política.

Aunque en 1850 el gobierno aún trataba de retrasar la abolición, los factores arriba mencionados unido a la presión que se ejerció desde el Congreso, la prensa, las sociedades democráticas y los propios esclavos llevaron finalmente a que el Gobierno liderado por José Hilario López se decantara por la abolición absoluta.

El debate parlamentario, que tuvo lugar entre marzo y mayo de 1851, se zanjó con la defensa de una fórmula de abolición con indemnización para los amos por parte del Estado. La Ley fue aprobada el 21 de julio de 1851 y entró en vigencia el 1 de enero de 1852, fecha en la que se liberaron unas 16 000 personas, después de un largo proceso de 40 años.

La integración de la población afrodescendiente ha sido desigual, siguen sufriendo ciertos estigmas y el racismo / clasismo aún no se ha erradicado del país, como puede comprobarse, por ejemplo, mediante un repaso de la prensa colombiana.

Se diferencia cuatro grupos principales de población afrocolombiana; (1) la comunidad chocoana que nos ocupa, (2) los raizales del Archipiélago de San Andrés, Providencia y Santa Catalina, con una fuerte identidad caribeña y rasgos socioculturales y lingüísticos claramente diferenciados, (3) la comunidad de San Basilio de Palenque ya mencionada que alcanzó su libertad en 1603 y que ha permanecido en un relativo aislamiento hasta hace poco, y (4) la población que reside en las cabeceras municipales y en grandes ciudades (Escobar, 2005; DANE, 2016).

El Chocó ha vivido tradicionalmente una situación de marginalidad económica (debido a las rutas limitadas de acceso a la región y al abandono político al que ha estado sometido, sufriendo un régimen cíclico de bonanza-recesión en función de la demanda de materias primas como el oro, el caucho o la madrea y en menor medida productos agrícolas, como el banano, plátano, coco y el arroz, y pesqueros, como el pescado seco (Whitten, 1974).

Actualmente en la región del Chocó colombiano las actividades económicas principales siguen basándose en la explotación de recursos naturales; la actividad depende en gran parte de la zona. En la franja costera los recursos minerales son escasos y predominan agricultura y pesca. En el interior se encuentra oro y platino, por lo que la minería se intercala con la agricultura y la pesca, entre otras actividades. Los cultivos principales son el



maíz, el plátano y el arroz. Recientemente se ha unido la industria maderera, habitualmente dirigida al mercado nacional y mundial.

La minería puede encontrarse en una diversa variedad de modalidades, desde el trabajo independiente con una batea en las playas de los ríos a agrupaciones más o menos profesionalizadas con aparatos especializados para esta actividad.

La agricultura tiene una importancia económica mayor a la minería y está ampliamente difundida en la región, aunque muy poco tecnificada. Además de los cultivos básicos señalados también se recoge yuca, caña de azúcar, chontaduro, coco, borjón y otras frutas. Una parte de los cultivos se destina a la subsistencia y otra se comercializa a nivel regional o nacional.

A pesar de la riqueza del Chocó en recursos naturales, su extracción se realiza principalmente mediante explotación directa por una compañía no local, con su propia maquinaria y trabajadores, que en ocasiones contrata a lugareños para los trabajos más pesados. Por esta vía empresas nacionales y multinacionales han obtenido metales preciosos y maderas tropicales fundamentalmente (Whitten, 1974).

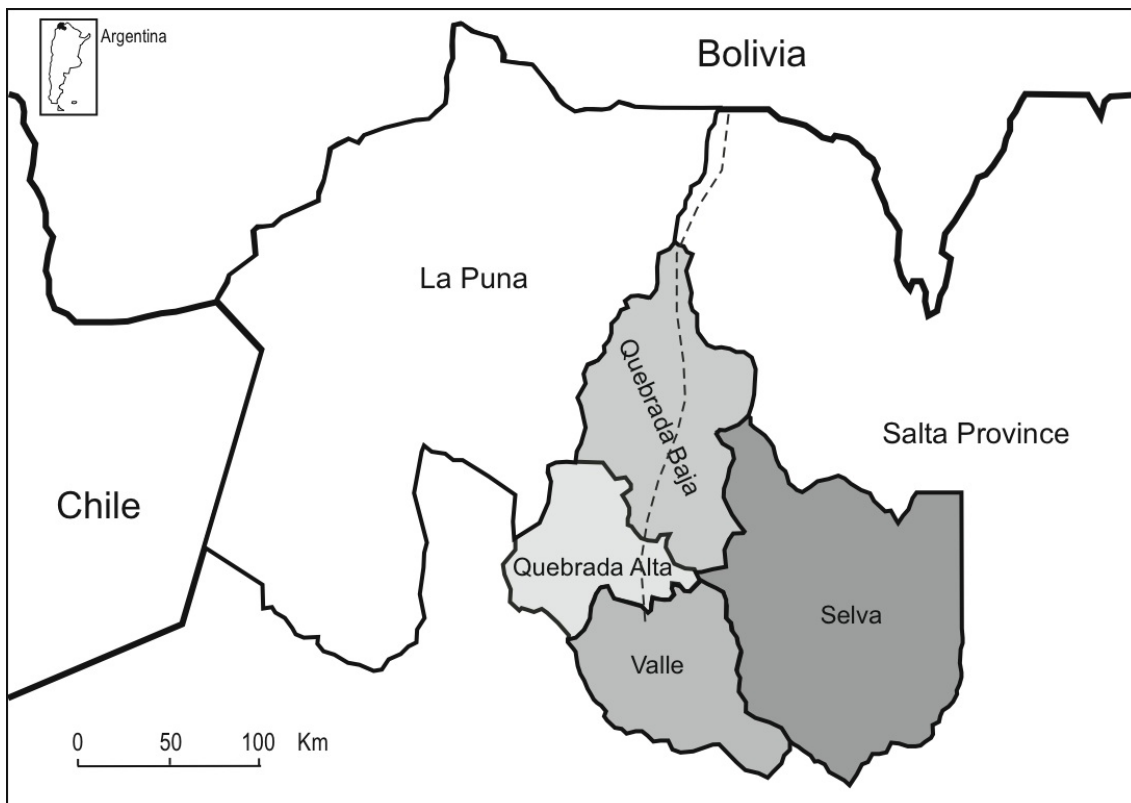
## **Mestizaje diferencial en Jujuy, Argentina.**

Argentina forma parte del denominado Cono Sur y está ubicada en el extremo sureste de Sudamérica. Su territorio continental se extiende entre la cordillera de los Andes al oeste, que ejerce de límite natural entre Chile y Argentina, y el océano Atlántico al este. Su frontera sur queda definida por el mar Antártico y la norte, de oeste a este, por Bolivia, Paraguay, Brasil y Uruguay.

Su orografía se caracteriza por una disminución de la altitud desde la zona montañosa del oeste hacia los llanos del este.

Jujuy es una de las 23 provincias argentinas. Ocupa 53.219 km<sup>2</sup> al noroeste del país, limita al norte con Bolivia, al oeste con Chile mediante los Andes y al este y al sur con otra provincia argentina, Salta. De acuerdo con el censo demográfico (INDEC, 2001) tiene una población de más 608.000 habitantes.

Como el resto del país, su altitud disminuye de oeste a este, desde los 6000 msnm de la cordillera de los Andes, pasando por el altiplano o Puna, de alrededor de 3.300 msnm de media, y la zona de las quebradas (2.500 msnm) antes de llegar a la región de valle (1.200 msnm), donde se ubica su capital, San Salvador de Jujuy, donde habita aproximadamente un tercio de la población de la provincia (INDEC, 2010)



**Figura I.9:** Mapa de la provincia Jujuy donde se especifican las principales comarcas y, con línea discontinua, la Quebrada del Humahuca (Gómez-Pérez, 2015).

Jujuy es por tanto una provincia de relieve eminentemente montañoso, con importantes variaciones altimétricas que favorecen la aparición de distintos tipos climáticos y sus hábitats correspondientes, y que han influido en la

distribución de las poblaciones humanas y en sus relaciones (Carnese et al, 2010; Gómez-Pérez, 2015).

Debido a las razones expuestas, la provincia de Jujuy se estructura en tres áreas geográfico-ecológicas diferentes conectadas entre sí por estrechos pasillos o desfiladeros, cuyo ejemplo más característico e históricamente relevante como se verá más adelante es la Quebrada del Humahuaca, de 170km de longitud y que recorre gran parte de la provincia en dirección norte sur de forma ascendente. Las áreas geográfico-ecológicas son (1) el valle subtropical, con una altitud media de 1 200 msnm y que se corresponde con la región de valle previamente descrita; (2) el área de la pre-Puna, que incluye regiones situadas entre 2 000 y 2 500 msnm y se corresponde con la zona de las quebradas y (3) la Puna de Jujuy, a más de 3 000 msnm (Gómez-Pérez, 2015).

La presencia humana en Jujuy se justifica en parte por su situación geográfica estratégica dado que comunica las tierras altas andinas con las vastas llanuras sudamericanas (Gómez-Pérez et al, 2011). Jujuy se halla en la región que, junto con la ruta de la costa, habría sido una de las principales vías para los primeros pobladores del continente en su camino hacia la conquista del extremo sur del continente (Rothhammer y Dillehay, 2009). Posteriormente, la Quebrada de Humahuaca se convertiría en uno de los caminos principales del Imperio Inca y más tarde, durante la época colonial, por Jujuy transcurría la vía que unía las minas de plata del Potosí en Bolivia con los puertos del Río de la Plata (Dipierri et al, 2000; Mendisco et al, 2014).

Los primeros asentamientos de cazadores-recolectores en la zona datan de al menos entre 10 600 y 8 400 años, teniendo en cuenta las fechas más tempranas de yacimientos como los de Hornillos-2 en la Puna jujeña (Yacobaccio et al, 2008), Inca Cueva 4 o Huachichocana III (Aschero, 2000). Algunos autores apuntan a que la presencia humana en la región podría remontarse hasta hace 11 000 años, cuando los primeros grupos de cazadores-recolectores habrían colonizado la parte meridional del continente (Tarrago, 2000; Carnese et al, 2010). Otra propuesta ha sido que grupos

humanos procedentes de las regiones andinas de Perú y Bolivia se habrían establecido en los territorios de Jujuy hace entre 6 000 y 8 000 años (Rothhammer et al, 1984; Rothhammer y Silva, 1992).

La ocupación pre-colombina del noroccidente de Argentina sigue un patrón común con el resto de comunidades del centro y sur de los Andes. A los asentamientos iniciales descritos, les sigue el periodo arcaico, entre los años 9000 y 1000 antes de nuestra era, caracterizado por la intensificación de la subsistencia que conlleva a la domesticación (Yacobaccio, 2001; Lagiglia, 2001). Posteriormente se establecen los modos de vida sedentarios basados en la agricultura y el pastoreo y organización política tribal, lo que se denomina periodo Formativo y abarca entre el final del periodo anterior y el siglo IX en Humahuaca (Albeck, 2000; Scattolin, 2006). Durante los siglos IX al XII aparecen las primeras evidencias de jerarquías sociales (Nielsen, 2001) y después, entre los siglos XIII y la llegada de los incas a la región en 1430, se diferencia una etapa en la que surgen pre-estados de políticas más complejas. Finalmente, el periodo inca en la región abarcaría desde el año 1430 a 1536 (Nielsen, 2001, 2005).

El imperio incaico, surgido en Perú a finales del siglo XII o principios del XIII, se expandió por el noroeste de Argentina al menos medio siglo antes de la llegada de los conquistadores españoles. Fue la civilización americana precolombina más grande – se estima que la componían unos 12 millones de personas – y de mayor complejidad, cuyo centro político, administrativo y militar se ubicaba en Cusco, Perú. Desde 1438 a 1533 conquistaron o asimilaron de manera pacífica una gran parte del occidente sudamericano, incluyendo Perú, un área amplia de Ecuador, el centro y el sur de Bolivia, el sur de Colombia, el noroccidente de Argentina y el norte de Chile (Gómez-Carvalla et al, 2015).

Esta civilización dejó una serie de construcciones de diversa índole en el territorio argentino, desde centros agrícolas y de producción de tejidos con sus asentamientos correspondientes a depósitos (*collcas*), postas (*tambos*), fortalezas (*pucarás*), santuarios en lo alto de las montañas donde se han

encontrado restos momificados de sacrificios humanos (Albeck, 2010, 2011). Uno de los más importantes es el llamado Camino del Inca.

El Camino del Inca o Camino Andino - y a cuya tronco principal se le denominaba *Qhapac Ñan* (Albeck, 2016) - era una vasta e intrincada red vial que unía todos los puntos del imperio incaico con su capital en Cusco, Perú, permitiendo la gobernabilidad y el control imperial al facilitar una comunicación fluida entre los diferentes puntos del imperio, desde el sur de Colombia pasando por Ecuador, Perú, Bolivia y llegando hasta Argentina y Chile (Gómez-Carvalla et al, 2015). La longitud total de estas vías usadas por peatones y animales de carga - llamas, no tenían carros ni caballos - era de 40 000 km. Fue construido - se cree que en menos de un siglo - siguiendo unas medidas determinadas y uniformes en todo el recorrido en cuanto al ancho, la superficie o la distancia entre postas. Discurre por la complicada orografía andina, cruzando ríos, valles abruptos y altos pasos entre las montañas.

Algunas de las construcciones mencionadas anteriormente estaban estrechamente relacionadas con esta red viaria como los *tambos*, que servían de alojamiento para los funcionarios incaicos y para los mensajeros (*chasquis*) que circulaban por ellos.

Actualmente sigue comunicando personas y pueblos, especialmente en las regiones más remotas, y ha sido incluido en la lista de Patrimonio Mundial de la UNESCO.

La provincia de Jujuy fue pues uno de los territorios por los que pasaba esta importante vía de comunicación, que se valió de los accidentes geográficos como la Quebrada del Humahuca para su trazado. La cabecera de esta región fue Tilcara, ubicada en este pasillo natural y donde se halla el *pucará* del mismo nombre, una fortaleza construida por los tilcara, etnia perteneciente a los omaguacas (Waddington, 2016). El *pucará* de Tilcara albergaba varias viviendas, corrales, una necrópolis y un lugar para las ceremonias religiosas. Su ocupación abarca desde el siglo XIII al XV, pero las evidencias predominantes pertenecen a la civilización Inca y datan de la

ventana de tiempo entre 1430 y 1536 (Belotti López de Medina, 2016). Cuando el imperio Inca llegó a Jujuy esta región pasó a formar parte del Collasuyo, la provincia meridional del mundo incaico. Unos cincuenta años después de su establecimiento llegaron los colonizadores españoles.

El primer europeo en transitar por el Camino del Inca en territorio argentino fue Diego de Almagro, socio de Francisco Pizarro en la conquista de Perú; partió desde Cusco en 1535 atravesando el norte de Argentina hasta llegar a Chile circulando por estas vías en su momento de mayor auge (Vitry, 2004, 2007).

Antes de la llegada casi simultánea de los conquistadores incas y de los españoles, Jujuy estaba habitado por numerosos grupos autóctonos como atacamas, casabindos, chicas, chiriguano, cochinos, diaguitas, jujuyes, maticos, mocovies, omaguacas, purmamarca, tilcaras o yalas entre otros (Dipierrri et al, 2000; Alfaro, 2010). El estudio de restos de DNA antiguo destaca una composición heterogénea en la población de Jujuy con un posible origen común andino (Mendisco et al, 2014). En esta zona se ha registrado una variedad de lenguas amerindias pertenecientes a familias lingüísticas tan diversas como la aimara, la maticaco, la quechua o la tupí. Algunos ejemplos de estas lenguas son el aimara, chiripá, guaraní, quechua sudboliviano y wichí lhamtés vejiz (Lewis et al, 2014).

De la misma manera que sucedió en el resto del continente, la llegada de los colonos europeos y posteriormente de inmigrantes forzados africanos y el establecimiento de la colonia española durante los siglos XVI al XIX, provocaron una serie de cambios cuyo impacto aún puede observarse, por ejemplo, en la diversidad genética de las poblaciones humanas que habitan Jujuy (Gómez-Pérez et al, 2011).

Los primeros esclavos africanos fueron introducidos en Argentina por el Río de la Plata en torno a 1588, capturados principalmente en las zonas costeras de los territorios actuales de Angola, República democrática del Congo, Guinea y Congo (Gómez-Pérez et al, 2011). Según un censo poblacional de 1778, Jujuy fue una de las provincias argentinas con menor densidad de

personas africanas, en torno al 13% de la población, cuya presencia iba menguando según los asentamientos aumentaban su altitud (Gomes, 2006); una tendencia observada también para los inmigrantes europeos (Martinez-Sarasola, 1992). Así pues, los autóctonos eran desplazados cada vez más a regiones de mayor altitud mientras que esclavos y colonizadores se concentraban principalmente en los bosques subtropicales de las tierras bajas de Jujuy (Martinez-Sarasola, 1992).

Los motivos que se barajan como principales causas de esta repartición no uniforme son (1) la dificultad de adaptación de africanos y europeos a altitudes elevadas y (2) las características de las tierras altas, más áridas y poco adecuadas para la agricultura y ganadería, con la excepción de la cría de llamas (Gómez-Pérez, 2013). A esta actividad y a sus oficios derivados (hilanderos y tejedores) se dedicaban fundamentalmente los integrantes de las poblaciones de la Puna, con una altitud media de 3 500 msnm, según un censo de 1859 (Gil, 2000-2001).

Así pues, las poblaciones situadas a mayor altitud tuvieron menos oportunidades de mestizaje (Martínez-Sarasola, 1992; Gomes, 2006 a, b); de hecho, varios estudios han reportado un flujo génico diferencial y distintos ratios de mestizaje en las poblaciones de Jujuy en función de la altitud (Dipierri et al, 1998, 2000; Gómez-Pérez, 2011). Hoy en día también los nativos americanos son el grupo humano predominante en las regiones de mayor altitud de Jujuy, donde el componente genético amerindio es el mayoritario (Dipierri et al, 2000). Adicionalmente y de manera más marcada en las áreas de difícil acceso, Jujuy ha sido tradicionalmente un área poco poblada, donde las consecuencias de la deriva genética pueden haberse visto exacerbadas (Gómez-Pérez, 2011), favoreciendo, junto con el impacto demográfico y los movimientos poblacionales resultantes de la llegada de los europeos la pérdida de algunas variantes genéticas de origen amerindio (Mendisco et al, 2014).

## La etnia Waorani, Ecuador.

Ecuador es un país situado en el noroccidente de América del Sur. Su extensión, incluyendo las islas Galápagos es de 238.561 km<sup>2</sup> y en el último censo (INEC, 2014) se contabilizaron casi 16 millones de habitantes.

Puede subdividirse en cuatro áreas geográficas/ecológicas, una insular, es decir, las Islas Galápagos, y tres continentales que discurren paralelas unas otras y a la costa del Pacífico. Desde el litoral y hacia el interior se sitúa primero la región costera, que coincide con la región biogeográfica del Chocó; la región interandina, caracterizada por elevaciones montañosas, volcanes y nevados con una altitud media de 4000 metros y la región oriental o Amazónica, limitada al oeste por la cordillera de los Andes y al este por las fronteras de Colombia y Perú.

Es en ésta última zona donde se ubica el Parque Nacional de Yasuní, compartido por las provincias de Pastaza y Orellana, un territorio principalmente cubierto por selva y con una densidad demográfica relativamente baja. Colonos e indígenas se dedican a la pesca fluvial, la caza y la recolección de cultivos tradicionales (yuca, banano y café).

El parque Nacional de Yasuní cubre casi 10 000 m<sup>2</sup>, entre el río Napo y el Curaray, en plena cuenca Amazónica, incluyendo también parte del territorio del pueblo Waorani.





**Figura I.10:** Mapa de Ecuador, donde se resalta el área que abarcan el Parque Nacional Yasuni y el territorio Waorani.

Los Waorani, en su grafía inglesa o Huaorani en la española, y conocidos también como Sabela, Auishiri o Waodani, son una de las trece nacionalidades indígenas que coexisten en Ecuador (Baeta, 2011; Gómez-Pérez, 2015). Los pueblos indígenas se caracterizan por su autodefinición, su autonomía organizativa y cultural, y por su política propia. Se originaron y constituyeron miles de años antes de la invasión europea pero actualmente la mayoría se encuentran en proceso de retroceso y extinción (Baeta, 2011).

Como prácticamente para la totalidad de América, el origen de los primeros pobladores de esta zona y las migraciones tempranas que se dieron es poco conocida. Evidencias arqueológicas apuntan a que hace unos 11 000 años la región ya estaría habitada (Dillehay et al, 1992), que se habría poblado vía el Callejón Interandino (o Avenida de los Volcanes, como bautizó a una parte de estos valles el geógrafo y científico alemán Alexander von Humbolt en 1892) y que desde allí estos grupos podrían haber hecho pequeñas incursiones a otros territorios del actual Ecuador (Dillehay, 1999).

Debido a la falta de exploraciones sistémicas de la Amazonía, hay pocas pruebas respecto a los asentamientos en esta región. Sí que existen vestigios que demuestran cierta ocupación de la costa ecuatoriana, principalmente de la península de Santa Elena, donde habitaban pequeños grupos de cazadores-recolectores (Stothert et al, 2003)

Se han descrito numerosas culturas en distintas regiones de Ecuador, como las Machalilla, Valdivia, Chorrera en la costa; Cotacollao y Chimba en la sierra; y Pastaza y Chiguaza en la región oriental. De economía arcaica y tecnología incipiente durante el primer periodo, implementaron nuevas formas de agricultura y desarrollaron la cerámica dando lugar posteriormente a los cacicazgos o señoríos étnicos a nivel regional en ciudades ubicadas principalmente en la costa y en la sierra (Baeta, 2011).

Se considera que la franja costera Pacífica podría haber representado una vía de comunicación para numerosos pueblos americanos, con la región ecuatoriana como importante punto de convergencia con grupos mesoamericanos (Lathrap y Marcos, 1975).

Como se ha comentado en el apartado anterior, con la llegada de los Incas en el siglo XV, el sur de Ecuador pasó a ser parte de este imperio, que se originó en Perú y en su periodo de máxima expansión, en torno a 1532, llegó a ocupar territorios de las actuales Colombia, Ecuador, Bolivia, Chile y Argentina. Se produjeron en consecuencia importantes cambios demográficos

debidos a los enfrentamientos, los flujos migratorios y la repoblación de las regiones anexadas.

Dos son las causas principales a las que se les achaca la caída del Imperio Inca, (1) la confrontación por el trono entre los hijos de Huayna Cápac, Huáscar y Atahualpa, que derivó en una guerra civil de la que el segundo salió vencedor en 1532 pero que debilitó profundamente el incanato y (2) la llegada de las tropas españolas al mando de Francisco Pizarro, que capturaron a Atahualpa y lo ejecutaron en 1533. Aunque los conocidos como Incas de Vilcabamba se rebelaron contra la Corona Española, se considera que el Imperio Inca como tal desapareció con la muerte de Atahualpa (Gómez-Carballa et al, 2015). El cabecilla de la rebelión y último líder inca, Túpac Amaru I, fue capturado y decapitado en 1572.

Como en el resto del continente la llegada de los colonizadores europeos supuso un impacto radical en los pueblos nativos y en sus sistemas de organización. Como se ha expuesto anteriormente, el conflicto bélico, la propagación de nuevas enfermedades y los trabajos forzados produjeron la reducción del número de pobladores originarios llegando a acabar con algunas de las culturas existentes y las relaciones entre amerindios e inmigrantes europeos y posteriormente con esclavos africanos supuso la creación de nuevos grupos étnicos, en los que actualmente está basada la estructura étnica y cultural del país.

Unas pocas etnias, en cambio, han permanecido relativamente ajenas al intercambio génico y cultural con europeos y africanos hasta hace relativamente poco tiempo. Una de ellas es la población que nos ocupa, los Huaorani.

El primer contacto pacífico entre occidentales y los Huaorani se produjo en 1958 (Wallis, 1973). Hasta ese momento habían practicado un modo de vida básicamente cazador-horticultor; se alimentaban de pequeños cultivos en

los claros del bosque, de plantas silvestres y de algunos animales, como grandes monos y pecaríes. Actualmente se les sigue considerando el último pueblo amerindio seminómada de cazadores recolectores habitando Ecuador; cambian sus asentamientos en función de las épocas de cosechas, aunque actualmente están adoptando progresivamente el sedentarismo dejando atrás su modo de vida tradicional (Rival, 2002; Papworth et al, 2013).

Tradicionalmente, la unidad básica de organización social de esta etnia ha sido el *nanicabo*, una agrupación doméstica autosuficiente y autónoma de entre seis y diez familias. A la unidad territorial compuesta por la unión de varios *nanicabos* se la conoce como *huaomoni* y tienen un marcado carácter endogámico. La mayoría de las comunidades Huaorani están conformadas por dos o tres *huaomonis* de diversa procedencia (Beckerman et al, 2009; González-Andrade et al, 2009)

El origen étnico y lingüístico de los Huaorani continúa siendo fuente de especulación. La construcción lingüística del *Huao Terero*, el idioma hablado por los Huaoranis, se considera única, sin similitudes con la fonología o estructuras de otras familias lingüísticas de la Amazonía, por lo que el *Huao Terero* permanece aislado y sin clasificar (Peeke, 1973). En cuanto a la procedencia de esta etnia se ha especulado que podría derivar de grupos de origen tupi-guaraní, como omaguas, yameos y oas, que se dispersaron desde Brasil hacia el Pacífico y se establecieron en la selva amazónica a lo largo del río Napo entre los siglos X y XVI al encontrarse con la barrera natural que es la Cordillera de los Andes (Cabodevilla, 1994).

La cuenca occidental del Amazonas, donde habitan los Huaorani, es una región conocida por los episodios violentos protagonizados por sus habitantes. Los Huaoranis no son una excepción a este tipo de comportamientos; han sido descritos como una de las sociedades más belicosas del mundo debido a la alta tasa de homicidios registrados con motivaciones tales como disputas sobre arreglos matrimoniales, represalias por homicidios anteriores, acusaciones de brujería, etc. (Yost, 1981; Beckerman et

al, 2009). La violencia y la venganza determinan de tal modo la mortalidad en este pueblo que se ha estimado que un 60% de las muertes de adultos son causadas por acciones de guerra y de ellas un 44% de carácter interno (Yost, 1981). Ante estos hechos resulta creíble que la agresividad y beligerancia de los Huaorani haya impedido su absorción por otros pueblos.

Como consecuencia de lo expuesto, los Huaorani son una población homogénea con un alto grado de endogamia (Cardoso et al. 2008, 2012; González-Andrade et al. 2009; Baeta et al. 2009, 2013; Baeta, 2011). Existe un predominio de matrimonios entre primos (Robarchek y Robarchek, 1992; Beckerman et al. 2009) y en las ocasiones en las que se da un desequilibrio demográfico debido a las guerras u otros eventos catastróficos o en alguna otra situación especial, se permite la poligamia, principalmente la poliginia.

Se estima que unos 3000 Huaorani se distribuyen en 24 comunidades (Baeta, 2011) aunque otros datos apuntan a un número menor, de unos 2000 individuos (Lewis et al, 2014). Dentro de la etnia Huaorani existen varios grupos diferenciados. Algunos de ellos, como los Tagaeri o los Taromenani evitan todo contacto con el exterior (Baeta, 2011). Habitan en una reserva étnica que comprende una pequeña parte del territorio tradicional de este pueblo, que se extendía a lo largo de un área de dos millones de hectáreas y abarcaba gran parte de las tres provincias más orientales de Ecuador, Napo, Orellana y Pastaza (Wallis, 1973). La Reserva Étnica Waorani está ubicada como se ha mencionado anteriormente junto al Parque Nacional Yasuní; el conjunto de ambos compone la Reserva Yasuní de la Biosfera (Gómez-Pérez, 2015), delimitada por los ríos Napo al norte y Curaray y Vilano al sur y con una temperatura constante de entre 18°C y 24°C a lo largo del año, lo que permite el desarrollo del bosque templado y húmedo que caracteriza la zona (Cabodevilla, 1994).

Esta reserva étnica, donde varias comunidades se ubican dispersas a lo largo de los ríos orientales de las provincias mencionadas, es el resultado de una descentralización iniciada a finales de la década de 1970. En ese

momento el 90% de la población había sido reubicada en un único asentamiento lo que provocó enfrentamientos intergrupales, enfermedades y agotamiento de recursos que desembocaron en una significativa reducción poblacional. Desde la llegada de los primeros misioneros en 1958 se había forzado al reagrupamiento y reducción artificial de los clanes Waorani (Baeta, 2011; Gómez-Pérez, 2015).

Desde entonces ha existido un contacto limitado pero creciente con extranjeros y otras etnias indígenas (Yost, 1981; Boster et al, 2003; Gómez-Pérez, 2015). Actualmente la mayor parte de los Huaorani están en contacto con otros pueblos indígenas vecinos, como los kichwas, con colonos y trabajadores de la industria petrolera y maderera (Fuentes, 1997). El pueblo Huaorani ha experimentado en consecuencia cambios en su estructura social, económica y cultural, como por ejemplo el paso del modelo seminómada tradicional al sedentarismo en muchos grupos o la realización de uniones monógamas e incluso exógamas, con individuos de fuera del grupo (Baeta, 2011).

El Gobierno Nacional de Ecuador declaró en 1999 y asignó en 2007 a través de dos decretos presidenciales la Zona Intangible Tagaeri Taromenane (ZITT) para proteger el territorio donde viven estas etnias (Pappalardo et al, 2013; Gómez-Pérez, 2015).







# HIPÓTESIS DE TRABAJO

---



La configuración MAPT\*H2 es la menos habitual de los haplotipos tau. Su peculiar distribución (prácticamente sólo presente en el continente europeo y a unas frecuencias relativamente bajas, del 0,2) ha dado lugar a una serie de hipótesis sobre su origen y evolución que se han detallado en el apartado de introducción y se resumen en:

- Origen en una especie de homínido de distribución europea e introgresión en *Homo sapiens* mediante hibridación (Evans et al, 2004; Hardy et al, 2005; Stefansson et al, 2005; Donnelly et al, 2010).
- Selección de uno de los dos haplotipos frente al otro debido a su protección contra enfermedades neurodegenerativas o a un posible aumento de la fertilidad (Zody et al, 2008; Stefansson et al, 2005; Boettger et al, 2012; Steinberg et al, 2012).
- Historia demográfica de las poblaciones (Zody et al, 2008; Donnelly et al, 2010; Steinberg et al, 2012; Alves et al, 2015), en donde destacan:
  - o Efecto de la recolonización de Europa desde el refugio franco-cantábrico tras la última glaciación. Éste habría determinado la actual distribución de frecuencias MAPT en el continente, más altas en la región franco-cantábrica y disminuyendo a medida que las poblaciones se alejan de allí (no publicada, primera hipótesis de este grupo).
  - o El origen de las “altas” frecuencias de MAPT\*H2 y la distribución de los haplotipos MAPT se deben al efecto de la expansión de la agricultura durante el Neolítico.

Así, la hipótesis principal de trabajo a lo largo de los años de desarrollo de este proyecto ha sido que la distribución de las frecuencias de los haplotipos MAPT en Europa tenían su origen en el efecto fundador asociado a la recolonización postglacial; es por ello que se le ha dado una notable importancia tanto en los apartados de introducción como en la discusión.

En cambio, a la luz de los últimos datos revelados, ésta primera hipótesis ha debido ser replanteada, proponiendo como hipótesis final que *el origen de los haplotipos MAPT en Europa, y presumiblemente en todo Eurasia occidental y también en el este de África, se deban a la expansión de la agricultura durante el Neolítico desde Oriente Próximo.*



# OBJETIVOS

---



El objetivo fundamental de esta tesis doctoral ha sido la aportación de nuevos datos acerca del origen y la distribución actual de los haplotipos MAPT en Europa, contrastando las principales hipótesis existentes, detalladas previamente, con el fin último de establecer el origen de las variaciones regionales observadas en la distribución de *MAPT\*H2*.

Ligado a este, se han propuesto los siguientes objetivos secundarios:

- 1- Aportar nuevos datos e interpretaciones que permitan entender la distribución de las frecuencias MAPT en otros continentes, particularmente en aquellos lugares donde dichas frecuencias alcanzan los valores más altos.
- 2- Valorar la importancia y establecer la secuencia de los procesos microevolutivos que han dado lugar a la distribución actual de los haplotipos MAPT en humanos.
- 3- Enriquecer la base de datos de polimorfismos de la región MAPT. Ampliar las bases de datos existentes relativas a las frecuencias de los marcadores indel rs72377089 (Baker et al, 1999), microsatélites MAPT (Donnelly et al, 2010), y SNPs (Donnelly et al, 2010; Steinberg et al, 2010), así como de los haplotipos y subhaplotipos de la región 17q21.31.





# MATERIAL Y MÉTODOS

---



## Poblaciones y áreas de estudio.

Para esta investigación se han analizado ocho poblaciones de varias regiones del mundo, agrupadas en zonas geográficas muy distantes. Por un lado, el extremo occidental de Europa, por otro el sur del continente americano. Todos los donantes dieron su consentimiento informado antes de su inclusión en la muestra, siguiendo las directrices éticas estipuladas por cada una de las instituciones involucradas en la investigación.

**Tabla M.1:** Población a la que pertenece cada muestra, ubicación geográfica y tamaño.

Población / etnia	Área Geográfica	Tamaño muestra
Vascos	Gipuzkoa, País Vasco	105
Vascos	Navarra, País Vasco	107
Residentes	País Vasco	36
Gitanos	Bizkaia, País Vasco	118
Valencianos	Valencia, España	118
Afrodescendientes	Chocó, Colombia	104
Indígenas jujeños	Jujuy, Argentina	104
Waorani	Pastaza, Ecuador	36

### PENÍNSULA IBÉRICA.

#### Vascos de Gipuzkoa.

La región del Goierri, de donde proviene la muestra, está ubicada en el sudeste de la provincia, en las zonas altas de los valles del río Oria y del Urumea.

La muestra de vascos de Gipuzkoa está formada por 105 individuos sanos, no emparentados, del Goierri gipuzkoano. Para evitar posibles sesgos

introducidos por la heterogeneidad de la muestra se incluyeron en el estudio sólo aquellos individuos autóctonos. El criterio empleado para aceptar a los voluntarios como autóctonos fue que los cuatro primeros apellidos tuvieran un origen vasco y que tanto su lugar de nacimiento como el de sus tres generaciones precedentes fuera guipuzcoano.

Las muestras se tomaron por punción venosa y el ADN genómico se extrajo de sangre periférica mediante el procedimiento estándar del fenol-cloroformo.

### **Residentes del País Vasco**

La muestra de residentes vascos se tomó de 36 individuos sanos, de ambos sexos y no emparentados residentes en la comarca del Gran Bilbao.

Las muestras se tomaron por punción venosa y el ADN genómico se extrajo de sangre periférica mediante el procedimiento estándar del fenol-cloroformo.

### **Gitanos vascos.**

La muestra de gitanos vascos está compuesta por 118 individuos de la etnia gitana no emparentados entre sí. Los individuos y/o sus familias se identificaron como pertenecientes a la etnia gitana. Para evitar la heterogeneidad de la muestra y el efecto mosaico se excluyeron los gitanos de origen rumano, conformando esta muestra sólo gitanos de origen ibérico residentes en la zona del Gran Bilbao. Se trata de una muestra de rroma predominantemente nacidos en la provincia de Bizkaia.

Las muestras de saliva de todos los participantes en el estudio se tomaron empleando el kit de Oragene OG-250 (DNA Genotek, Ottawa, Ontario, Canadá), un sistema “todo-en-uno” que recoge, preserva, transporta y purifica el ADN de la saliva. El ADN fue extraído de una alícuota de 500 µl de saliva usando el protocolo del fabricante para la purificación de ADN, PD-PR-015 Issue 2.0. El ADN se cuantificó mediante el espectrofotómetro Nanodrop (NanoDrop Technologies, USA).

### **Vascos de Navarra.**

Los 107 voluntarios que integran esta muestra son oriundos de cuatro pueblos rurales de la región vascofona de Navarra: Bera, Elizondo, Lekarotz y Leitza.

Las muestras fueron tomadas mediante punción venosa de estudiantes de *Ikastolas* – colegios donde se imparte la enseñanza totalmente en euskera. Para ser considerados autóctonos sus apellidos y el lugar de nacimiento de sus antecesores hasta la tercera generación debían ser de origen vasco. La extracción de ADN se realizó mediante el método estándar de fenol-cloroformo.

### **Valencia.**

Las muestras de ADN se obtuvieron de 118 voluntarios sanos y no emparentados de dicha provincia. El criterio de inclusión en la muestra fue el lugar de nacimiento de los donantes, de sus padres y de sus abuelos. La extracción se realizó a partir de las manchas de sangre en papel de filtro mediante el procedimiento estándar del fenol-cloroformo (Sambrook et al., 1989).

## **SUDAMÉRICA**

### **Afrodescendientes de Chocó, Colombia.**

La muestra empleada en esta tesis procede de diferentes núcleos poblacionales del departamento del Chocó colombiano y consiste en muestras de ADN de 104 individuos de ascendencia africana, voluntarios informados y sanos. La clasificación fue realizada en función de su fenotipo morfológico y su auto-reconocimiento, siguiendo los criterios utilizados en el censo de 2005 (DANE, 2005), el más reciente disponible en el momento del estudio. Se tuvo también en cuenta su lugar de nacimiento y el de sus ascendientes hasta la

tercera generación para considerar a los voluntarios como originarios del departamento.

La muestra fue tomada por un equipo dirigido por la Dra. Bravo (GENES, Medellín, Colombia).

### **Waoranis de Ecuador.**

La muestra de la tribu Waorani consiste en 36 individuos de diferentes casas comunales o *nanacaibos* – grupos domésticos tradicionales – asentados precisamente en este Parque Nacional. Un estudio reciente de microsátelites de ambos cromosomas sexuales (X-STRs e Y-STRs) confirma la existencia de varios linajes paternos / maternos (Cardoso et al, 2012) asegurando por tanto la ausencia de parentesco cercano entre los individuos incluidos en esta muestra (Gómez-Pérez, 2013).

Los voluntarios eran individuos sanos e informados acerca del propósito de la investigación. Las muestras fueron tomadas por un equipo dirigido por la Dra. Matínez-Jarreta, de la Universidad de Zaragoza.

El ADN fue extraído de manchas de sangre en papel siguiendo el procedimiento estándar del fenol-cloroformo.

### **Indígenas de Jujuy, Argentina.**

Se han recogido muestras de sangre sobre papel (Whatman y de filtro) de 104 individuos sanos y no emparentados de distintos núcleos poblacionales de la provincia de Jujuy. Todos los individuos eran voluntarios informados y sanos. Su linaje ha sido verificado mediante información biogeográfica comprobada hasta al menos la tercera generación; principalmente los lugares de nacimiento del donante, de sus padres y sus abuelos.

La muestra fue tomada por un equipo dirigido por el Dr. Dipierri (Universidad Nacional de Jujuy, Argentina) El ADN fue extraído de manchas de sangre en papel siguiendo el procedimiento estándar del fenol-cloroformo (Sambrook et al.,1989).

**Tabla M.2:** Resumen del origen y el tipo de extracción / purificación del material genético.

Nombre de la muestra	Origen	Tipo de extracción / purificación
Vascos de Gipuzkoa	Sangre periférica	Extracción orgánica fenol-cloroformo
Vascos de Navarra	Sangre periférica	Extracción orgánica fenol-cloroformo
Residentes vascos	Sangre periférica	Extracción orgánica fenol-cloroformo
Gitanos vascos	Saliva	Qiagen
Valencianos	Sangre en papel	Extracción orgánica fenol-cloroformo
Afrodescendientes de Chocó	Banco de ADN	Banco ADN
Indígenas de Jujuy	Sangre en papel	Extracción orgánica fenol-cloroformo
Worani de Ecuador	Sangre en papel	Extracción orgánica fenol-cloroformo

## Marcadores y análisis molecular.

Para este proyecto de investigación se han estudiado el indel rs72377086, 4 microsatélites y diez SNPs; un total de 15 marcadores localizados en la región 17q21.31 que cubren algo más de 1 Mb de la misma, incluyendo la inversión del cromosoma 17 (figura M.1.).

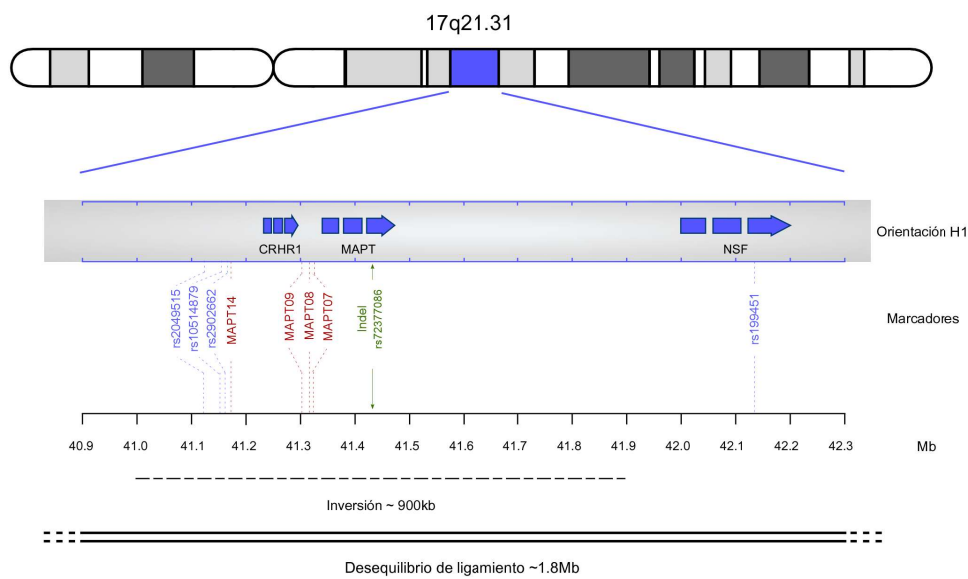


Figura M.1.: Marcadores analizados.

## Indel, PCR y electroforesis en gel de agarosa.

Se ha analizado el indel rs72377086 en todas las poblaciones del estudio. La presencia o ausencia del fragmento intrónico de 238 pb fue determinada mediante la visualización del producto de la PCR junto con un marcador de peso molecular en un gel de agarosa.

## Amplificación por PCR y electroforesis en gel de agarosa.

La región seleccionada fue amplificada mediante PCR (Mullis et al, 1986) empleando los siguientes cebadores (Baker et al, 1999):



F (5'-3) GCA AGA CGT TCT CAC TGA TCT G

R (5'-3) AGG AGT CTG GCT TCA GTC TCT C

La reacción en cadena de la polimerasa se llevó a cabo en un volumen total de 10  $\mu$ l por muestra que contenían 1 $\mu$ l de tampón de PCR 10x de Bioline, 0.3 $\mu$ l de MgCl<sub>2</sub> 50mM de Bioline, 0.05 $\mu$ l de cada uno de los dNTPs 100mM de Bioline, 0.5 $\mu$ l de glicerol, 0.2 $\mu$ l de cada uno de los cebadores a una concentración final de 0.75 $\mu$ M y 0.12 $\mu$ l de la enzima Taq Polimerasa de Bioline a una concentración de 0.06 U/ $\mu$ l, en un termociclador Perkin Elmer GeneAmp PCR System 9700 de Applied Biosystems.

El protocolo de amplificación constó de 3 minutos a 95°C, de 10 ciclos de 95°C durante 30 segundos, 62°C durante 30 segundos y 72°C durante 45 segundos, de 25 ciclos de 95°C durante 30 segundos, 60°C durante 30 segundos y 72°C durante 45 segundos y una extensión final de 10 minutos a 72°C.

Se cargó un volumen de 5 $\mu$ l a 7 $\mu$ l de amplificado en el gel de agarosa (1.5%) tratado con bromuro de etidio, así como 4 $\mu$ l de una solución diluida en agua ultrapura del marcador molecular *DNA Molecular Weight Marker V 100 bp ladder* de Roche (0.225  $\mu$ g / $\mu$ l) (1:15), con un rango de 8 pb a 587 pb, para la identificación de los alelos “inserción” y “no-inserción”, cuyo tamaño es de 484pb y 246pb respectivamente.

Durante el transcurso de esta investigación se sustituyó el bromuro de etidio por un colorante más seguro, GelRed de Biotium, que presenta una menor toxicidad, una susceptibilidad mayor y un espectro de emisión prácticamente idéntico, que permite mantener el mismo equipo de detección y captura de imagen. Ambos son colorantes que se intercalan entre las bases del ADN y que emiten fluorescencia cuando son iluminados por luz ultravioleta.

Las muestras se dejaron migrar 20 minutos a 115 voltios. Finalizada la electroforesis se visualizaron en un transiluminador de luz ultravioleta. Los

resultados han sido documentados mediante un sistema de fotografía digital (Kodak Digital Science EDAS 120).

### Microsatélites, PCR y electroforesis en capilar.

Se han analizado cuatro microsatélites- MAPT 07, MAPT 08, MAPT 09 y MAPT 14 - dentro de los límites de la inversión en todas las poblaciones del estudio. La longitud de los productos de la PCR fue determinada mediante el secuenciador ABI Prism 310 Genetic Analyzer y un marcador del tamaño molecular.

La región seleccionada fue amplificada mediante PCR empleando los siguientes cebadores (Donnelly et al, 2010), de los cuales uno de cada par estaba marcado por un fluorocromo:

**Tabla M.3:** Microsatélite, cebadores y modificación correspondiente.

Marcador	Primer	Modificación
STRP MAPT 07	F: 5' TCA GAT TCG TTT CGG GGG TCG TTA TCG 3'	5FAM
	R: 3' GCT GGA GGG TCT TCA CCA CCA GAT TCG C 5'	
STR MAPT 08	F 5' CAA TGA GCC AAG ATC ACA CCT ACT G 3'	5HEX
	R: 3' GGG TCC TAT CAT TCT CTC TGC TAT CA 5'	
STRPs MAPT 09	F 5' TTC TTC TTG CCA ATA GCA AGG ATA C 3'	5FAM
	R: 3' CAG TAT ATA CCC AGA AAA GTG CAA TGT G 5'	
STRP MAPT 14	F 5' CTC ACA AAT GTT CTC TGT CTC TTC TGT C 3'	5TET
	R: 3' GAT GCA TCT AAT TTA AGG CAT CTT ATG A 5'	

La reacción en cadena de la polimerasa se llevó a cabo en un volumen total de 10 µl por muestra (concentración aproximada de ADN, 2 ng) con los siguientes volúmenes y concentraciones de reactivos de Bioline:

**Tabla M.4:** Volúmenes y concentraciones de los reactivos empleados para la PCR.

Reactivo	MAPT 07	MAPT 08	MAPT 09	MAPT 14
[concentración inicial]	Vol (μl)	Vol (μl)	Vol (μl)	Vol (μl)
PCR Buffer [10x]	1	1	1	1
MgCl <sub>2</sub> [50mM]	0.16	0.17	0.17	0.16
dNTPs [100mM]	0.02	0.15	0.1	0.15
Primer A [5μM]	0.35	0.37	0.35	0.4
Primer B [5μM]	0.35	0.37	0.35	0.4
H <sub>2</sub> O <sub>up</sub>	Hasta 10 μl de volumen total			
Taq Polimerasa [5 U/μl]	0.1	0.1	0.08	0.1
ADN	0.1 - 0.2			

La amplificación se realizó mediante el termociclador Gene Amp PCR System 9700 PE Applied Biosystems. Los programas empleados se detallan a continuación.

**Tabla M.5:** Programas empleados para la PCR.

	MAPT 07	MAPT 08	MAPT 09	MAPT 14
1	95 30 seg	94 60 seg	94 60 seg	94 60 seg
2	45 20 seg 32	64,5 90 seg 7	62 90 seg 7	65 90 seg 7
3	72 30 seg	72 90 seg	72 90 seg	72 90 seg
1		95 60 seg	94 60 seg	94 60 seg
2		63 60 seg 25	60 60 seg 25	62 60 seg 25
3		72 60 seg	72 60 seg	72 60 seg

En todos los casos se llevó a cabo una desnaturalización previa de 10 minutos a 94°C – salvo para el microsatélite MAPT07, para el cual la temperatura fue de 95°C – y una extensión final de 10 minutos a 72°C.

Los números del 1 al 3 de la primera columna hacen referencia a los procesos de desnaturalización (1), anillamiento (2) y elongación (3). Dentro de las columnas de cada marcador se detallan, de izquierda a derecha, las temperaturas en °C, los tiempos en segundos y el número de repeticiones del ciclo.

Todos los programas, menos el del marcador MAPT07, tienen dos grupos de ciclos, el primero a una temperatura de anillamiento superior al segundo, para prevenir la aparición de picos inespecíficos en el electroferograma.

El amplificado resultante de la PCR se mezcló con el fluorocromo Gene Scan 500 TAMRA (Applied Biosystems) como estándar interno y Hi-Di Formamida en proporción 1:1:18. El estándar permite detectar el tamaño y la cantidad de cada producto de PCR, ya que se usa como referencia para extrapolar el tamaño de los fragmentos amplificados, eliminando la variación entre capilares. Se ha empleado también como control positivo. La formamida se emplea para desnaturalizar las muestras de ADN antes de colocarlas en el analizador genético. La presencia de cantidades variables de agua y de iones no deseados en la formamida estándar puede causar problemas debido a la competencia de los citados iones con los iones del ADN o a la reacción de estos iones con el ADN, provocando la degradación de la muestra. Por ello se ha empleado formamida altamente desionizada.

Esta mezcla se sometió a una incubación a 95°C durante 5 minutos para favorecer la desnaturalización del ADN e inmediatamente después se enfrió durante 3 minutos a -20°C para detener la reacción (proceso también denominado estabilización).

La detección de fragmentos se realizó mediante el secuenciador automático de ADN ABI Prism 310 Genetic Analyzer (Applied Biosystems) por electroforesis capilar a una diferencia de potencial de 15.0 kV y a una temperatura de 60°C, tras un tiempo de inyección que ha variado entre cinco y diez segundos. Las muestras han migrado durante quince a treinta minutos en función del tamaño de los alelos de cada marcador.

En cada análisis se incluyó como blanco una muestra de Gene Scan 500 TAMRA (Applied Biosystems) sin amplificado, además de una prueba de color para comprobar el adecuado funcionamiento del equipo, protocolo de uso recomendado por Applied Biosystems.

Se empleó el programa GeneMapper® ID Software v3.2.1 (Applied Biosystems) para la lectura de los resultados que con el marcador estándar Gene Scan 500 TAMRA como referencia realizó la asignación alélica para cada marcador estudiado.

**Tabla M.6:** En la primera columna se representa el nombre del alelo, que coincide con el número de repeticiones del mismo. En la segunda el símbolo.

MAPT 07		MAPT 08		MAPT 09		MAPT 14	
15	201 bp	7	187 bp	13	107 bp	9	99 bp
16	203 bp	8	191 bp	14	109 bp	10	103 bp
17	205 bp	9	195 bp	15	111 bp	11	107 bp
18	207 bp	10	199 bp	16	113 bp	12	111 bp
19	209 bp	11	203 bp	17	115 bp	13	115 bp
20	211 bp	12	207 bp	18	117 bp	14	119 bp
21	213 bp	13	211 bp	19	119 bp	15	123 bp
22	215 bp			20	121 bp	16	127 bp
23	217 bp			21	123 bp	17	131 bp
24	219 bp			22	125 bp	18	135 bp
25	221 bp			23	127 bp	19	139 bp
26	223 bp			24	129 bp	20	143 bp
						21	147 bp
						22	151 bp
						23	155

### SNPs y HRM, High Resolution Melt Analysis.

Se seleccionaron diez SNPs de la bibliografía de los cuales cinco (rs2049515, rs10491144, rs17651213, rs10514879 y rs2902662) discriminan entre los haplotipos principales H1 y H2, debiendo por tanto sus resultados

ser idénticos a los obtenidos con el análisis del indel rs72377086, y otros cinco (rs199448, rs199451, rs199456, rs199457 y rs199533) que discriminan entre los subhaplotipos H2´ y H2D.

La región seleccionada fue amplificada durante el proceso de PCR a tiempo real que precede al análisis HRM empleando los cebadores de la casa Eurofins MWG Operon indicados en la tabla M.4, diseñados con el programa PerlPrimer (Marshall OJ, 2004), y en presencia de *EvaGreen* (Biotium), un fluorocromo de nueva generación que emite fluorescencia únicamente cuando está unido a ADN de doble hebra.

**Tabla M.7:** En la primera columna se indica el nombre del marcador, en la segunda los cebadores correspondientes y en la tercera la referencia bibliográfica.

Marcador	Primer	Referencia
SNP rs2049515	F: 5' TAG GAA TGT CAA AGT CCA AAC C 3' R: 5' CAA AGT GAA GAC TAT CTG TCT CC 3'	
SNP rs10491144	F: 5' TTC CAC CTC CTT CCT AAG AC 3' R: 5' TTG AGC TTG ACT ATT TGA CCA G 3'	
SNP rs17651213	F: 5' GAT CCC AGA AGG AAC CAC AG 3' R: 5' GCT ACT CAC AAC CCA AGA TTC C 3'	Donnelly <i>et al</i> 2011
SNP rs10514879	F: 5' TGA GAT CCG GCA GAT AAA TG TG 3' R: 5' AAT GGA CTC TGA AAT CTC ACT GTC 3'	
SNP rs2902662	F: 5' CTA CCG TGT ACC TAG ATG GAG 3' R: 5' GAG TGT TGA TGT TTC CTG GC 3'	
SNP rs199448	F: 5' TGA GGG TGA GCA TCC AAG GTA CG 3' R: 5' CTT ACA TAT TTC TAT ATT CTT C 3'	
SNP rs199451	F: 5' TCA GAG ACT CAA GCT AAT AG 3' R: 5' TGA TTC ACC ATA CTC CTT TCC C 3'	
SNP rs199456	F: 5' AGT TTA GGT CCA CTT CCT GAG 3' R: 5' GGG TGC TAT CAG TAA TAA TCT CGT 3'	Steinberg <i>et al</i> 2012
SNP rs199457	F: 5' CAG AGT GAC ACC AGT AAT GCC 3' R: 5' CCA TTT GGC TTC TCC CTC AG 3'	
SNP rs199533	F: 5' CTG TAG GGA CAT CTC GAT CAG 3' R: 5' GGG CAA CTT CAA GGA TAA GG 3'	

El análisis HRM permite la caracterización de muestras de ácidos nucleicos en función de su perfil de desnaturalización. Este método post-PCR se basa en que las diferentes muestras tienen distintas secuencias de ADN y por tanto una temperatura de desnaturalización específica para cada una de ellas. En el caso de los SNPs se espera encontrar tres tipos de amplicones, dos homocigotos para cada uno de los alelos del marcador y uno heterocigoto. Esta identificación es posible gracias a la combinación de fluorocromos saturantes de nueva generación que no inhiben la PCR a pesar de encontrarse en grandes cantidades, de instrumentos de alta resolución para la detección de la fluorescencia y de un sofisticado software con un algoritmo capaz de analizar las formas de las curvas de desnaturalización y clasificarlas en distintos grupos por su similaridad, idealmente en tres: homocigoto salvaje, homocigoto mutante y heterocigoto.

El termociclador empleado para este proyecto, el BIO-RAD C1000 con módulo óptico CFX96, lleva a cabo el proceso de pre-amplificación y de análisis HRM en un solo ensayo. Esto permite visualizar el “*amplification plot*” o curva de amplificado, lo que supone una gran ventaja a la hora de interpretar los resultados de HRM, ya que aquellas muestras que presenten una amplificación defectuosa, anormal o que no la presenten en absoluto pueden ser identificadas y en su caso eliminadas del análisis posterior (*downstream HRM analysis*) evitando artefactos, además de permitir la optimización de los cebadores.

#### Optimización de cebadores / Puesta a punto del protocolo.

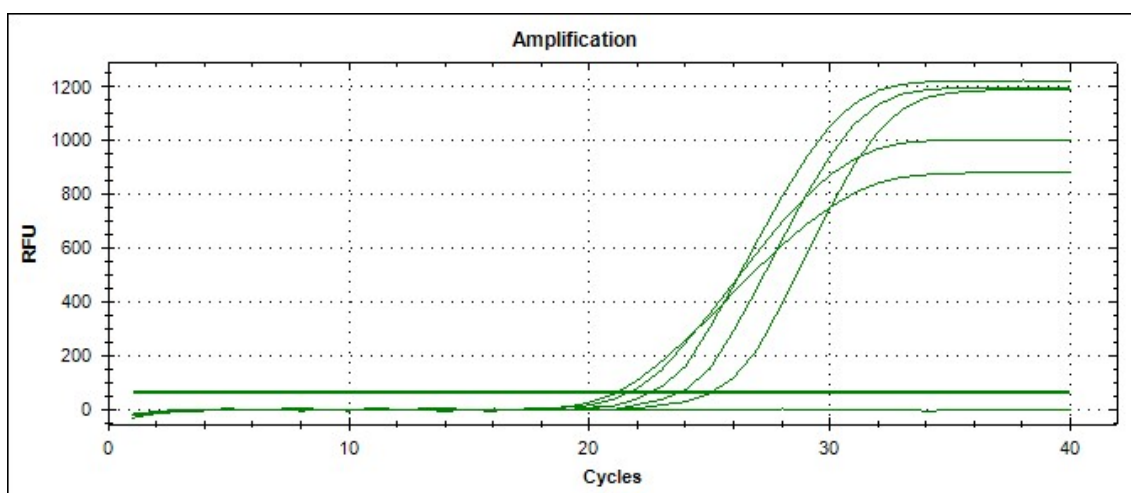
Previo al estudio de las muestras se realizó una optimización de los cebadores para determinar (1) la temperatura, (2) la concentración de cebador y (3) la concentración de ADN óptimas para el análisis.

Para ello se cargaron distintas concentraciones de ADN (diluciones seriadas: 10ng/μl, 5ng/μl, 2.5ng/μl, 1.25 g/μl y 0.625 g/μl) en las columnas de una placa y empleando la función de gradiente del termociclador BIO-RAD C1000 con módulo óptico CFX96 se probaron distintas temperaturas en cada

fila, tomando como medida orientativa la temperatura de disociación de los cebadores.

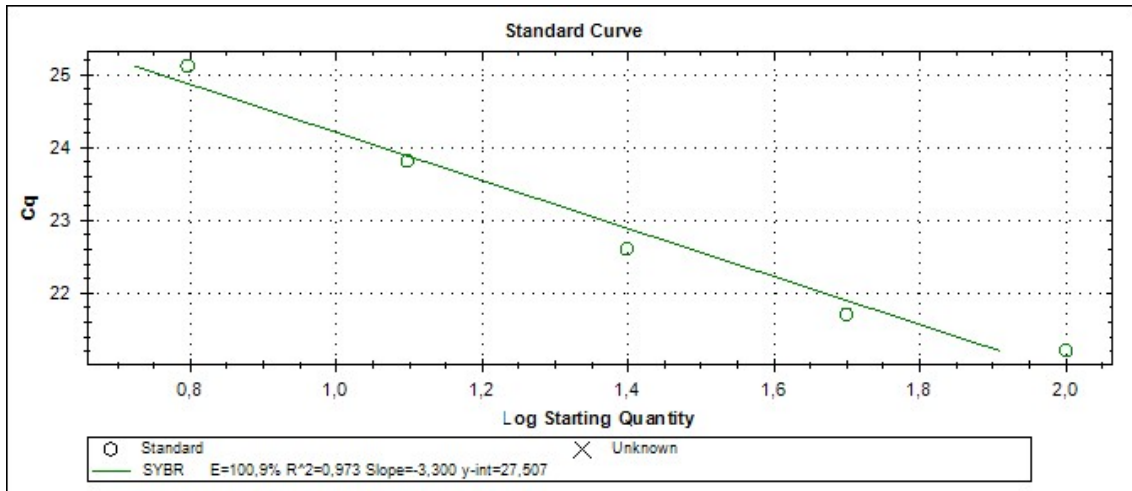
Dado que cada placa empleada tenía doce columnas y ocho filas, se analizaron en cada prueba dos primers con temperaturas de disociación similares, uno de la columna 1 a la 5 y otro de la 7 a la 11, dejando las columnas 6 y 12 como controles negativos, y ocho temperaturas.

Los datos obtenidos fueron procesados mediante el software Bio-Rad CFX Manager versión 2.0.

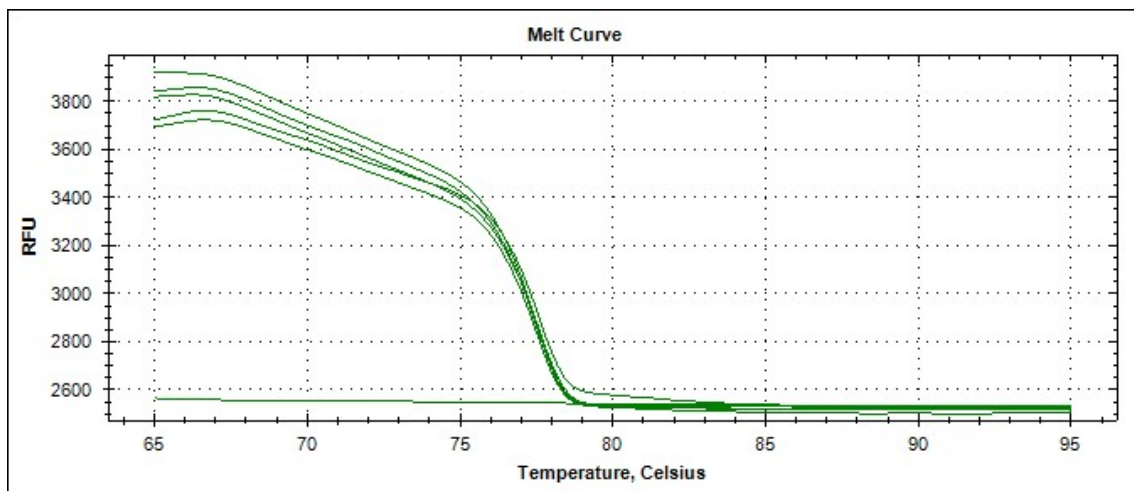


**Figura M.2:** Curva de amplificación o *amplification plot* de una de las réplicas de un ensayo de optimización de cebadores. En el eje de las ordenadas se representan las unidades de fluorescencia relativa (*Relative Fluorescence Unit*), en el eje de las abscisas el número de ciclos. La línea verde horizontal es la medida de la “fluorescencia umbral”, que es la resta de la fluorescencia de base (*baseline fluorescence*) a la fluorescencia de la muestra. El punto en el que la curva de amplificación de la muestra corta esta línea se llama *quantification cycle*, **C<sub>q</sub>** (también **C<sub>T</sub>**, *threshold cycle* o ciclo umbral). Para aceptar un protocolo las curvas de amplificación deben llegar a la fase estacionaria, situarse preferiblemente entre las 800 – 1000 RFUs y comenzar su amplificación entre los ciclos 22 y 28. Ciclos umbrales por encima del 30 son considerados no válidos por deberse probablemente a la amplificación de uniones entre los propios cebadores.

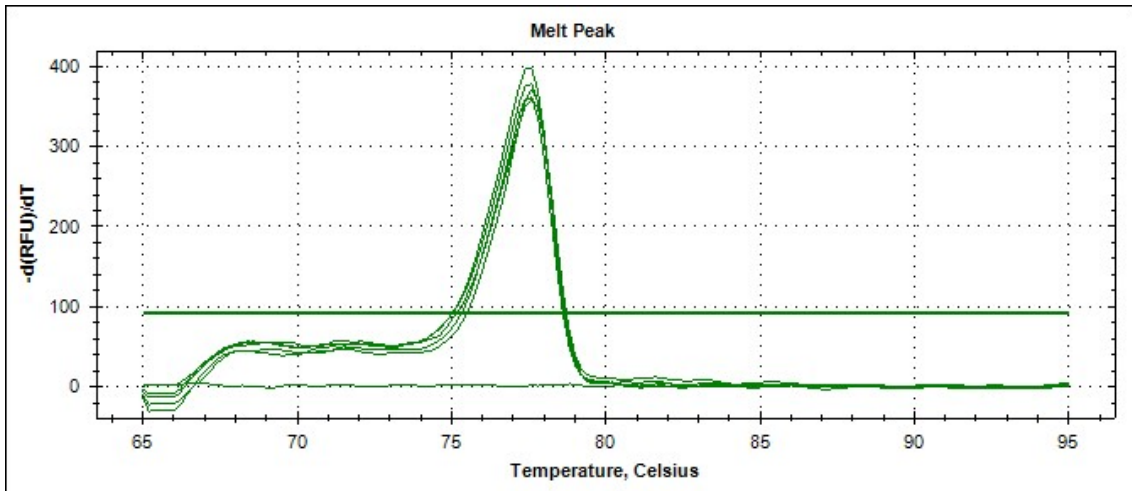




**Figura M.3:** Curva estándar. En el eje de ordenadas se representa el ciclo umbral o ciclo de cuantificación ( $C_q$ , *quantification cycle*) y en el de abscisas el logaritmo de la concentración inicial de ADN. La recta verde es la representación de la ecuación del fluorocromo Eva Green, mientras que los círculos son las muestras de la réplica. “E” es la medida de la eficacia del proceso;  $R^2$  el coeficiente de determinación. Se buscan aquellas condiciones de concentración y temperatura en las que la eficiencia de cebador se sitúe lo más cerca posible del 100% con una  $R^2$  similar a 1.

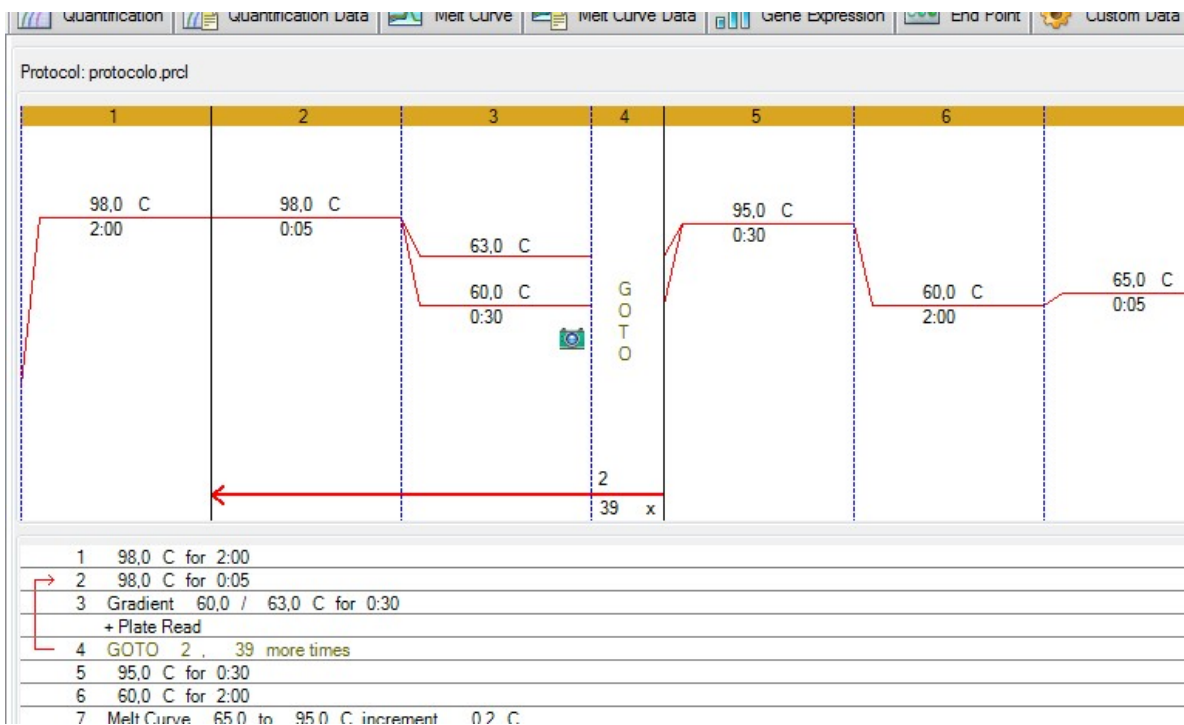


**Figura M.4:** Curva de desnaturalización o *melting curve* donde se aprecia cómo la fluorescencia, medida en RFUs y representada en el eje de ordenadas, disminuye a medida que aumenta la temperatura, en °C y en el eje de las abscisas, dado que el fluorocromo sólo emite fluorescencia cuando está unido a ADN de doble hebra y éste se desnaturaliza progresivamente con el incremento de la temperatura. La  $T_m$  queda entre los 76°C y los 78°C.



**Figura M.5:** En esta figura se representa la  $T_m$  en forma de pico de desnaturalización o *melt peak*. Aquellas muestras cuyo pico de desnaturalización es muy bajo o queda por debajo del umbral se desestiman.

Todas las directrices para la optimización de protocolo expuestas sirven también para la interpretación de los resultados de HRM. Las condiciones finales se detallan en el siguiente apartado.



**Figura M.6:** Pantalla resumen del protocolo de PCR de puesta a punto de los cebadores.

Análisis de las muestras mediante HRM.

La reacción en cadena de la polimerasa se llevó a cabo en un volumen total de 5  $\mu\text{l}$  por muestra con los siguientes volúmenes y concentraciones de reactivos:

**Tabla M.8:** Volúmenes y concentraciones de reactivos empleados en la PCR.

Reactivo	Volumen ( $\mu\text{l}$ ) por muestra
Master Mix con EvaGreen	2.5
PrimerA [1 $\mu\text{M}$ ]	0.75
PrimerB [1 $\mu\text{M}$ ]	0.75
ADN [15 $\eta\text{g}$ / $\mu\text{l}$ ]	1
FINAL	5

La amplificación se realizó en el termociclador BIO-RAD C1000 con modulo óptico CFX96. El programa conjunto de PCR y HRM es el que se detalla a continuación, donde  $T_a$  es la temperatura descrita en la tabla X, específica de cada par de primers.

Fase PCR: 98°C x 3 minutos; [98°C x 10 segundos,  $T_a$  x 30 segundos] 39 ciclos.

Fase HRM: 95°C x 10 segundos, 65°C x 5 segundos, aumento de la temperatura hasta los 95°C con lectura de la fluorescencia cada incremento de 0.2°C.

**Tabla M.9:** Relación de cebadores, alelos del marcador y temperatura de hibridación o *annealing* ( $T_a$ ).

H1 ' vs H2 '			H2 ' vs H2D		
Ref.: <i>Donnelly</i> 2010			Ref.: <i>Steinberg</i> 2012		
Oligo name	SNP	T (°C) HRM	Oligo name	SNP	T (°C) HRM
rs2049515	C/T	60,1	rs199533	C/T	56,6
rs1051487	G/A	61,2	rs199451	G/A	58,2
rs10491144	A/C	58,2	rs199456	C/T	60
rs2902662	C/T	61,2	rs199448	A/G	????
rs17651213	G/A	58,3	rs199457	C/T	60

Se empleó el programa Bio-Rad Precision Melt Analysis para la lectura de los resultados.

## Análisis estadísticos.

“En aquel imperio, el arte de la cartografía logró tal perfección que el mapa de una sola provincia ocupaba toda una ciudad, y el mapa del imperio, toda una provincia. Con el tiempo, esos mapas desmesurados no satisficieron y los colegios de cartógrafos levantaron un mapa del imperio, que tenía el tamaño del imperio y coincidía puntualmente con él. Menos adictas al estudio de la cartografía, las generaciones siguientes entendieron que ese dilatado mapa era inútil y no sin impiedad lo entregaron a las inclemencias del sol y de los inviernos. En los desiertos del Oeste perduran despedazadas ruinas del mapa, habitadas por animales y por mendigos; en todo el país no hay otra reliquia de las disciplinas geográficas”

Jorge Luis Borges (El hacedor, 1960)

Se define **frecuencia alélica** como la proporción relativa de cada alelo frente al total de alelos examinados. Las frecuencias alélicas de todos los marcadores han sido calculadas por recuento directo y su **error estándar** ha sido hallado mediante la fórmula

$$ES = \sqrt{\frac{p(1-p)}{2N}}$$

donde  $p$  es la frecuencia del alelo y  $N$  es el tamaño de la muestra.

En algún caso se ha empleado también el programa STRs desarrollado en el laboratorio para tal fin.

Dado que para los haplotipos autosómicos no puede conocerse la fase gamética de los alelos, el cálculo de las **frecuencias haplotípicas** se ha realizado mediante inferencias basadas en el método de máxima verosimilitud empleando el programa Arlequín v3.5 (Excoffier y Lischer, 2010).

Éste mismo programa se ha usado para analizar el **equilibrio Hardy-Weinberg** en las poblaciones del estudio. La comprobación de esta ley se

realiza mediante la comparación de la heterocigosidad observada y la esperada para la muestra, partiendo de la hipótesis de que en ausencia de selección natural, mutación y migración, en una población suficientemente grande como para que no se aprecien los efectos de la deriva genética y en dónde los apareamientos son aleatorios, las frecuencias alélicas y genotípicas van a permanecer constantes generación tras generación.

Para reducir el riesgo de cometer un error estadístico de tipo I (identificar una hipótesis como significativa cuando en realidad no lo es), cuya probabilidad de ocurrencia aumenta al realizar contrastes múltiples, se ha aplicado la **corrección de Bonferroni** al análisis del equilibrio de Hardy-Weinberg, de tal manera que si se efectúa un número K de comparaciones, para mantener la probabilidad global p de rechazar incorrectamente la hipótesis nula, el nivel de corte a emplear en cada contraste debe ser  $p/K$ .

El objetivo de este método es por tanto no aumentar la probabilidad global de hallar resultados sólo por efectuar una gran cantidad de análisis en diferentes variables obtenidas en nuestra muestra.

Para medir la variabilidad genética de los microsatélites analizados se obtuvieron los valores de **diversidad genética** (gene diversity, GD) y el **contenido de información polimórfica** (Polymorphic Information Content, PIC) mediante el programa Power Marker v. 3.0 (Liu & Muse, 2004). Para comprobar si se cumple el equilibrio de Hardy y Weinberg, se llevó a cabo el test exacto de Fisher para estimar los valores p (Guo & Thompson, 1992) usando Arlequin 3.0 (Excoffier et al., 2005)

Las frecuencias alélicas y haplotípicas pueden verse alteradas por la existencia de *loci* ligados. Para revelar la existencia de este tipo de relaciones se ha llevado a cabo el cálculo del **desequilibrio de ligamiento** con el programa Arlequin v3.5 (Excoffier y Lischer, 2010). El valor del desequilibrio de

ligamiento (LD) aporta una medida de la desviación de las frecuencias de los haplotipos en una población frente a la frecuencia esperada cuando los alelos de los *loci* que conforman el haplotipo están asociados de manera aleatoria. Este valor resulta significativo cuando los loci analizados se encuentran ligados entre sí. Este hecho puede deberse, además de a la cercanía de los marcadores analizados, a procesos como la deriva genética, el efecto fundador, la selección o las mezclas poblacionales, causando un resultado contrario a los procesos de recombinación (Reich et al., 2001).

Para estimar el grado de estructuración que presentan las poblaciones estudiadas en este proyecto de investigación se ha realizado el Análisis Molecular de la Varianza (*Analysis of Molecular Variance*, AMOVA) empleando el programa Arlequín v3.5 (Excoffier y Lischer, 2010) para realizar los cálculos.

A partir de ellos se obtienen la variación genética intergrupala ( $F_{CT}$ ), la variación genética entre poblaciones dentro de grupos ( $F_{SC}$ ) y la variación genética dentro de la población ( $F_{ST}$ ).

En algunos casos se observa una variación de las frecuencias alélicas o haplotípicas en función de una componente geográfica como la longitud, la latitud, la insolación o la altitud (Gómez-Pérez, 2015), por citar algunas. A este cambio a lo largo de un gradiente geográfico de un rasgo determinado genéticamente se le denomina **clina** y su detección y análisis puede revelar información muy valiosa ya que puede aparecer a consecuencia de presiones selectivas, de procesos migratorios o debido a la deriva genética.

Cuando se ha sospechado que alguno de los marcadores analizados podría presentar una clina, se ha realizado un análisis de correlación de las frecuencias alélicas o haplotípicas respecto a las coordenadas geográficas de las poblaciones mediante el programa GeDis v1.8 (Peña et al, 2009). La orientación de aquel eje que presente una asociación significativa es considerada como clina. Las clinas van acompañadas de una medida de

bondad del ajuste del modelo, el coeficiente de determinación  $R^2$ , que indica el porcentaje de ajuste que se ha conseguido con el modelo lineal presentado y por una medida de la significación estadística de la regresión,  $p$ .

Se han realizado **mapas sintéticos** para una mejor y más inmediata visualización de los resultados. Se trata de mapas geográficos del área estudiada en los que se han representado las variaciones locales de las frecuencias de un alelo o de un haplotipo mediante diferencias en un gradiente de color. Cada muestra queda asignada al punto del mapa correspondiente a su ubicación. A partir de las frecuencias de las diferentes muestras analizadas y de las disponibles en la bibliografía, se realiza una extrapolación para los puntos intermedios GeDis v1.8 (Peña et al, 2009).

Con el objetivo de estudiar la **autocorrelación espacial**, es decir, la existencia de patrones no meramente aleatorios de distribución espacial en los valores registrados en diferentes localizaciones, se ha utilizado el **Índice de Moran (IM)**, que compara la localización geográfica de una muestra con los valores de una variable, en este caso, de las frecuencias alélicas.

El cálculo del IM se basa en la covariación de los valores de cada una de las variables utilizadas en el mapa de muestreo poblacional y puede presentar valores de -1 a +1 con un valor esperado de  $E(IM) = -1/(n-1)$ , donde “n” es el tamaño de la muestra correspondiente.

Cuando  $IM > 0$ , la variable estudiada presenta una autocorrelación espacial positiva, es decir, los puntos que presentan valores similares para esa variable estarán yuxtapuestos en el espacio poblacional. Cuando  $IM < 0$ , la variable estudiada presenta una autocorrelación espacial negativa, donde los puntos con valores similares tenderían a estar separados en el espacio.

Los valores del índice de Moran pueden representarse frente a la distancia dando lugar a una figura denominada correlograma. El patrón esperado de acuerdo con el modelo de **aislamiento por la distancia** es el de



una progresiva disminución del índice de Moran en las sucesivas clases de distancia. Puede además estimarse la significación de los valores de autocorrelación, de modo que, aunque no es una condición decisiva, el gráfico será más fiable si muestra significación en la mayor parte de sus valores.

En condiciones ideales, esta disminución del parentesco debería ajustarse a una curva exponencial negativa

$$\Phi (\text{parentesco}) = a \cdot e^{-bd}$$

donde  $a$  es la tasa local de parentesco,  $b$  la pendiente de la curva y  $d$  la distancia.

Las poblaciones biológicas están habitualmente distribuidas en zonas geográficas más o menos continuas en vez de ser fácilmente identificables como grupos discretos y sus individuos, por tanto, dispersos – uniformemente o en núcleos – por ellas. Miembros de una misma localidad tienen más probabilidad de contribuir a la descendencia que individuos de cualquier otra localidad. Se conoce como aislamiento por la distancia al proceso de diferenciación que se da en la situación arriba descrita. Dado que el grado en que estas subpoblaciones difieran entre sí dependerá, entre otros factores, de la cantidad de flujo migratorio, la acumulación de este fenómeno a lo largo de generaciones puede dar lugar a la aparición de una correlación entre la distancia geográfica que separa las subpoblaciones y su distancia genética, provocando que, en general, se encuentren más emparentadas genéticamente dos subpoblaciones cuanta menos distancia medie entre ellas.

Con el fin de valorar las relaciones entre las poblaciones, se han calculado las **distancias genéticas**  $F_{ST}$  de Reynolds y  $R$  de Harpending y Jenkins. La distancia genética puede definirse como una medida de la afinidad existente entre diferentes taxones, poblaciones o individuos, calculada

a partir de datos con origen genético. Las matrices de distancia genética entre poblaciones para los dos coeficientes se han obtenido mediante los programas Phylip v3.69 (Felsenstein, 1989) y GeDis v1.8 (Peña et al. 2009).

Distancia genética de Reynolds (Reynolds et al. 1983). Asume que no existen mutaciones y que la deriva genética es el único factor responsable del cambio en las frecuencias de los haplotipos. A diferencia de otras distancias como la distancia genética de Nei y Li (1979), esta medida no asume un tamaño efectivo de población constante e igual para todas las poblaciones. La distancia genética de Reynolds se calcula mediante la fórmula:

constante e igual para todas las poblaciones. La distancia genética de Reynolds se calcula mediante la fórmula:

donde  $m$  se refiere al conjunto de los loci,  $i$  al conjunto de alelos,  $p_{1mi}$  las frecuencias del alelo  $i$  del locus  $m$  en la población 1 y  $p_{2mi}$  las frecuencias del alelo  $i$  del locus  $m$  en la población 2.

Distancia genética de Harpending y Jenkins (Harpending y Jenkins, 1973). Se basa en la matriz  $R$  de parentesco, que se define como:

locus  $m$  en la población 2.

- o Distancia genética de Harpending y Jenkins (Harper

Donde  $p_{ik}$ ,  $p_{jk}$  son las frecuencias del alelo  $k$  en las poblaciones  $i$  y  $j$  y  $\bar{p}_k$  es la frecuencia promedio del alelo  $k$ .

A partir de la matriz  $R$  de parentesco se construye una distancia genética mediante la fórmula:

$$d_{ij} = r_{ii} + r_{jj} - 2r_{ij}$$

Para comparar la matriz de distancias geográficas y las matrices de distancias genéticas  $R$  de Harpending y Jenkins y  $F_{ST}$  de Reynolds se ha realizado el **test de Mantel**, empleando el programa Past v2.16 (Hammer *et al*, 2001). Se trata de un método que aporta un coeficiente de correlación específico para datos interdependientes, como es el caso de los valores de una matriz. Puesto que no se conoce su distribución nula, es preciso construir una distribución esperada mediante un proceso de remuestreo. La asociación entre ambas matrices de distancias, y por tanto de los datos que contienen, puede resultar significativa (aceptando que la posición geográfica condiciona su patrimonio genético) o no.

Con los datos obtenidos en el laboratorio y los recopilados de la bibliografía se construyó una base de datos a partir de la cual se establecieron las distancias genéticas entre poblaciones mediante el cálculo del **coeficiente  $R$  de Harpending y Jenkins** (1973). Sobre la matriz de distancias obtenida de este modo se realizó un **análisis de escalamiento multidimensional** (MultiDimensional Scaling, **MDS**) (Kruskal 1964) con el paquete estadístico SPSS. Gracias a esta técnica se obtiene una representación en dos dimensiones de la similaridad o disimilaridad genética entre las poblaciones de interés en forma de distancias euclídeas, más fácilmente interpretables.

El escalamiento multidimensional es un conjunto de técnicas en las que, dada una ordenación de las distancias entre puntos, puede encontrarse una configuración de puntos en un espacio euclidiano de baja dimensión cuyas inter distancias euclidianas reproducen prácticamente la ordenación original (Shepard, 1962).

En algunos casos y con el objetivo de caracterizar en mayor profundidad los factores que pueden haber determinado la distribución de frecuencias de los haplotipos MAPT, se ha realizado un **análisis del centroide**

(Harpending y Ward, 1982), mediante el programa GeDis v1.8 (Peña *et al.* 2009). Este modelo valora la intensidad del flujo génico que ha recibido una población con relación al resto de poblaciones analizadas. Se establece una relación lineal simple entre la Heterocigosidad Observada en cada población y la distancia al centroide de cada una de las poblaciones ( $r_i$ ).

Esta distancia,  $r_i$ , se corresponde con la aportación de cada población a la varianza de Wahlund del conjunto y se calcula mediante la fórmula

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

donde  $p_i$  es la frecuencia del alelo en la población  $i$  y  $P$  es la frecuencia promedio en la población total. Es decir, el cociente entre la diferencia al cuadrado de la frecuencia alélica de cada población y la media aritmética de las frecuencias alélicas del conjunto de poblaciones analizadas [  $(p_i - P)^2$  ] y el valor máximo de la varianza [  $P(1 - P)$  ].

El cálculo se realiza para cada alelo y después es promediado para todos los loci analizados. Puede representarse en un gráfico en el que los valores de  $r_i$  aparecen en el eje de las abscisas y la heterocigosidad observada en el de las ordenadas.

En las ocasiones en las que se ha considerado necesario se ha realizado un **diagrama de Sephard** para comprobar de manera gráfica si la distribución de las poblaciones en las representaciones topológicas de las matrices de distancias eran o no adecuadas. En estos diagramas se representan las distancias genéticas frente a las distancias de la figura obtenida a partir del análisis de escalamiento multidimensional (MDS). Cuanto más lineal sea el resultado, mejor será la representación.

## Análisis de bases de datos de ADN antiguo.

### Origen de los datos.

Además de las muestras de ADN extraídas de poblaciones actuales, para el presente proyecto se han analizado los genotipos MAPT de varios cientos de muestras de ADN antiguo obtenidos a partir de diferentes bases de datos de Genoma Completo disponibles en la bibliografía (Olalde, 2018; Lazaridis, 2016; Lipson, 2017; Mathieson, 2018; Haak, 2015; Fu, 2016). De ellos, 539 han podido agruparse en muestras caracterizadas por una cierta homogeneidad temporal y espacial.

### Pre-procesamiento de los datos.

Las bases de datos se procesaron mediante el programa GWEdit (Peña, comunicación personal), que utiliza como entrada de datos el formato EIGENSOFT (Patterson et al, 2006).

Cada base de datos se compone de tres ficheros, que incluyen información sobre los individuos, sobre los SNPs analizados y sobre los genotipos obtenidos, respectivamente.

Al programa se le indicó que identificara los genotipos individuales para el SNP rs10514879, que, como se ha indicado anteriormente, discrimina entre los haplotipos H1 y H2 (Donnelly, 2010).

Los resultados genotípicos individuales fueron agrupados en 33 poblaciones caracterizadas geográfica y temporalmente en el contexto de Europa y Oeste de Asia. Para cada una de las 33 poblaciones se calculó la frecuencia del alelo MAPT\*H2.

### Análisis estadísticos.

Como se ha explicado en un apartado anterior, se ha estimado el grado de estructuración que presentan las poblaciones estudiadas mediante el Análisis Molecular de la Varianza (*Analysis of Molecular Variance*, AMOVA) empleando el programa Arlequín v3.5 (Excoffier y Lischer, 2010) para realizar los cálculos.

En este caso, para realizar la propuesta de los agrupamientos poblacionales se han empleado no sólo criterios geográficos sino también temporales; el rango de estudio ha sido el periodo comprendido entre los

años 8500 y 2000 antes del presente. Se han diferenciado tres grupos, cada uno de los cuales se extiende a lo largo de aproximadamente dos milenios.

También se ha realizado un análisis exploratorio de las **correlaciones parciales** mediante el programa Past 3.0 (Hammer *et al.*, 2001) para tratar de determinar el motivo de dichas diferencias.

En este tipo de análisis se estudia el efecto de la eliminación de ciertas variables que pueden perturbar la comprensión de la relación entre los factores de interés. Al eliminar el influjo de una tercera variable se pretende observar si las variables consideradas siguen manteniendo la relación detectada en análisis previos.

Mediante este método se han valorado las relaciones entre las frecuencias haplotípicas, antigüedad, posición geográfica y patrimonio cultural.

Finalmente se han elaborado una serie de **mapas sintéticos** con el programa MapChart (<https://mapchart.net>) para facilitar el estudio de la distribución geográfica del haplotipo MAPT\*H2 en diferentes periodos (8500 BP – 6000 BP, 6000 BP – 4000 BP y 4000 BP – 2000 BP).

# RESULTADOS

---





## Sección 1. Frecuencias alélicas y haplotípicas

Se han analizado tres marcadores bialélicos y cuatro STRPs en ocho muestras poblacionales: Gipuzkoa, Navarra, Residentes del País Vasco, Gitanos del País Vasco, Valencia, Chocó, Jujuy y Waorani.

### 1.1. Marcadores bialélicos

Los marcadores bialélicos son el indel rs72377086 y los SNPs rs10514879 y rs199451.

La inserción o delección de 238 pares de bases que caracteriza al marcador indel rs72377086, descrito por Baker *et al.* (1999), se define por la existencia de un alelo con la inserción (designado como 1), que identifica el haplotipo H1, y otro con la delección (designado como 2), que identifica el haplotipo H2.

Los alelos G y A del SNP rs10514879 se corresponden, según Donnelly *et al.* (2010), con los haplotipos H1 y H2 respectivamente. En el conjunto de 410 muestras de diferentes procedencias para las que se han identificado el genotipo tanto del indel como del SNP, el porcentaje de coincidencia en el diagnóstico de ambos marcadores es del 98,04%. En la **tabla R1** puede verse el grado de coincidencia desglosado por poblaciones. No se incluye la población de afrocolombianos de Chocó, debido a que no se analizó en esta muestra el SNP rs10514879.

**Tabla R1.** Porcentaje de coincidencia de los resultados obtenidos para los marcadores indel rs72377086 y SNP rs10514879, por poblaciones. N: número de individuos analizados para ambos marcadores.

	Gipuzkoa	Navarra	ResidentesPV	GitanosPV	Valencia	Jujuy	Waorani	Total
N	89	85	24	56	69	63	24	410
% coincidencia	100	99,40	100	98,21	92,75	98,41	97,91	98,04

Valencia, la población en la que el porcentaje de coincidencia ha sido menor, era también la que presentaba unas muestras de ADN más degradadas.

En el tercer marcador (rs199451) el alelo G es diagnóstico de los haplotipos sin duplicaciones (H1' y H2) mientras que el alelo A es diagnóstico del haplotipo H2D, la variante con duplicaciones del haplotipo H2. En combinación con alguno de los otros dos marcadores (indel o rs10514879), puede discriminarse entre los tres subhaplotipos (H1', H2' y H2D).

En la tabla R2 se muestran las frecuencias estimadas de los alelos de los 3 marcadores. Excepto en los Waorani, donde los alelos que caracterizan el haplotipo H1 se han fijado, la frecuencia máxima de la inserción (alelo 1 del

indel rs72377086) se encuentra en la población de Jujuy (0,993) y la mínima en Gipuzkoa (0,506). Esta misma tendencia se observa en el marcador rs10514879, para el cual la máxima frecuencia del alelo G aparece en Jujuy, (1,000) y la mínima en Gipuzkoa (0,500). De forma similar, para el marcador rs199451, la frecuencia máxima del alelo G aparece en la población de Jujuy (0,993) y la mínima en Gipuzkoa (0,511).

En general, las frecuencias más bajas de los alelos diagnóstico del haplotipo H1 se encuentran entre las poblaciones europeas, siendo consecuentemente más altas las frecuencias de los alelos que caracterizan el haplotipo H2.

Debido a la escasa cantidad de DNA disponible en las muestras de afrocolombianos del Chocó, no se pudieron realizar los análisis correspondientes a rs10514879. La asignación de los haplotipos MAPT se realizó en esta población combinando los resultados del indel y el SNP rs199451. Para el resto de poblaciones se utilizaron preferentemente los SNPs rs10514879 y rs199451.

En la **tabla R2** se observan diferencias en las frecuencias de *rs72377086\*1* y *rs10514879\*G* incluso en las poblaciones en las que no se observaron discordancias entre ambos marcadores. Ello es debido al diferente número de individuos analizado para cada marcador.

**Tabla R2.** Frecuencias estimadas ( $\pm$  error estándar) de 3 marcadores bialélicos en las 8 poblaciones analizadas. N: número de individuos analizados.

	Gipuzkoa	Navarra	ResidentesPV	GitanosPV	Valencia	Chocó	Jujuy	Waorani
indel rs72377086								
N	105	105	31	112	100	100	101	34
1	0,505	0,652	0,677	0,853	0,690	0,915	0,955	0,926
	$\pm 0,035$	$\pm 0,033$	$\pm 0,059$	$\pm 0,023$	$\pm 0,033$	$\pm 0,020$	$\pm 0,015$	$\pm 0,032$
2	0,495	0,348	0,323	0,147	0,310	0,085	0,045	0,074
	$\pm 0,035$	$\pm 0,033$	$\pm 0,059$	$\pm 0,023$	$\pm 0,033$	$\pm 0,020$	$\pm 0,015$	$\pm 0,032$
rs10514879								
N	90	90	23	67	91		70	24
G	0,506	0,689	0,717	0,873	0,742	-	0,993	1,000
	$\pm 0,037$	$\pm 0,035$	$\pm 0,066$	$\pm 0,029$	$\pm 0,032$		$\pm 0,007$	$\pm 0,000$
A	0,494	0,311	0,283	0,127	0,258	-	0,007	0,000
	$\pm 0,037$	$\pm 0,035$	$\pm 0,066$	$\pm 0,029$	$\pm 0,032$		$\pm 0,007$	$\pm 0,000$
rs199451								
N	90	90	23	67	91	35	70	24

G	0,511	0,722	0,739	0,873	0,764	0,971	0,993	1,000
	±0,037	±0,033	±0,065	±0,029	±0,031	±0,020	±0,007	±0,000
A	0,489	0,278	0,261	0,127	0,236	0,029	0,007	0,000
	±0,037	±0,033	±0,065	±0,029	±0,031	±0,020	±0,007	±0,000

Las frecuencias de los haplotipos MATP H1, H2' y H2D en las ocho poblaciones de este estudio se muestran en la **tabla R3**. Para su estima se tuvieron en cuenta aquellos individuos que mostraron un resultado legible para rs10514879 y para rs199451, excepto para las muestras de Chocó.

La frecuencia más alta del haplotipo *MAPT\*H1* se encuentra en la población Waorani, donde es el único haplotipo presente. En la otra población americana estudiada, Jujuy, la frecuencia de este haplotipo es muy cercana a 1 (0,993). La población afroamericana del Chocó presenta también una frecuencia muy alta de *MAPT\*H1* (0,915), seguida de la población gitana del País Vasco (0,873). Las otras poblaciones de la Península Ibérica de origen no vasco presentan unos valores intermedios (Valencia: 0,740; Residentes del País Vasco: 0,717) en tanto que los vascos de Navarra y de Gipuzkoa muestran las frecuencias más bajas de este haplotipo (0,690 y 0,500 respectivamente).

El haplotipo invertido y sin duplicación (*MAPT\*H2*) es el menos común de los tres. No aparece en las poblaciones americanas ni en la población de gitanos del País Vasco. Sus frecuencias más altas se encuentran en la población afrodescendiente del Chocó (0,056) y en la población vasca de Navarra (0,033). En Valencia y en los residentes del País Vasco existe con una frecuencia algo más baja (0,022) y en vascos de Gipuzkoa su presencia se reduce a una frecuencia de 0,006.

El haplotipo invertido con duplicación presenta su máxima frecuencia en la población vasca de Gipuzkoa (0,489). Navarra presenta una frecuencia de 0,278. Residentes vascos y Valencia tienen frecuencias de 0,261 y 0,236 respectivamente mientras que la población gitana del País Vasco muestra una frecuencia más baja (0,127). En Chocó se encuentra una frecuencia mucho más discreta (0,029) pero mayor que en las poblaciones indígenas americanas, Jujuy (0,007) y Waorani (0,000).

**Tabla R3.** Frecuencias estimadas ( $\pm$  error estándar) de los haplotipos MATP directo (H1), invertido (H2') e invertido con duplicación (H2D) en las 8 poblaciones analizadas. N: tamaño de la muestra expresado en número de individuos.

	Gipuzkoa	Navarra	ResidentesPV	GitanosPV	Valencia	Chocó	Jujuy	Waorani
--	----------	---------	--------------	-----------	----------	-------	-------	---------

Haplotipos MAPT en *Homo sapiens*

N	90	105	23	118	91	100	101	24
H1'	0,506	0,689	0,717	0,873	0,742	0,915	0,993	1,000
	±0,037	±0,035	±0,066	±0,029	±0,032	±0,028	±0,007	±0,000
H2'	0,006	0,033	0,022	0,000	0,022	0,056	0,000	0,000
	±0,006	±0,013	±0,022	±0,000	±0,011	±0,020	±0,000	±0,000
H2D	0,489	0,278	0,261	0,127	0,236	0,029	0,007	0,000
	±0,037	±0,033	±0,065	±0,029	±0,031	±0,020	±0,007	±0,000

En la **tabla R4** se presentan las frecuencias estimadas de los haplotipos MAPT de poblaciones incluidas en el proyecto *1000 Genomes*, para las que no se habían calculado previamente. Se observan unos patrones similares a los descritos en los párrafos anteriores; son las poblaciones de origen europeo las que presentan frecuencias más altas de los haplotipos invertidos (H2'+H2D), por encima de 0,200, con la excepción de los Fineses (0,107), aunque en ninguna de ellas son los haplotipos predominantes, mientras que en las poblaciones de origen africano, americano y asiático, a excepción de aquellas poblaciones con un mayor índice de mestizaje con poblaciones de origen europeo, la frecuencia de los haplotipos invertidos no supera el 10%.

**Tabla R4.** Frecuencias estimadas ( $\pm$  error estándar) de los haplotipos MAPT directo (H1), invertido (H2') e invertido con duplicación (H2D) en las 26 poblaciones incluidas en el proyecto 1000 Genomes. N: tamaño de la muestra expresado en número de individuos.

	Afroamer. Barbados	Afroamer. USA	Bengalí Bangladesh	Chinos Dai	Residentes Utah	Han Beijing	Han S. China	Colombianos Medellín	Esan Nigeria
N	192	132	170	198	198	206	216	188	198
H1'	0,953	0,879	0,953	1,000	0,793	1,000	1,000	0,819	1,000
	±0,011	±0,020	±0,011	±0,000	±0,020	±0,000	±0,000	±0,020	±0,000
H2'	0,026	0,045	0,006	0,000	0,010	0,000	0,000	0,005	0,000
	±0,008	±0,013	±0,004	±0,000	±0,005	±0,000	±0,000	±0,004	±0,000
H2D	0,021	0,076	0,041	0,000	0,197	0,000	0,000	0,176	0,000
	±0,007	±0,016	±0,011	±0,000	±0,020	±0,000	±0,000	±0,020	±0,000

	Fineses	Británicos	Gujarati USA	Gambia	Españoles	Telugu UK	Japoneses	Kinh Vietnam	Luhya Kenia
N	196	184	210	226	212	206	208	202	202
H1'	0,893	0,755	0,910	0,987	0,736	0,951	0,995	1,000	0,980

	$\pm 0,016$	$\pm 0,022$	$\pm 0,014$	$\pm 0,005$	$\pm 0,021$	$\pm 0,011$	$\pm 0,003$	$\pm 0,000$	$\pm 0,007$
H2'	0,005	0,022	0,010	0,009	0,005	0,010	0,000	0,000	0,020
	$\pm 0,004$	$\pm 0,008$	$\pm 0,005$	$\pm 0,004$	$\pm 0,003$	$\pm 0,005$	$\pm 0,000$	$\pm 0,000$	$\pm 0,007$
H2D	0,102	0,223	0,081	0,004	0,259	0,039	0,005	0,000	0,000
	$\pm 0,015$	$\pm 0,022$	$\pm 0,013$	$\pm 0,003$	$\pm 0,021$	$\pm 0,010$	$\pm 0,003$	$\pm 0,000$	$\pm 0,000$

	Mende Sierra Leona	Mejicanos California	Peruanos	Punjab	Portorriqueños	Tamiles UK	Toscanos	Yoruba
N	170	134	172	192	208	206	214	218
H1'	0,994	0,828	0,930	0,901	0,803	0,971	0,636	0,995
	$\pm 0,004$	$\pm 0,023$	$\pm 0,014$	$\pm 0,015$	$\pm 0,020$	$\pm 0,008$	$\pm 0,023$	$\pm 0,003$
H2'	0,006	0,022	0,000	0,031	0,005	0,005	0,056	0,005
	$\pm 0,004$	$\pm 0,009$	$\pm 0,000$	$\pm 0,009$	$\pm 0,003$	$\pm 0,003$	$\pm 0,011$	$\pm 0,003$
H2D	0,000	0,149	0,070	0,068	0,192	0,024	0,308	0,000
	$\pm 0,000$	$\pm 0,022$	$\pm 0,014$	$\pm 0,013$	$\pm 0,019$	$\pm 0,008$	$\pm 0,022$	$\pm 0,000$

## 1.2. Marcadores STRPs

Las frecuencias alélicas para los cuatro microsatélites analizados, así como sus errores estándar se muestran en la **tabla R5**. En este estudio se han identificado un total de 11 alelos para el locus MAPT07, 6 para MAPT08 y MAPT09 y 14 para MAPT14. En la bibliografía aparecen descritos 11 para MAPT07, 7 para MAPT08, 12 para MAPT09 y 14 para MAPT14. En el caso de MAPT08 el alelo no identificado es el 13 (211 bp), sólo presente en los Masai. En MAPT09 hay una diferencia de seis alelos, probablemente debido a que para este marcador existen sólo de dos a cuatro alelos predominantes en todas las poblaciones y el resto son muy poco frecuentes. El alelo 12 (111 bp) de MAPT14 pertenecen también al grupo de alelos muy poco frecuentes.

Los dos alelos más frecuentes del marcador MAPT07 en todas las poblaciones son *MAPT07\*18* (207 bp) y *MAPT07\*19* (209 bp). Excepto en Chocó, la suma de ambos alelos en todas poblaciones representa más del 80% de la variabilidad. El alelo *MAPT07\*18* es el mayoritario entre los Gitanos del País Vasco (0,534), en Jujuy (0,416) y en Chocó (0,362) mientras que *MAPT07\*19* es el mayoritario en ambas poblaciones vascas (0,682 en Gipuzkoa y 0,618 en Navarra), en Residentes del País Vasco (0,500), en Valencia (0,559) y en Waorani (0,638).

Los alelos más frecuentes del marcador MAPT08 son el 9 (195 pb) y el 10 (199 pb). En las poblaciones vascas el alelo más frecuente es *MAPT08\*10* (199 pb) (0,693 en Gipuzkoa y 0,668 en Navarra). En estas poblaciones también aparece el alelo 9 (195pb) (0,193 y 0,179) y el alelo 11 (203pb) (0,115 y 0,153). En Valencia también aparecen estos tres alelos con proporciones similares (0,515 *MAPT08\*10*, 0,318 *MAPT08\*9* y 0,167 *MAPT08\*11*). Entre los Gitanos vascos, la diferencia entre las frecuencias de los alelos 9 y 10 no es tan notable (0,496 y 0,465 respectivamente) y además del alelo 11 (0,035) aparece también un nuevo alelo minoritario, *MAPT08\*12* (207 pb, 0,004). En la población afrodescendiente de Chocó la frecuencia del alelo 9 es de 0,627 y la del alelo 10 de 0,313. Aparecen además otros tres alelos minoritarios *MAPT08\*7* (187pb) (0,020), *MAPT08\*8* (191 pb) (0,027) y *MAPT08\*11* (0,013). De las poblaciones estudiadas, Jujuy es la que menos diversidad de alelos muestra; sólo se han hallado el alelo 9 (0,370) y el 10 (0,630). En la otra población indígena americana, los Waorani, se han encontrado 4 alelos, *MAPT08\*8* (0,190), *MAPT08\*9* (0,466), *MAPT08\*10* (0,293) y *MAPT08\*11* (0,052).

Excepto en las poblaciones de origen europeo, el alelo 22 (125 pb) del microsatélite MAPT09 presenta frecuencias superiores a 0,8 (0,881 Gitanos de País Vasco; 0,925 afrodescendientes de Chocó; 0,961 en Jujuy; en la muestra de Waorani es el único alelo presente). En las poblaciones vascas de Gipuzkoa y de Navarra, las frecuencias de este alelo (0,522 y 0,702) permiten frecuencias más altas del alelo 14 (109pb) que presenta una frecuencia de 0,452 y 0,287 respectivamente. *MAPT09\*14* aparece también en residentes vascos (0,297) y en Valencia (0,190) y con frecuencias mucho más bajas en Gitanos vascos (0,114), Chocó (0,052) y Jujuy (0,034). En Gipuzkoa, Navarra y Valencia aparece también el alelo 13 (107pb) con frecuencia 0,011, 0,005 y 0,044 respectivamente. En Gipuzkoa, Gitanos vascos, Chocó y Jujuy aparece también el alelo 21 (123pb) y en Navarra, Residentes vascos y Valencia *MAPT09\*23* (129 pb). *MAPT09\*24* sólo aparece en Chocó (0,006).

El microsatélite MAPT14 muestra una mayor diversidad de alelos en todas las poblaciones y una distribución de los mismos más equilibrada. En casi todas las poblaciones el alelo más frecuente es *MAPT14\*17* (131 pb) con la excepción de la población gitana, donde es más frecuente *MAPT14\*18* (135 pb) (0,283) y en Residentes vascos, donde los alelos 17 y 18 tienen la misma frecuencia (0,205).

**Tabla R5.** Frecuencias estimadas ( $\pm$  error estándar) de 4 marcadores STRs en las 8 poblaciones analizadas. N: tamaño de la muestra expresado en número de individuos.

	Gipuzkoa	Navarra	ResidentesPV	GitanosPV	Valencia	Chocó	Jujuy	Waorani
N	(105)	(105)	(35)	(118)	(118)	(100)	(101)	(34)

## MAPT07

15	0,000	0,000	0,000	0,004	0,000	0,000	0,000	0,000
	±0,000	±0,000	±0,000	±0,004	±0,000	±0,000	±0,000	±0,000
16	0,025	0,018	0,030	0,038	0,000	0,061	0,006	0,000
	±0,011	±0,009	±0,020	±0,012	±0,000	±0,017	±0,005	±0,000
17	0,030	0,012	0,106	0,000	0,053	0,036	0,062	0,000
	±0,012	±0,008	±0,037	±0,000	±0,015	±0,013	±0,017	±0,000
18	0,207	0,288	0,333	0,534	0,340	0,362	0,416	0,310
	±0,028	±0,031	±0,056	±0,032	±0,031	±0,034	±0,035	±0,056
19	0,682	0,618	0,500	0,312	0,559	0,291	0,393	0,638
	±0,032	±0,034	±0,060	±0,030	±0,032	±0,032	±0,034	±0,058
20	0,040	0,041	0,030	0,064	0,037	0,087	0,022	0,000
	±0,014	±0,014	±0,020	±0,016	±0,012	±0,020	±0,010	±0,000
21	0,010	0,024	0,000	0,013	0,011	0,117	0,090	0,052
	±0,007	±0,011	±0,000	±0,007	±0,007	±0,023	±0,020	±0,027
22	0,000	0,000	0,000	0,009	0,000	0,015	0,011	0,000
	±0,000	±0,000	±0,000	±0,006	±0,000	±0,009	±0,007	±0,000
23	0,000	0,000	0,000	0,026	0,000	0,026	0,000	0,000
	±0,000	±0,000	±0,000	±0,010	±0,000	±0,011	±0,000	±0,000
24	0,005	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	±0,005	±0,000	±0,000	±0,000	±0,000	±0,000	±0,000	±0,000
26	0,000	0,000	0,000	0,000	0,000	0,005	0,000	0,000
	±0,000	±0,000	±0,000	±0,000	±0,000	±0,005	±0,000	±0,000

## MAPT08

7	0,000	0,000	0,016	0,000	0,000	0,020	0,000	0,000
	±0,000	±0,000	±0,015	±0,000	±0,000	±0,010	±0,000	±0,000
8	0,000	0,000	0,016	0,000	0,000	0,027	0,000	0,190
	±0,000	±0,000	±0,015	±0,000	±0,000	±0,011	±0,000	±0,048
9	0,193	0,179	0,281	0,496	0,318	0,627	0,370	0,466
	±0,027	±0,026	±0,054	±0,033	±0,030	±0,034	±0,034	±0,060
10	0,693	0,668	0,609	0,465	0,515	0,313	0,630	0,293
	±0,032	±0,032	±0,058	±0,032	±0,033	±0,033	±0,034	±0,055
11	0,115	0,153	0,078	0,035	0,167	0,013	0,000	0,052
	±0,022	±0,025	±0,032	±0,012	±0,024	±0,008	±0,000	±0,027

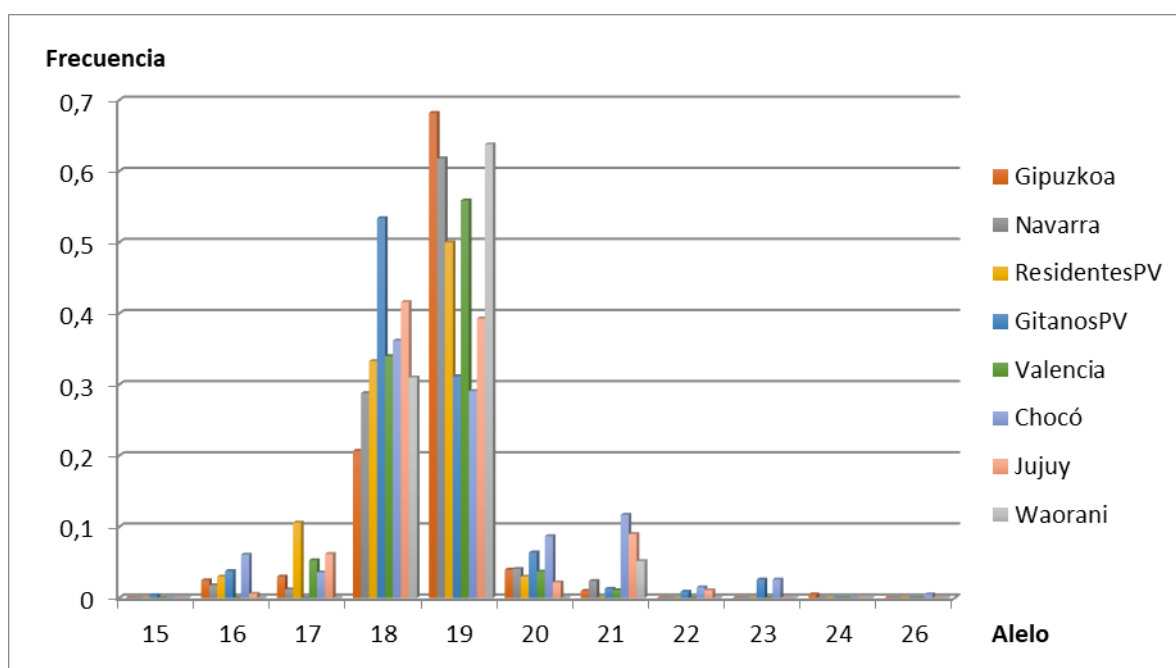
Haplotipos MAPT en *Homo sapiens*

12	0,000	0,000	0,000	0,004	0,000	0,000	0,000	0,000
	±0,000	±0,000	±0,000	±0,004	±0,000	±0,000	±0,000	±0,000
MAPT09								
13	0,011	0,005	0,000	0,000	0,044	0,000	0,000	0,000
	±0,007	±0,005	±0,000	±0,000	±0,013	±0,000	±0,000	±0,000
14	0,452	0,287	0,297	0,114	0,190	0,052	0,034	0,000
	±0,034	±0,031	±0,055	±0,021	±0,026	±0,016	±0,013	±0,000
21	0,016	0,000	0,000	0,004	0,000	0,017	0,006	0,000
	±0,009	±0,000	±0,000	±0,004	±0,000	±0,009	±0,005	±0,000
22	0,522	0,702	0,688	0,881	0,759	0,925	0,961	1,000
	±0,034	±0,032	±0,055	±0,021	±0,028	±0,019	±0,014	±0,000
23	0,000	0,005	0,016	0,000	0,006	0,000	0,000	0,000
	±0,000	±0,005	±0,015	±0,000	±0,005	±0,000	±0,000	±0,000
24	0,000	0,000	0,000	0,000	0,000	0,006	0,000	0,000
	±0,000	±0,000	±0,000	±0,000	±0,000	±0,005	±0,000	±0,000
MAPT14								
9	0,050	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	±0,015	±0,000	±0,000	±0,000	±0,000	±0,000	±0,000	±0,000
10	0,151	0,060	0,091	0,035	0,012	0,032	0,000	0,000
	±0,025	±0,016	±0,034	±0,012	±0,007	±0,012	±0,000	±0,000
11	0,000	0,000	0,000	0,000	0,000	0,000	0,006	0,000
	±0,000	±0,000	±0,000	±0,000	±0,000	±0,000	±0,005	±0,000
13	0,016	0,000	0,000	0,000	0,000	0,022	0,006	0,000
	±0,009	±0,000	±0,000	±0,000	±0,000	±0,010	±0,005	±0,000
14	0,047	0,120	0,068	0,139	0,113	0,129	0,080	0,000
	±0,015	±0,022	±0,030	±0,023	±0,021	±0,024	±0,019	±0,000
15	0,078	0,102	0,250	0,178	0,131	0,183	0,216	0,397
	±0,019	±0,021	±0,052	±0,025	±0,022	±0,027	±0,029	±0,059
16	0,057	0,072	0,000	0,178	0,083	0,151	0,111	0,000
	±0,016	±0,018	±0,000	±0,025	±0,018	±0,025	±0,022	±0,000
17	0,318	0,283	0,205	0,126	0,345	0,204	0,296	0,448
	±0,032	±0,031	±0,048	±0,022	±0,031	±0,028	±0,032	±0,060
18	0,177	0,235	0,205	0,283	0,196	0,167	0,216	0,103
	±0,026	±0,029	±0,048	±0,029	±0,026	±0,026	±0,029	±0,037



19	0,141	0,120	0,091	0,043	0,101	0,054	0,062	0,017
	±0,024	±0,022	±0,034	±0,013	±0,020	±0,016	±0,017	±0,016
20	0,010	0,006	0,091	0,000	0,012	0,059	0,006	0,000
	±0,007	±0,005	±0,034	±0,000	±0,007	±0,017	±0,005	±0,000
21	0,000	0,000	0,000	0,000	0,006	0,000	0,000	0,000
	±0,000	±0,000	±0,000	±0,000	±0,005	±0,000	±0,000	±0,000
22	0,000	0,000	0,000	0,004	0,000	0,000	0,000	0,034
	±0,000	±0,000	±0,000	±0,004	±0,000	±0,000	±0,000	±0,022
23	0,000	0,000	0,000	0,013	0,000	0,000	0,000	0,000
	±0,000	±0,000	±0,000	±0,007	±0,000	±0,000	±0,000	±0,000

Como puede observarse en las **figuras R1 a R4**, la distribución de frecuencias de MAPT07 y MAPT08 es básicamente unimodal, la de MAPT09 es claramente bimodal y la de MAPT14, más compleja, podría considerarse trimodal.



**Figura R1.** Histograma de frecuencias de MAPT07 en las 8 poblaciones analizadas.

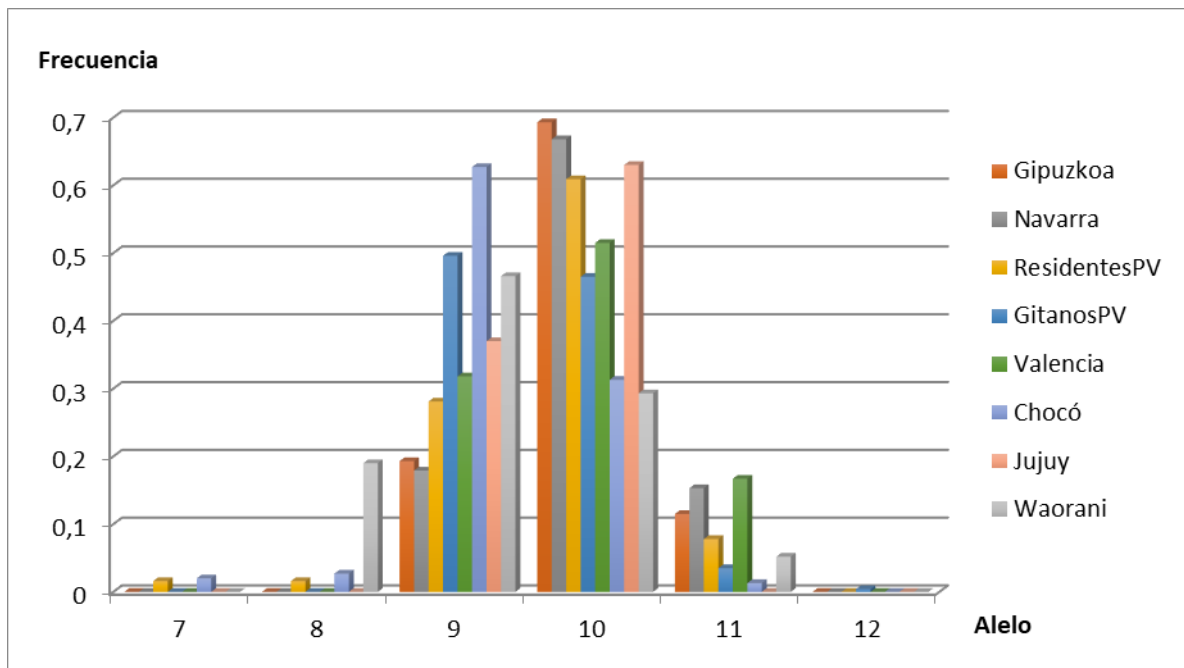


Figura R2. Histograma de frecuencias de MAPT08 en las 8 poblaciones analizadas.

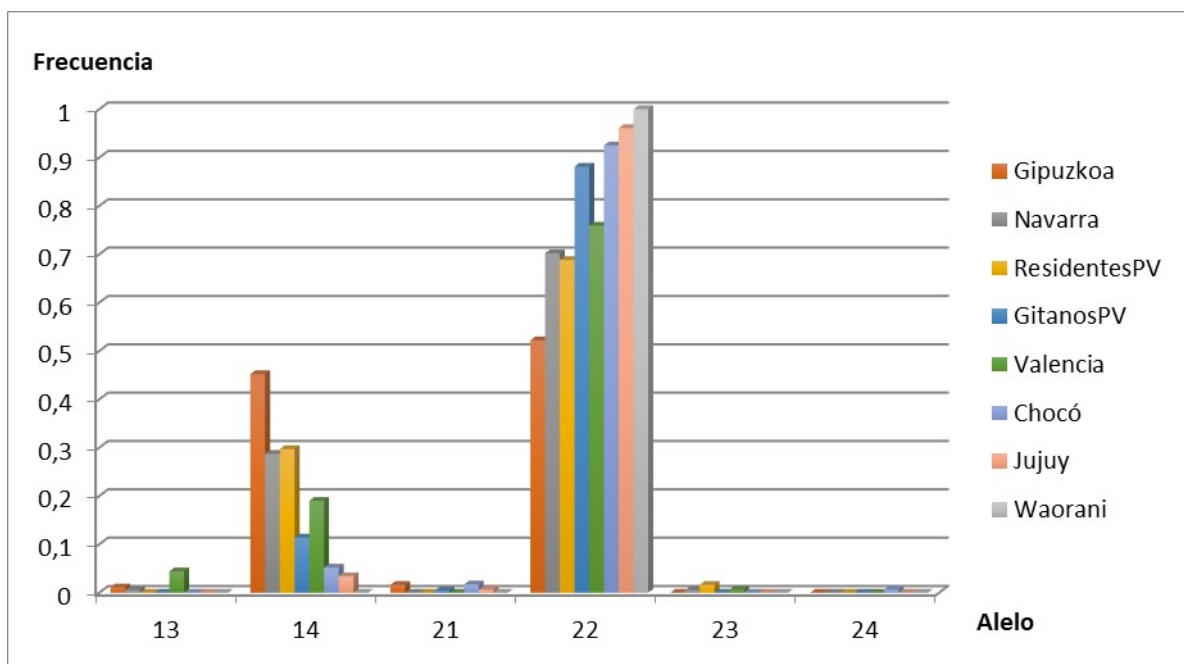


Figura R3. Histograma de frecuencias de MAPT09 en las 8 poblaciones analizadas.

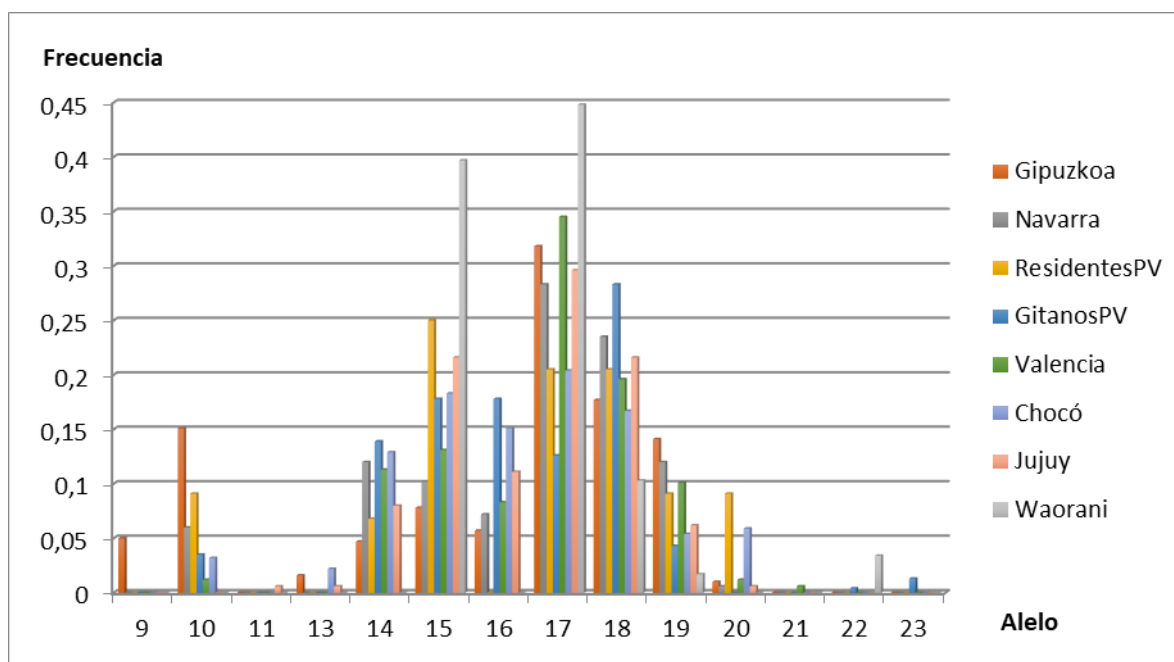


Figura R4. Histograma de frecuencias de MAPT14 en las 8 poblaciones analizadas.

### 1.3. Test de equilibrio HW

En la **tabla R6** se detallan los resultados del test de equilibrio Hardy-Weinberg. Excepto para el microsatélite MAPT14, existen pocos valores significativos. Tras aplicar la corrección de Bonferroni sólo se mantienen los valores de  $p < 0,05$  en Jujuy para rs72377086, en Gipuzkoa para MAPT09 y en 6 muestras para MAPT14. Por tanto, las poblaciones analizadas se encuentran en equilibrio de Hardy-Weinberg, si bien parece existir algún factor perturbador en MAPT14.

Destaca entre los valores significativos el caso de MAPT09 en Gipuzkoa, ya que como puede observarse, los valores de heterocigosidad observada y esperada son muy similares. El motivo es la existencia de varios valores de frecuencias genotípicas observadas y esperadas por debajo de 5 y más concretamente de entre 0 y 1. Tanto el método del test exacto de  $p$  como el basado en el test  $\chi^2$  han dado un valor de  $p < 0,0001$ , incluso aplicando la corrección de Hedrick (2004). Si no se consideran estos genotipos, o bien se agrupan en una sola clase, el test no resulta significativo.

**Tabla R6.** Test de equilibrio Hardy-Weinberg para los 3 marcadores bialélicos y los 4 marcadores STRs en las 8 poblaciones analizadas. Se muestran en negrita los valores significativos. \*: Significación después de aplicar la corrección de Bonferroni.

	Gipuzkoa	Navarra	ResidentesPV	GitanosPV	Valencia	Chocó	Jujuy	Waorani
rs72377086								
Het. obs.	0,476	0,410	0,452	0,205	0,400	0,130	0,030	0,147
Het. esp.	0,502	0,456	0,444	0,252	0,428	0,156	0,086	0,138
P	0,696	0,386	1,000	0,061	0,492	0,139	<b>0,000*</b>	1,000
rs10514879								
Het. obs.	0,484	0,400	0,333	0,134	0,353	-	0,000	0,040
Het. esp.	0,503	0,439	0,422	0,200	0,397	-	0,000	0,040
P	0,834	0,471	0,346	<b>0,027</b>	0,406	-	1,000	1,000
rs199451								
Het. obs.	0,456	0,386	0,348	0,191	0,326	0,057	0,015	0,000
Het. esp.	0,503	0,409	0,394	0,220	0,366	0,056	0,015	0,000
p	0,406	0,607	0,609	0,264	0,387	1,000	1,000	1,000
MAPT07								
Het. obs.	0,556	0,588	0,606	0,578	0,596	0,684	0,730	0,724
Het. esp.	0,491	0,536	0,635	0,609	0,571	0,761	0,664	0,503
p	0,729	0,117	0,197	0,059	0,465	0,135	0,915	<b>0,020</b>
MAPT08								
Het. obs.	0,484	0,474	0,469	0,509	0,525	0,507	0,481	0,793
Het. esp.	0,471	0,501	0,552	0,547	0,609	0,511	0,469	0,670
p	0,728	0,298	0,312	0,255	<b>0,027</b>	0,110	1,000	0,367
MAPT09								
Het. obs.	0,533	0,404	0,406	0,186	0,392	0,125	0,079	-
Het. esp.	0,527	0,427	0,446	0,211	0,388	0,140	0,076	-
p	<b>0,000*</b>	0,153	0,351	0,270	0,070	0,067	1,000	-
MAPT14								
Het. obs.	0,379	0,422	0,393	0,591	0,500	0,559	0,765	0,690
Het. esp.	0,817	0,821	0,845	0,822	0,802	0,854	0,801	0,641
p	<b>0,000*</b>	<b>0,000*</b>	<b>0,000*</b>	<b>0,000*</b>	<b>0,000*</b>	<b>0,000*</b>	0,626	0,946

#### 1.4. Desequilibrio de ligamiento

Las **tablas R7 a R14** muestran los resultados del análisis de desequilibrio de ligamiento (LD) entre los siete marcadores examinados para las ocho poblaciones de este estudio. Estos datos han revelado varias asociaciones. En concreto,

- Se han detectado valores significativos de desequilibrio de ligamiento entre todos los marcadores en la población de **Gipuzkoa**, salvo entre el microsatélite MAPT07 con MAPT08 y con MAPT14.
- En **Navarra** se han detectado valores significativos de desequilibrio de ligamiento entre el indel y el resto de marcadores, excepto con MAPT08; entre MAPT07 con todos los marcadores salvo MAPT08 y MAPT14; MAPT08 sólo muestra LD con MAPT09; entre MAPT09 y todo el resto de marcadores; entre MAPT14 con el resto de marcadores salvo con MAPT07 y MAPT08; los SNPs muestran LD con todos los marcadores salvo con MAPT08.
- En la población de **Residentes vascos** el indel muestra LD con todo el resto de marcadores; MAPT07 con todos salvo con MAPT14 y el SNP rs199451; MAPT08 sólo con el indel y MAPT07; MAPT09 con todos los marcadores salvo con MAPT08; MAPT14 con el indel, con MAPT09 y con el SNP rs10514879; el SNP rs10514879 con todos los marcadores salvo con MAPT08; el SNP rs199451 con el indel, con MAPT09 y con el el SNP rs10514879.
- En la población de **Gitanos vascos** todos los marcadores están en desequilibrio de ligamiento unos con otros salvo el microstélite MAPT08 con MAPT09.
- En **Valencia** el marcador indel está en LD con ambos SNPs y con el microsatélite MAPT09. MAPT07 y MAPT 08 sólo muestran LD uno con el otro. MAPT09, además de con el indel, presenta LD con los SNPs; MATP14 no presenta LD.
- En la población afrocolombiana de **Chocó**, el indel muestra LD con MAPT09 y con el SNP rs199451; MAPT09, además de con el indel, también presenta LD con el SNP rs199451; El SNP rs199451 muestra por tanto LD con el indel y con el microsatélite MAPT09.
- En **Jujuy** sólo se observa LD entre los microsatélites MAPT08 y MAPT14;
- En los **Worani** sólo se observa LD entre los microsatélites MAPT07 y MAPT08.

**Tabla R7.** Análisis del desequilibrio de ligamiento entre los 7 marcadores analizados en la muestra de Gipuzkoa. Se muestran en negrita los valores de p significativos.

	rs72377086	MAPT07	MAPT08	MAPT09	MAPT14	rs10514879	rs199451
		<b>0.00000</b>	<b>0.00098</b>	<b>0.00000</b>	<b>0.00000</b>	<b>0.00000</b>	<b>0.00000</b>
rs72377086	*	<b>± 0.00000</b>	<b>± 0.00098</b>	<b>± 0.00000</b>	<b>± 0.00000</b>	<b>± 0.00000</b>	<b>± 0.00000</b>
			0.95308	<b>0.00000</b>	0.17889	<b>0.00000</b>	<b>0.00000</b>
MAPT07		*	± 0.00768	<b>± 0.00000</b>	± 0.01291	<b>± 0.00000</b>	<b>± 0.00000</b>
				<b>0.00098</b>	<b>0.02835</b>	<b>0.00000</b>	<b>0.00000</b>
MAPT08			*	<b>± 0.00098</b>	<b>± 0.00561</b>	<b>± 0.00000</b>	<b>± 0.00000</b>
					<b>0.00000</b>	<b>0.00000</b>	<b>0.00000</b>
MAPT09				*	<b>± 0.00000</b>	<b>± 0.00000</b>	<b>± 0.00000</b>
						<b>0.00000</b>	<b>0.00000</b>
MAPT14					*	<b>± 0.00000</b>	<b>± 0.00000</b>
							<b>0.00000</b>
rs10514879						*	<b>± 0.00000</b>
rs199451							*

**Tabla R8.** Análisis del desequilibrio de ligamiento entre los 7 marcadores analizados en la muestra de Navarra. Se muestran en negrita los valores de p significativos.

	rs72377086	MAPT07	MAPT08	MAPT09	MAPT14	rs10514879	rs199451
		<b>0.00098</b>	0.07136	<b>0.00000</b>	<b>0.00098</b>	<b>0.00000</b>	<b>0.00000</b>
rs72377086	*	<b>±0.00098</b>	±0.00696	<b>±0.00000</b>	<b>±0.00098</b>	<b>±0.00000</b>	<b>±0.00000</b>
			0.33822	<b>0.00587</b>	0.42229	<b>0.00098</b>	<b>0.00293</b>
MAPT07		*	±0.01704	<b>±0.00260</b>	±0.01460	<b>±0.00098</b>	<b>±0.00164</b>
				<b>0.01955</b>	0.27664	0.05963	0.09384
MAPT08			*	<b>±0.00386</b>	±0.01301	±0.00813	±0.00870
					<b>0.00293</b>	<b>0.00000</b>	<b>0.00000</b>
MAPT09				*	<b>±0.00164</b>	<b>±0.00000</b>	<b>±0.00000</b>
						<b>0.00196</b>	<b>0.00000</b>
MAPT14					*	<b>±0.00136</b>	<b>±0.00000</b>
							<b>0.00000</b>
rs10514879						*	<b>±0.00000</b>
rs199451							*

**Tabla R9.** Análisis del desequilibrio de ligamiento entre los 7 marcadores analizados en la muestra de Residentes en el País Vasco. Se muestran en negrita los valores de p significativos.

	rs72377086	MAPT07	MAPT08	MAPT09	MAPT14	rs10514879	rs199451
		<b>0.04692</b>	<b>0.04203</b>	<b>0.00000</b>	<b>0.00587</b>	<b>0.00000</b>	<b>0.00000</b>
rs72377086	*	<b>±0.00657</b>	<b>±0.00669</b>	<b>±0.00000</b>	<b>±0.00219</b>	<b>±0.00000</b>	<b>±0.00000</b>
			<b>0.00293</b>	<b>0.01857</b>	0.51124	<b>0.04399</b>	0.06843
MAPT07		*	<b>±0.00164</b>	<b>±0.00538</b>	±0.01438	<b>±0.00504</b>	±0.00807
				0.13294	0.66178	0.06354	0.09873
MAPT08			*	±0.00939	±0.01441	±0.00847	±0.00877
					<b>0.00489</b>	<b>0.00000</b>	<b>0.00000</b>
MAPT09				*	<b>±0.00203</b>	<b>±0.00000</b>	<b>±0.00000</b>
						<b>0.00391</b>	0.08113
MAPT14					*	<b>±0.00233</b>	±0.00998
							<b>0.00000</b>
rs10514879						*	<b>±0.00000</b>
rs199451							*

**Tabla R10.** Análisis del desequilibrio de ligamiento entre los 7 marcadores analizados en la muestra de Gitanos del País Vasco. Se muestran en negrita los valores de p significativos.

	rs72377086	MAPT07	MAPT08	MAPT09	MAPT14	rs10514879	rs199451
		<b>0.00000</b>	<b>0.02151</b>	<b>0.00000</b>	<b>0.00000</b>	<b>0.00000</b>	<b>0.00000</b>
rs72377086	*	<b>±0.00000</b>	<b>±0.00426</b>	<b>±0.00000</b>	<b>±0.00000</b>	<b>±0.00000</b>	<b>±0.00000</b>
			<b>0.00000</b>	<b>0.00489</b>	<b>0.00293</b>	<b>0.00000</b>	<b>0.00000</b>
MAPT07		*	<b>±0.00000</b>	<b>±0.00203</b>	<b>±0.00164</b>	<b>±0.00000</b>	<b>±0.00000</b>
				0.05279	<b>0.00000</b>	<b>0.00196</b>	<b>0.00587</b>
MAPT08			*	±0.00730	<b>±0.00000</b>	<b>±0.00136</b>	<b>±0.00219</b>
					<b>0.04594</b>	<b>0.00000</b>	<b>0.00000</b>
MAPT09				*	<b>±0.00701</b>	<b>±0.00000</b>	<b>±0.00000</b>
						<b>0.00098</b>	<b>0.00098</b>
MAPT14					*	<b>±0.00098</b>	<b>±0.00098</b>
							<b>0.00000</b>
rs10514879						*	<b>±0.00000</b>
rs199451							*

**Tabla R11.** Análisis del desequilibrio de ligamiento entre los 7 marcadores analizados en la muestra de Valencia. Se muestran en negrita los valores de p significativos.

	rs72377086	MAPT07	MAPT08	MAPT09	MAPT14	rs10514879	rs199451
		0.99316	0.22776	<b>0.00000</b>	0.39492	<b>0.00000</b>	<b>0.00000</b>
rs72377086	*	±0.00271	±0.01096	<b>±0.00000</b>	±0.01484	<b>±0.00000</b>	<b>±0.00000</b>
			<b>0.03128</b>	0.92082	0.51124	0.69892	0.62561
MAPT07		*	<b>±0.00409</b>	±0.00684	±0.01879	±0.01545	±0.01457
				0.20723	0.33040	0.09091	0.25415
MAPT08			*	±0.01265	±0.01252	±0.00807	±0.01396
					0.66373	<b>0.00000</b>	<b>0.00000</b>
MAPT09				*	±0.01394	<b>±0.00000</b>	<b>±0.00000</b>
						0.41642	0.51320
MAPT14					*	±0.01507	±0.01500
							<b>0.00000</b>
rs10514879						*	<b>±0.00000</b>
rs199451							*

**Tabla R12.** Análisis del desequilibrio de ligamiento entre los 6 marcadores analizados en la muestra de afrocolombianos de Chocó. Se muestran en negrita los valores de p significativos.

	rs72377086	MAPT07	MAPT08	MAPT09	MAPT14	rs199451
		0.92571	0.53177	<b>0.00098</b>	0.20528	<b>0.00684</b>
rs72377086	*	±0.00728	±0.01614	<b>±0.00098</b>	±0.01239	<b>±0.00271</b>
			0.75660	0.10557	0.81623	0.69795
MAPT07		*	±0.01402	±0.01158	±0.01051	±0.01530
				0.14565	0.25024	0.33920
MAPT08			*	±0.01083	±0.01520	±0.01523
					0.67253	<b>0.00196</b>
MAPT09				*	±0.01329	<b>±0.00136</b>
						0.09482
MAPT14					*	±0.00950
rs199451						*



**Tabla R13.** Análisis del desequilibrio de ligamiento entre los 7 marcadores analizados en la muestra de Jujuy. Se muestran en negrita los valores de p significativos.

	rs72377086	MAPT07	MAPT08	MAPT09	MAPT14	rs10514879	rs199451
		1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
rs72377086	*	±0.00000	±0.00000	±0.00000	±0.00000	±0.00000	±0.00000
			0.21896	0.07234	0.18768	1.00000	1.00000
MAPT07		*	±0.01247	±0.00766	±0.01299	±0.00000	±0.00000
				0.36461	<b>0.02248</b>	1.00000	1.00000
MAPT08			*	±0.01423	<b>±0.00421</b>	±0.00000	±0.00000
					0.23265	1.00000	1.00000
MAPT09				*	±0.01314	±0.00000	±0.00000
						1.00000	1.00000
MAPT14					*	±0.00000	±0.00000
							1.00000
rs10514879						*	±0.00000
rs199451							*

**Tabla R14.** Análisis del desequilibrio de ligamiento entre los 7 marcadores analizados en la muestra de Waorani. Se muestran en negrita los valores de p significativos.

	rs72377086	MAPT07	MAPT08	MAPT09	MAPT14	rs10514879	rs199451
		0.24536	0.48094	1.00000	0.69013	1.00000	1.00000
rs72377086	*	±0.01529	±0.01403	±0.00000	±0.01637	±0.00000	±0.00000
			<b>0.04399</b>	1.00000	0.07136	1.00000	1.00000
MAPT07		*	<b>±0.00626</b>	±0.00000	±0.00872	±0.00000	±0.00000
				1.00000	0.00098	1.00000	1.00000
MAPT08			*	±0.00000	±0.00098	±0.00000	±0.00000
					1.00000	1.00000	1.00000
MAPT09				*	±0.00000	±0.00000	±0.00000
						1.00000	1.00000
MAPT14					*	±0.00000	±0.00000
							1.00000
rs10514879						*	±0.00000
rs199451							*

En conjunto, se observa una distribución irregular de los valores significativos, que puede verse afectada por la existencia de dos modelos de cromosomas, H1 y H2, que no experimentan entrecruzamiento y por la existencia de alelos fijados o próximos a la fijación en los marcadores bialélicos de Jujuy y Waorani, haciendo inviable el test. En todo caso, es notablemente mayor el número de combinaciones con desequilibrio de ligamiento significativo en los marcadores bialélicos y en MAPT09, en relación a los otros 3 microsatélites. Ello es probablemente debido a la existencia de alelos específicos de cada uno de los haplotipos MAPT, H1 y H2. En los SNPs y el indel es evidente, ya que son diagnósticos, pero también se observa claramente en MAPT09.

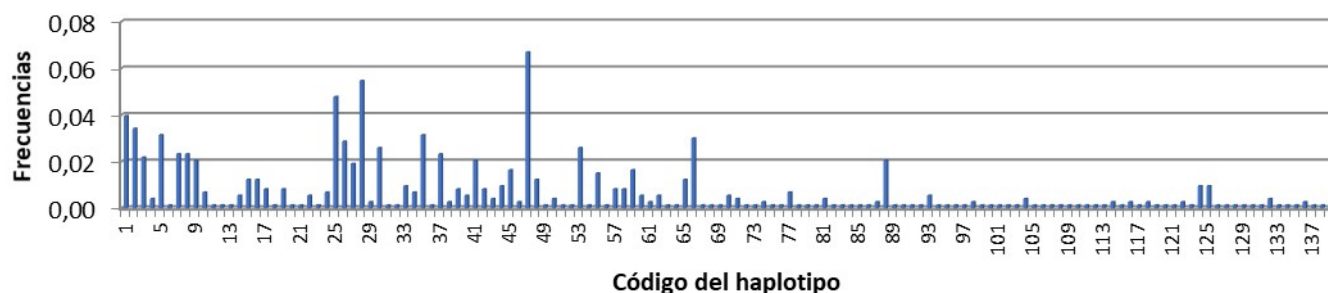
### 1.5. Haplotipos extendidos

Una vez comprobada la existencia de ligamiento entre los marcadores considerados, se han estimado las frecuencias de los haplotipos extendidos (**Anexo: Tabla A2a**). Puesto que no se pudo conocer la fase gamética, las frecuencias haplotípicas se estimaron mediante Arlequin (Excoffier & Lischer, 2010).

Se aprecia que, de una parte, los alelos rs72377086\*1, MAPT09\*22 (125 bp) y rs199451\*G, y de otra, los alelos rs72377086\*2 y MAPT09\*14 (109 bp), aparecen juntos en prácticamente todos los cromosomas. En sólo 18 de los 732 analizados (2,5%) los haplotipos son rs72377086\*1, MAPT09\*14 (109 bp) y rs199451\*G por un lado y rs72377086\*2 y MAPT09\*22 (125 bp) por otro (haplotipos identificados con los códigos 22, 51, 54, 73, 74, 75, 80, 86, 87, 126, 129, 130 y 131).

Se han encontrado también 10 cromosomas en los que los marcadores bialélicos diagnósticos del haplotipo H1 están asociados a alelos con un número de repeticiones inmediatamente inferior o superior a MAPT09\*22; siete de ellos con el alelo MAPT09\*21 (123 bp) (haplotipos 11, 12, 13, 89, 107, 108, 119), dos con el alelo 23 (127 bp) (haplotipos 72, 138) y uno con el alelo 24 (129 bp) (haplotipo 109). En el caso de los marcadores bialélicos diagnóstico del haplotipo H2, se han hallado cinco cromosomas que presentan el alelo MAPT09\*13 (107bp) (haplotipos 67, 81 y 82). En total un 2% de los cromosomas analizados.

Los haplotipos extendidos más abundantes, con frecuencias superiores a 0,05, son el 47 y el 28 (**Fig.: R5; Tabla A2a**), pertenecientes al haplotipo H1. Los haplotipos extendidos más frecuentes pertenecientes al grupo H2 son el 2 y el 5, con frecuencias en torno a 0,03, siendo el quinto y sexto más abundantes respectivamente en la clasificación global.



**Fig. R5:** Frecuencias de los haplotipos extendidos hallados en las ocho poblaciones analizadas.

En las **figuras R6 a R9** se muestran las frecuencias alélicas de los 4 marcadores microsatélites, en función de su pertenencia a un haplotipo MAPT.

Se observan unas distribuciones de frecuencias con un gran solapamiento en 3 de los 4 microsatélites. Sin embargo, como se observa también en las frecuencias de los haplotipos extendidos que se han descrito en los párrafos anteriores, en MAPT09 el alelo 22 (125 pb) es prácticamente exclusivo de H1, ya que son cromosomas H1 570 de los 575 que portan este alelo, en tanto que el alelo 14 (109 pb) lo es de H2, encontrándose en este haplotipo en la mayor parte de los casos. Los 5 cromosomas *H2/MAPT09\*22* corresponden a muestras de Gipuzkoa (3) y Gitanos (2), en tanto que los cromosomas *H1/MAPT09\*14* corresponden a Gipuzkoa, Gitanos, Valencia, Chocó y Waoranis, por lo que no pude asignárseles un origen geográfico concreto.

Puede destacarse también el alelo MAPT14\*10, que, con una frecuencia notable, es exclusivo de los cromosomas H2.

Donnelly y colaboradores (2010) observaron que los alelos más comunes en el haplotipo H2 para los microsatélites MAPT07, MAPT08, MAPT09 y MAPT14 eran, respectivamente, 19 (209 bp), 11 (203 bp), 14 (109 bp) y 10 (103 bp), coincidiendo con los resultados de este estudio en MAPT07 y MAPT08, en tanto que para el microsatélite MAPT08 el alelo más abundante es el 10 (199 bp) y en MAPT14 es el 17 (131 bp).

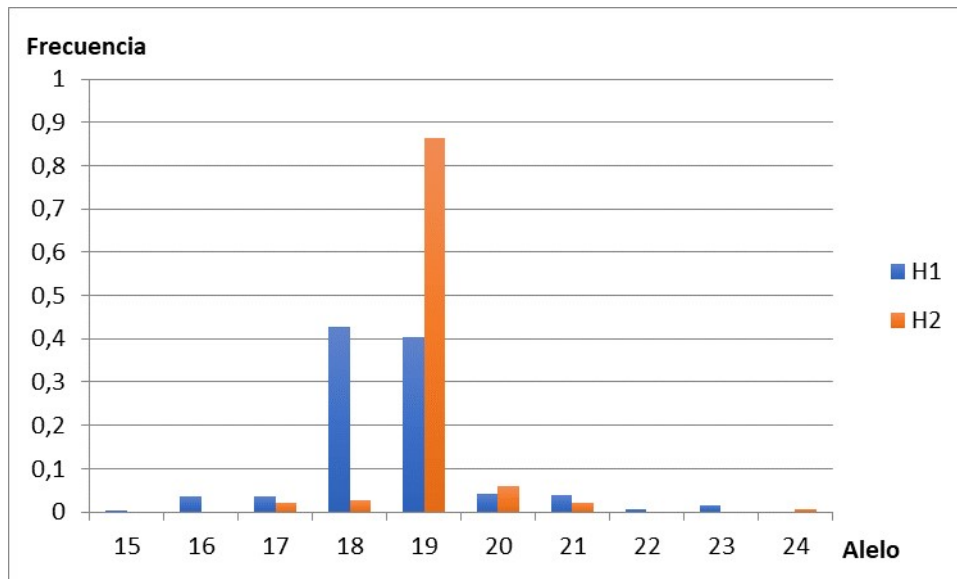


Figura R6. Histograma de frecuencias de MAPT07 para el conjunto de poblaciones analizadas, en función del haplotipo MAPT.

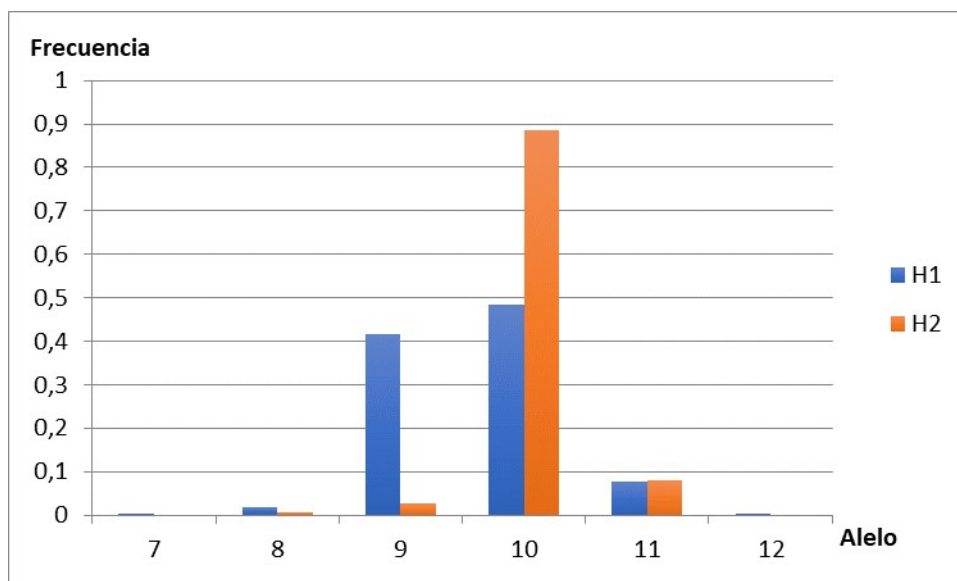
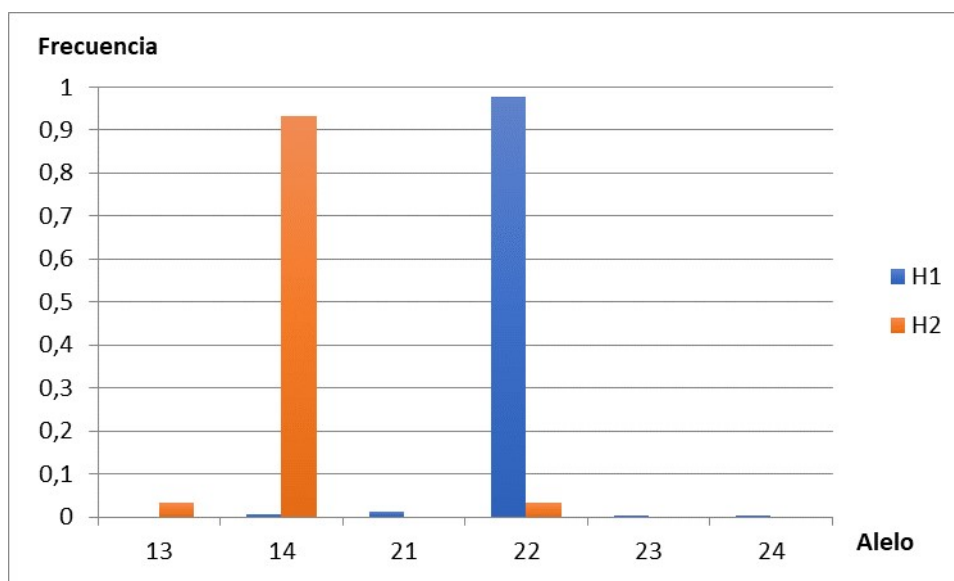
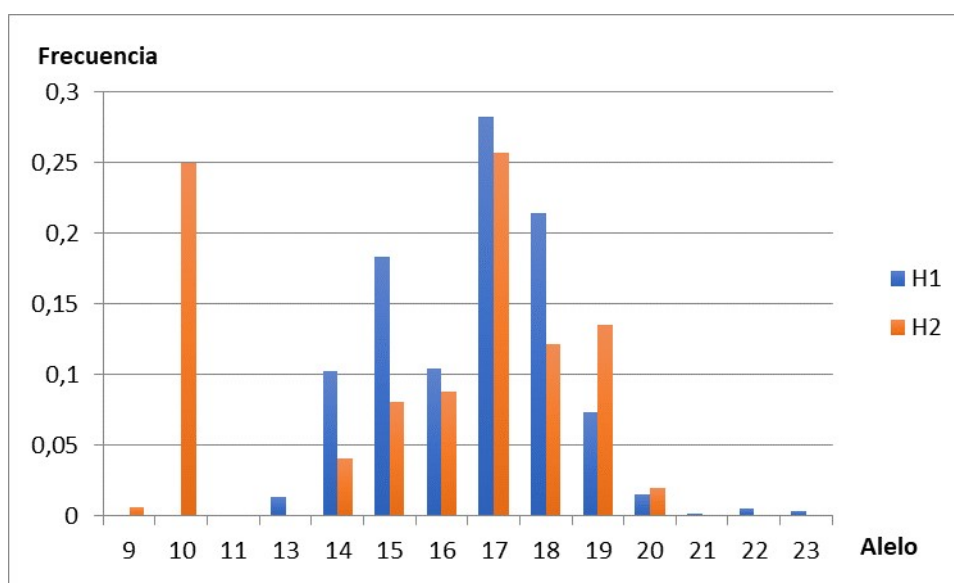


Figura R7. Histograma de frecuencias de MAPT08 para el conjunto de poblaciones analizadas, en función del haplotipo MAPT.



**Figura R8.** Histograma de frecuencias de MAPT09 para el conjunto de poblaciones analizadas, en función del haplotipo MAPT.



**Figura R9.** Histograma de frecuencias de MAPT14 para el conjunto de poblaciones analizadas, en función del haplotipo MAPT.

## 1.6. Test de neutralidad

En la **tabla R15** se muestran los valores p del estadístico F del test de neutralidad de Ewens-Watterson para todos los marcadores analizados en las diferentes poblaciones, resaltándose en **negrita** los valores significativos.

Las diferencias significativas se dan en los marcadores bialélicos en Gipúzcoa, en el microsatélite MAPT14 en Navarra, Gitanos Vascos, Valencia y Jujuy. En la población de residentes vascos, aparecen valores significativos para este estadístico en el microsatélite MAPT08. Una vez realizada la corrección de Bonferroni, tan sólo continúan mostrando significación rs72377086 y rs199451

en Gipuzkoa y MAPT14 en Valencia. Es preciso mencionar aquí que el test de Ewens-Watterson da valores significativos sistemáticamente en todos los marcadores bialélicos cuyas frecuencias se encuentren próximas a 0,5, como es el caso de Gipuzkoa para los tres marcadores bialélicos.

**Tabla R15.** Valores p del estadístico F del test de neutralidad de Ewens-Watterson. \*: Significación después de aplicar la corrección de Bonferroni.

	Gipuzkoa	Navarra	ResidentesPV	GitanosPV	Valencia	Chocó	Jujuy	Waorani
rs72377086	<b>0,0060*</b>	0,1090	0,1910	0,3040	0,1450	0,3850	0,5330	0,5640
MAPT07	0,7630	0,5800	0,3620	0,6080	0,2170	0,2980	0,2620	0,3780
MAPT08	0,2670	0,1930	<b>0,0430</b>	0,2510	0,5420	0,1010	0,4410	0,0980
MAPT09	0,3350	0,5280	0,6010	0,5470	0,8820	N.A.	0,4120	0,8510
MAPT14	0,0670	<b>0,0140</b>	0,0900	<b>0,0190</b>	<b>0,0000*</b>	0,3500	<b>0,0110</b>	0,0870
rs10514879	<b>0,0110</b>	0,1260	0,2670	0,3980	0,1850	--	N.A.	N.A.
rs199451	<b>0,0040*</b>	0,1540	0,3120	0,3710	0,2180	0,7850	1,0000	N.A.

## 1.7. AMOVA

Para estudiar cómo se estructura en el espacio la heterogeneidad genética observada, se ha realizado un análisis molecular de la varianza (AMOVA). Se han empleado todas las poblaciones de la bibliografía para las que se dispone de datos de frecuencias de los haplotipos H1', H2y H2D de MAPT, así como las del presente estudio y se han agrupado en función del continente de origen. Los resultados se muestran en la **tabla R16**.

En relación a las diferencias de las cinco regiones en que se agruparon las 96 poblaciones, el porcentaje de la varianza entre los cinco grupos propuestos ( $F_{CT}$ ) ha sido del 10,26%. En segundo lugar, se ha observado un porcentaje de varianza dentro de las poblaciones ( $F_{ST}$ ) del 83,25%. La varianza de las poblaciones dentro de los grupos propuestos ( $F_{SC}$ ) ha sido del 6,50%. Los tres componentes del AMOVA resultaron altamente significativos ( $p < 0,00001$ ), siendo los valores mencionados similares a los promedios obtenidos para diferentes marcadores genéticos (Barbujani y Goldstein, 2004).

Puede inferirse, por tanto, que existe una heterogeneidad significativa en la distribución de los haplotipos MAPT tanto entre continentes como dentro de los continentes.

**Tabla R16.** Análisis molecular de la varianza de los haplotipos H1', H2y H2D de MAPT. Porcentaje de variación y significación. Cinco grupos en función del continente de origen.

$F_{ST}$	$F_{SC}$	$F_{CT}$
83,25*	6,50*	10,26*

$F_{ST}$ : Variación genética dentro de poblaciones.  $F_{SC}$ : Variación genética entre poblaciones dentro de grupos.  $F_{CT}$ : Variación genética entre grupos. \*  $p < 0,00001$

## Sección 2. Variabilidad geográfica de los haplotipos SNP

Se han tomado de la bibliografía las frecuencias publicadas para los 3 haplotipos analizados en poblaciones de diferentes continentes. Se incluyen, junto con las frecuencias obtenidas en este trabajo, en la **Tabla Anexa A1**.

En las **figuras R10, R11 y R12** se representan los valores máximo, mínimo y medio de las frecuencias de los haplotipos H1', H2' y H2D por subcontinentes, respectivamente. Las **figuras R13 a R15** muestran esta misma información, pero excluyendo las poblaciones mestizas o desubicadas (afroamericanos, europeos residentes en EEUU, mejicanos, colombianos, peruanos y portorriqueños).

Se observan unos valores relativamente bajos de H1' en Europa y Oeste de Asia. Además, en Europa el rango de variación es muy superior al de otras regiones.

H2' muestra valores bajos en todos los casos, pero con un mayor rango en África, Europa y Sur de Asia.

Al contrario que para H1', para H2D Europa y Oeste de Asia presentan unos valores superiores al resto de regiones. En este caso, también Europa ofrece un rango mayor de valores, con un máximo en la población de Gipuzkoa. El valor medio de H2D es superior en el Oeste de Asia en relación a Europa. Esto puede indicar un problema de muestreo, ya que la región Oeste de Asia se encuentra representada tan sólo por 3 poblaciones de Oriente Próximo, (Ashkenazi, Drusos y Samaritanos) con características de aislados genéticos.

Se observan unos valores relativamente altos de H2D en América, que bajan ostensiblemente cuando se excluyen las poblaciones de mestizos, al mismo tiempo que se reduce drásticamente su rango de variabilidad.

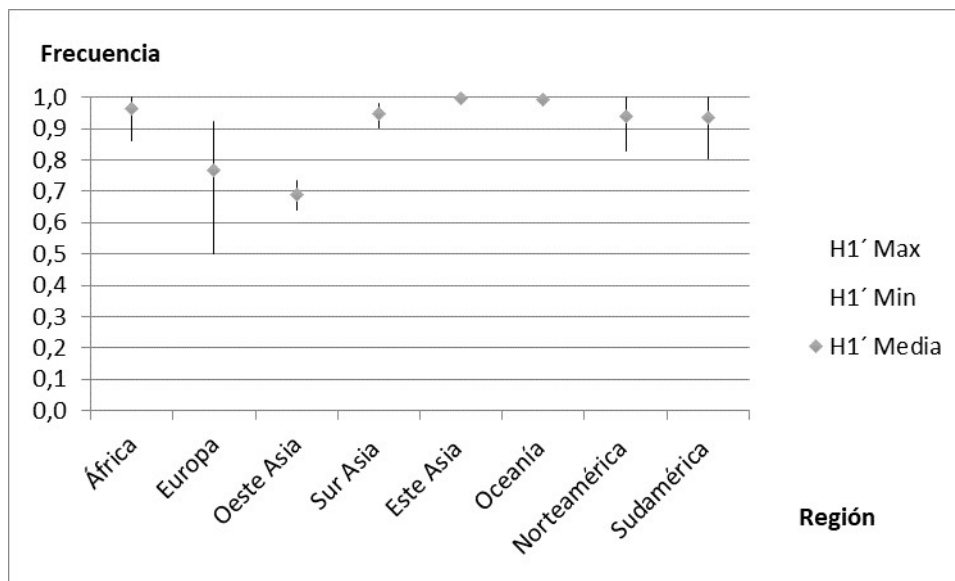


Figura R10. Valores máximo, mínimo y medio de las frecuencias del haplotipo H1' por subcontinentes.

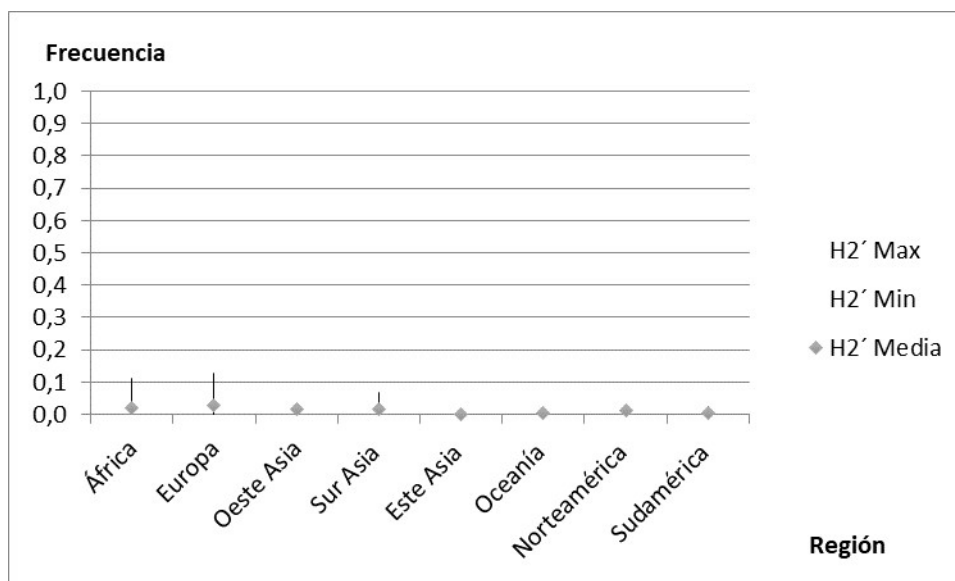


Figura R11. Valores máximo, mínimo y medio de las frecuencias del haplotipo H2' por subcontinentes.



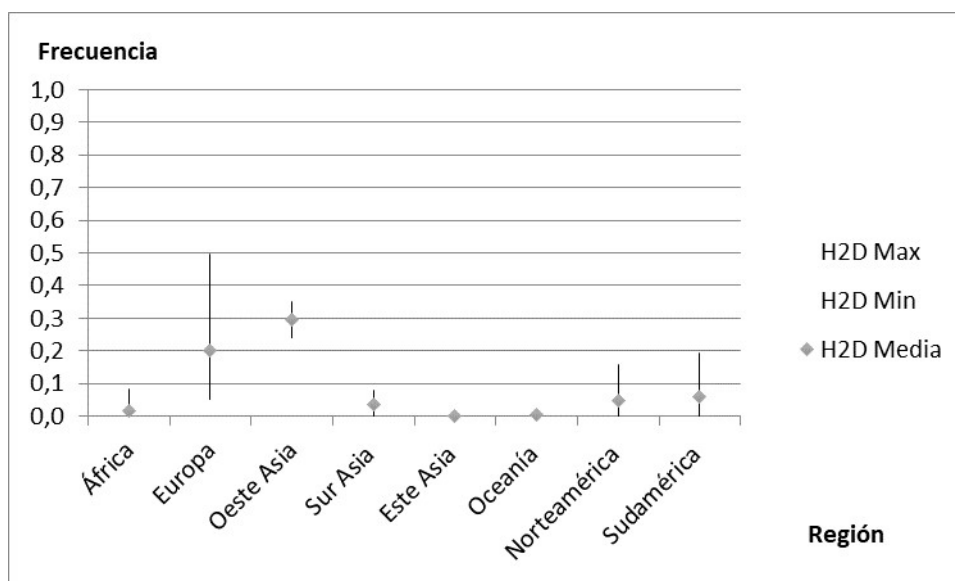


Figura R12. Valores máximo, mínimo y medio de las frecuencias del haplotipo H2D por subcontinentes.

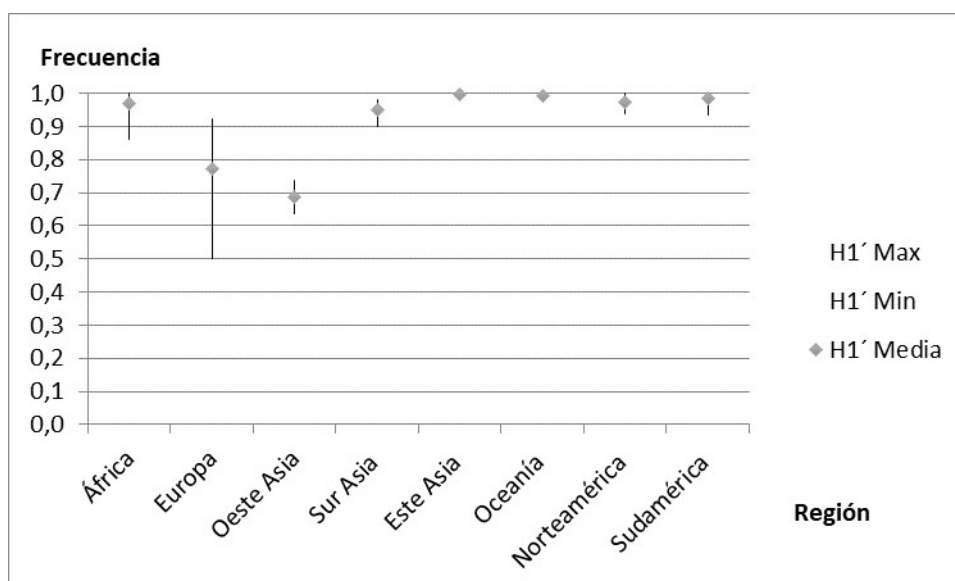
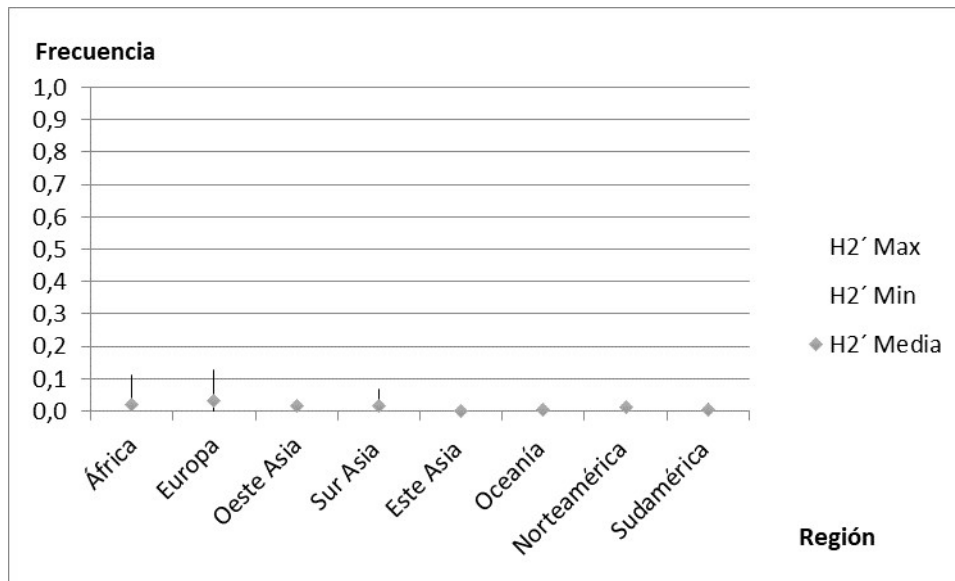
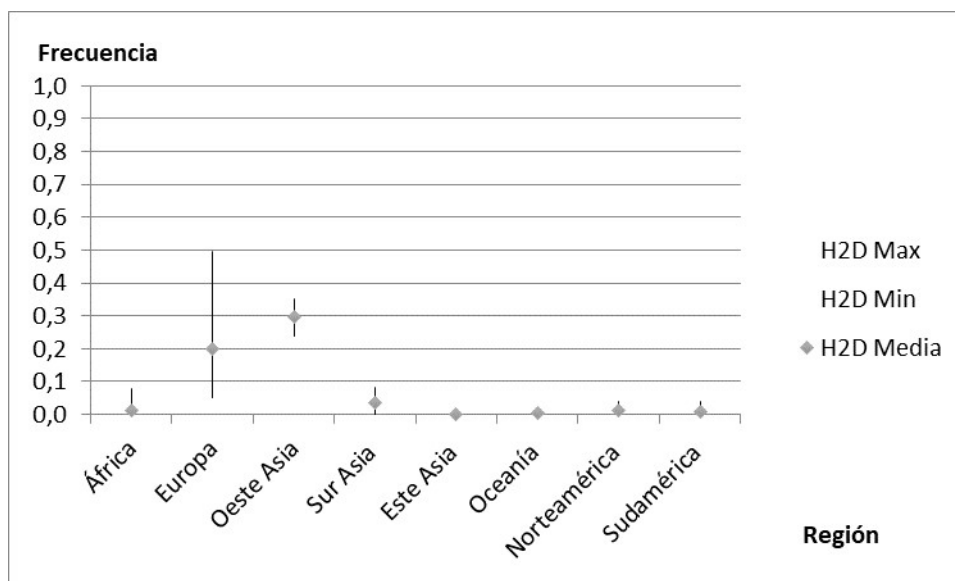


Figura R13. Valores máximo, mínimo y medio de las frecuencias del haplotipo H1' por subcontinentes, excluyendo poblaciones con un alto grado de mestizaje.



**Figura R14.** Valores máximo, mínimo y medio de las frecuencias del haplotipo H2' por subcontinentes, excluyendo poblaciones con un alto grado de mestizaje.

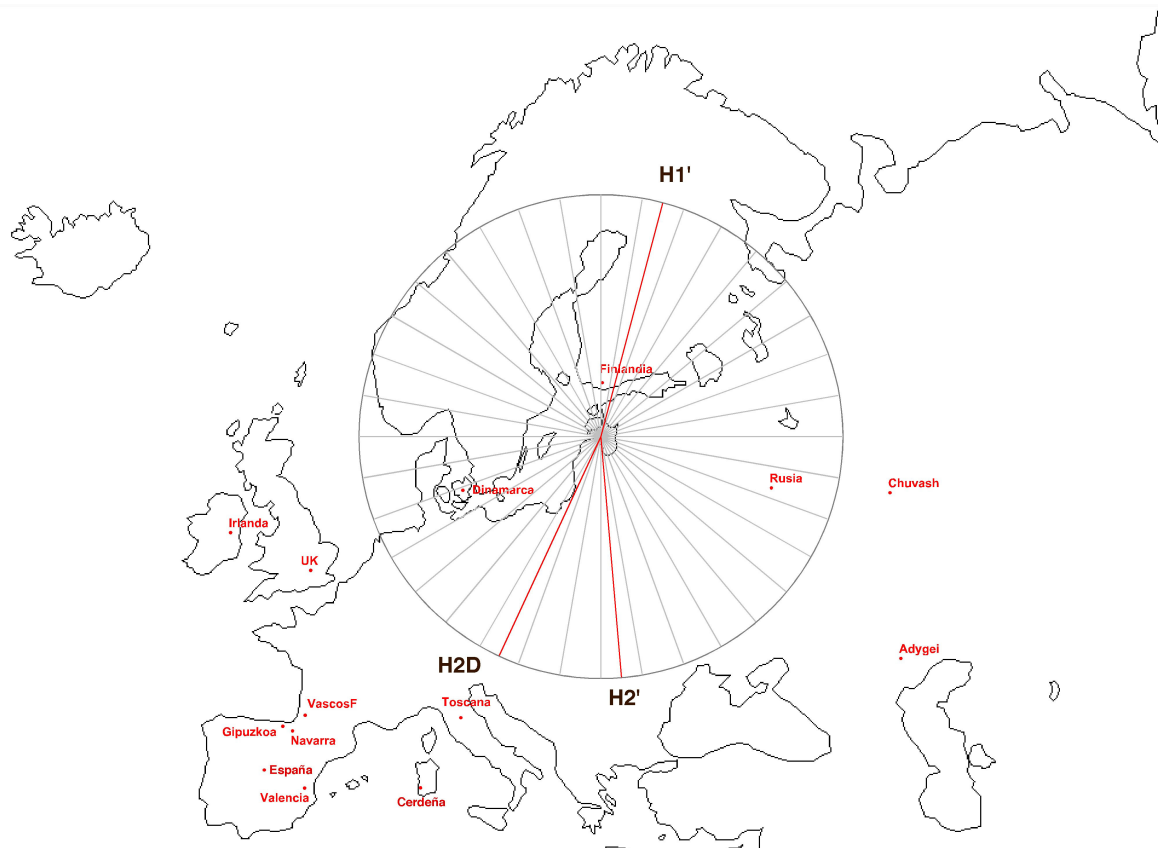


**Figura R15.** Valores máximo, mínimo y medio de las frecuencias del haplotipo H2D por subcontinentes, excluyendo poblaciones con un alto grado de mestizaje.

Con el fin de realizar un análisis más minucioso de la distribución geográfica de las frecuencias de los haplotipos H1', H2' y H2D, se ha realizado una serie de análisis estadísticos en diferentes ámbitos geográficos, incluyendo Europa, África, Asia-Pacífico y América. Se presentan en primer lugar los resultados de Europa, por ser el continente con la mayor heterogeneidad genética para estos marcadores, para seguir con África, Asia-Pacífico y América.

## 2.1. Europa

En Europa, se ha observado una correlación significativa para las frecuencias de los tres haplotipos MAPT con las coordenadas geográficas. Los gradientes de frecuencias obtenidos descubren clinas de orientación suroeste-nordeste para H1', nordeste-suroeste para H2D y noroeste-sudeste para H2' (Fig. R16).



**Figura R16.** Orientación de las clinas correspondientes a los haplotipos MAPT en Europa.

La clina de H1' (**figura R17**), con una orientación de 12 grados en relación al norte, muestra su máxima frecuencia en el extremo nordeste del continente, representado por Rusia (0,906), los Chuvash (0,893), una etnia túrquica que habita en la región rusa del Volga y Finlandia (0,893). Estos valores decaen, siguiendo la orientación de la clina, al analizar poblaciones situadas en zonas cada vez más al oeste y al sur de Europa, como los Adygei, un grupo étnico habitante de la región rusa del Cáucaso (0,843), Dinamarca (0,843), Irlanda (0,825) y Reino Unido (0,755). Los valores más bajos del haplotipo H1' se encuentran en el extremo sudoeste de Europa en poblaciones italianas – Toscana (0,636) y Cerdeña (0,643) – y en los vascofranceses (0,645).

Entre las muestras analizadas para este trabajo, la población de Valencia es la que presenta una mayor frecuencia de este haplotipo (0,742), consistente con la frecuencia calculada previamente para una muestra de España (0,736). Los vascos de Navarra presentan una frecuencia más baja, 0,689, mientras que el valor mínimo aparece en Gipuzkoa, con una frecuencia extraordinariamente baja, 0,506.

El coeficiente de determinación de la regresión lineal es  $R^2$ : 0,680, con 13 grados de libertad y la significación estadística de la regresión es  $p$ : 0,0003.

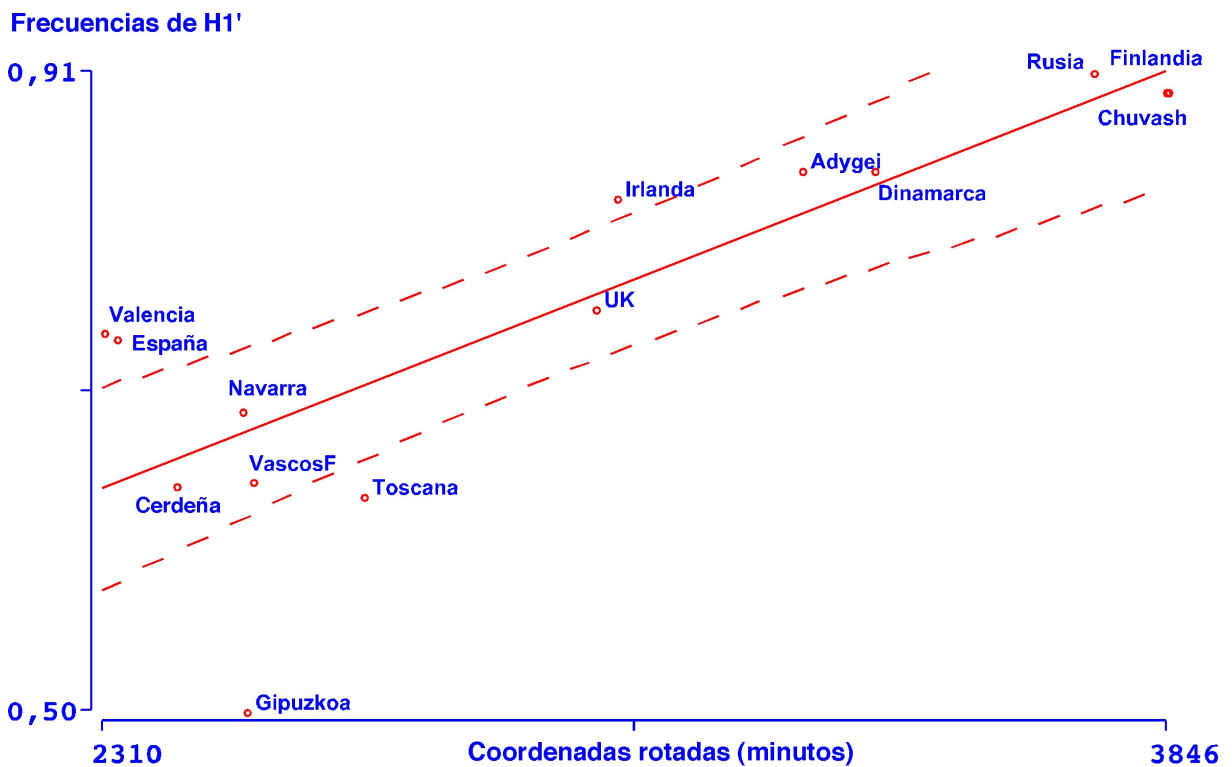


Figura R17. Recta de regresión con intervalos de confianza del 95% de la clina observada para H1' en Europa.

En el caso del haplotipo MAPT\*H2, la clina obtenida (figura R18) tiene una orientación de 170 grados, es decir, las frecuencias de este haplotipo aumentan de norte a sur en un eje ligeramente inclinado en dirección noroeste – sudeste, pero la significación de la misma es escasa ( $R^2$ : 0,340;  $p$ : 0,0279), de manera que, aunque la frecuencia máxima de H2 aparece en la población más al sur (Cerdeña, 0,127), los valores inmediatamente más altos de este haplotipo aparecen en Vascofranceses (0,105) y en los Adygei (0,077); es decir, conforman este grupo tres poblaciones reconocidas por poseer una cierta peculiaridad genética.

En regiones tan alejadas en este eje norte-sur como España y Finlandia, el haplotipo H2' está ausente y se registran valores mínimos del mismo en Rusia (0,004), Irlanda (0,005) y entre los Chuvash (0,007).

Poblaciones como Dinamarca, Reino Unido o Toscana, muestran unos valores intermedios, aunque también muy bajos, por debajo de 0,04 (0,017; 0,022; 0,037 respectivamente).

En cuanto a las poblaciones analizadas para este trabajo, la población vasca de Gipuzkoa, con una frecuencia de 0,006, comparte rango de valores con poblaciones tan dispares en el contexto europeo como Rusia, Irlanda y los Chuvash, mientras que la población vasca de Navarra se sitúa en el grupo de valores intermedios (0,026). En este mismo grupo encaja Valencia (0,027), a pesar de no haberse detectado este haplotipo en la población española tomada de la bibliografía. Es sorprendente la diferencia en las frecuencias de H2' de la población vascofrancesa con las poblaciones vascas peninsulares.

#### Frecuencias de H2'

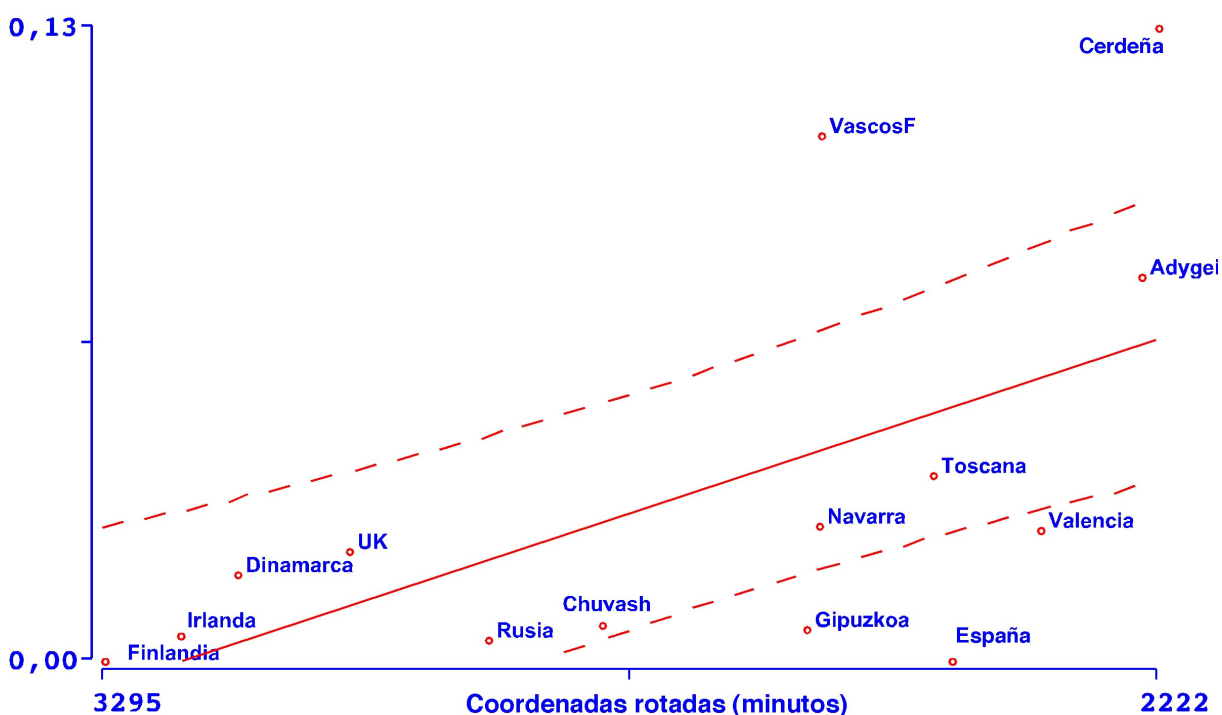


Figura R18. Recta de regresión con intervalos de confianza del 95% de la clina observada para H2' en Europa.

La clina observada para el haplotipo H2D en Europa presenta una orientación de 204 grados (figura R19), es decir, el gradiente de frecuencias de este haplotipo aumenta en dirección noroeste – sudoeste. Esta clina, como la de H1', se ajusta bastante bien a la distribución geográfica de las poblaciones

analizadas, con una regresión estadísticamente significativa ( $R^2$ : 0,590;  $p$ : 0,0014).

Las poblaciones ubicadas en el extremo noroeste de la clina presentan frecuencias en torno a 0,1 (Adygei, 0,080; Rusia, 0,090; Chuvash, 0,100 y Finlandia, 0,107). Estas frecuencias aumentan paulatinamente en poblaciones más occidentales como Dinamarca (0,140), Reino Unido (0,223) e Irlanda (0,170). Siguiendo esta tendencia, en el extremo sudoeste de Europa aparecen las poblaciones con las frecuencias más altas, Cerdeña (0,230), vascofranceses (0,250), España (0,264) y la que hasta el presente estudio mostraba la frecuencia más alta de este haplotipo, la Toscana (0,327).

Las poblaciones de Valencia (0,234) y Navarra (0,283) exhiben frecuencias coherentes con el contexto geográfico en el que se encuentran, en tanto que la frecuencia de vascos del *goiherri* guipuzcoano establece un nuevo máximo europeo y mundial.

**Frecuencias de H2D**

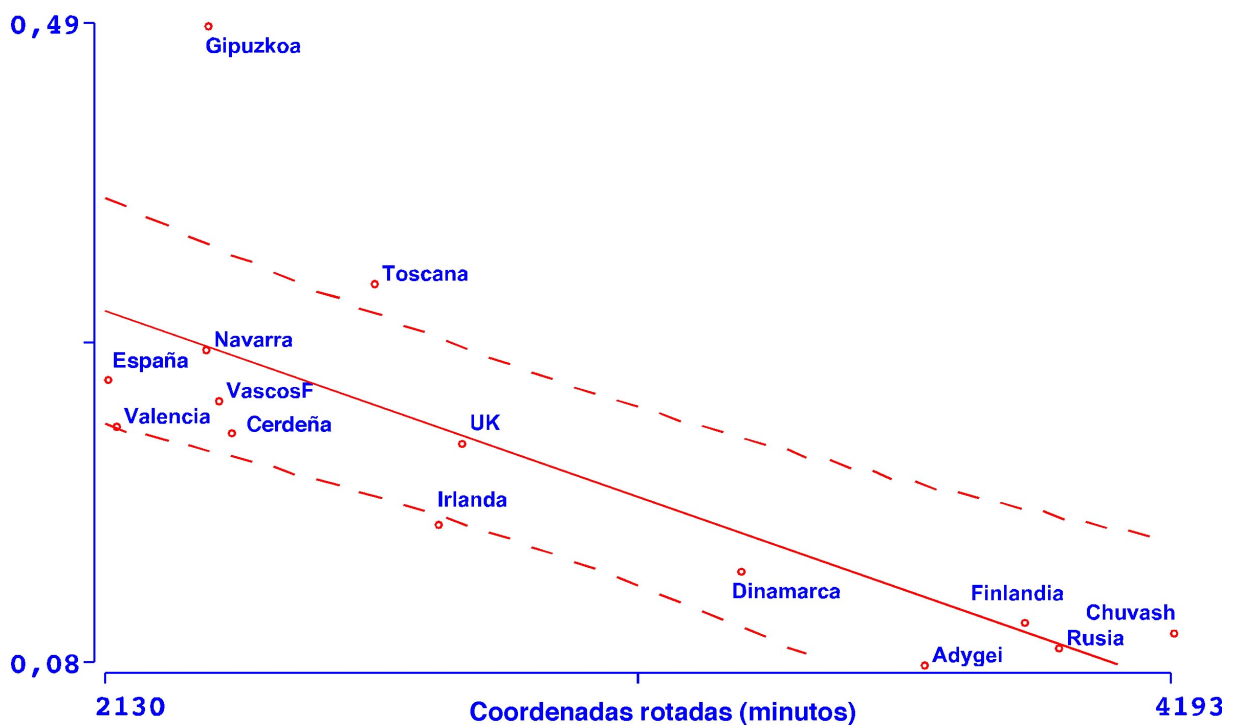


Figura R19. Recta de regresión con intervalos de confianza del 95% de la clina observada para H2D en Europa.

Según algunos autores (Hartl y Clark, 1997), la transformación angular  $\arcsin\sqrt{p}$  determina que la varianza de una estima (en este caso la frecuencia alélica) sea independiente de la propia estima, lo que redundaría en una mayor eficacia en la detección de clinas. Se realizaron regresiones sobre la

transformación angular de las frecuencias alélicas, obteniéndose unas clinas muy similares, tanto en su orientación (13, 170 y 207 grados respectivamente) como en su significación ( $R^2$ : 0,590, 0,350 y 0,670).

Con el fin de aumentar el número de poblaciones incluidas en los análisis, se ha ampliado la base de datos considerando también aquellas muestras analizadas sólo para los dos haplotipos básicos, H1 y H2. De este modo, se ha incrementado la representatividad de los resultados, particularmente por la inclusión de poblaciones de Europa central. En la **figura R20** se muestra la clina obtenida para H1, observándose una tendencia muy similar a la mostrada en los párrafos anteriores ( $R^2$ : 0,560. p: 0,0000). En este caso, el azimut de la clina es de 20 grados, por 12 del análisis anterior.

La mayor parte de las poblaciones añadidas encajan perfectamente dentro del intervalo de confianza de la clina (Grecia, Brescia, Serbia, Hungría, Islandia). Entre las principales excepciones, Cataluña muestra una frecuencia de H1 algo superior a la media de la región y las Islas Orcadas, en cambio, inferior.

La clina obtenida para H2 es estrictamente complementaria de la de H1, con una orientación de 200 grados y la misma significación (**Figura R21**).

Dado que las clinas para H1 y H2 son complementarias, sucede lo mismo en el sentido inverso con H2. Cataluña presenta una frecuencia de H2 menor que España (0,209 frente a 0,264), que la comunidad autónoma vecina, Valencia (0,258) y por supuesto que ambas poblaciones vascas (0,311 Navarra y 0,494 Gipuzkoa). Las Islas Orcadas, por el contrario, presentan una mayor frecuencia del haplotipo H2 (0,280) que la esperada por su ubicación geográfica.

Frecuencias de H1

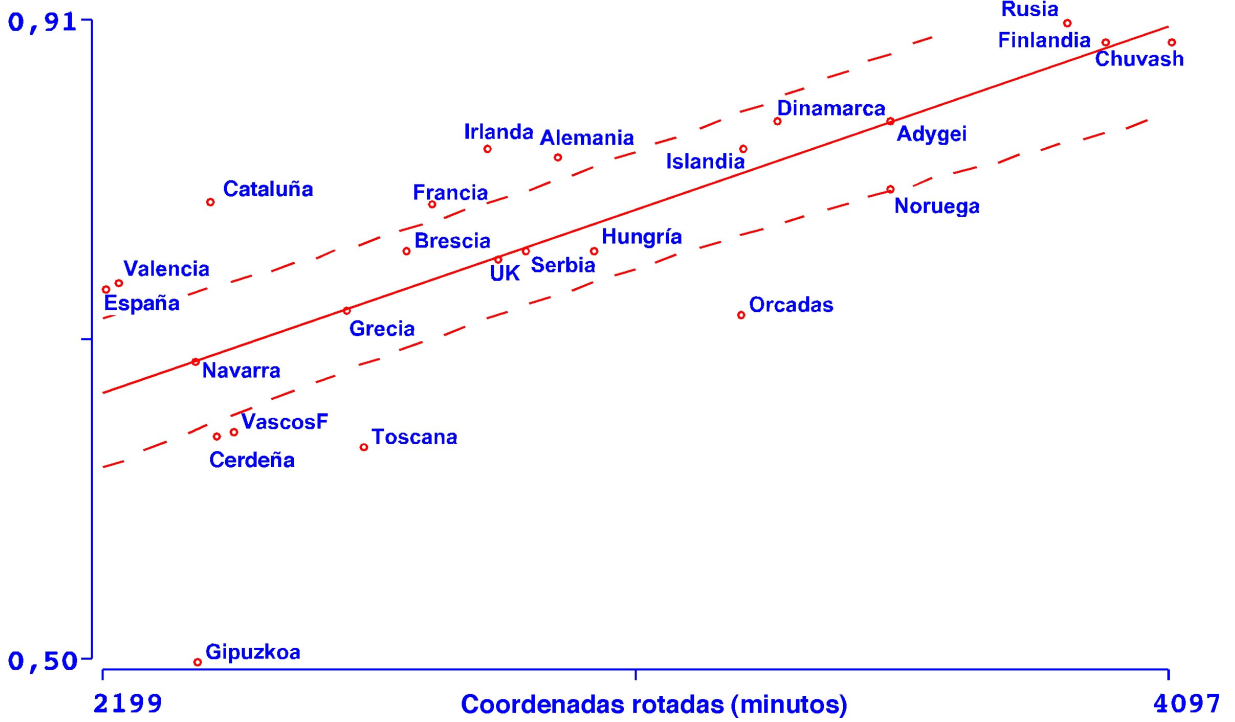


Figura R20. Recta de regresión con intervalos de confianza del 95% de la clina observada para H1 en Europa.

Frecuencias de H2

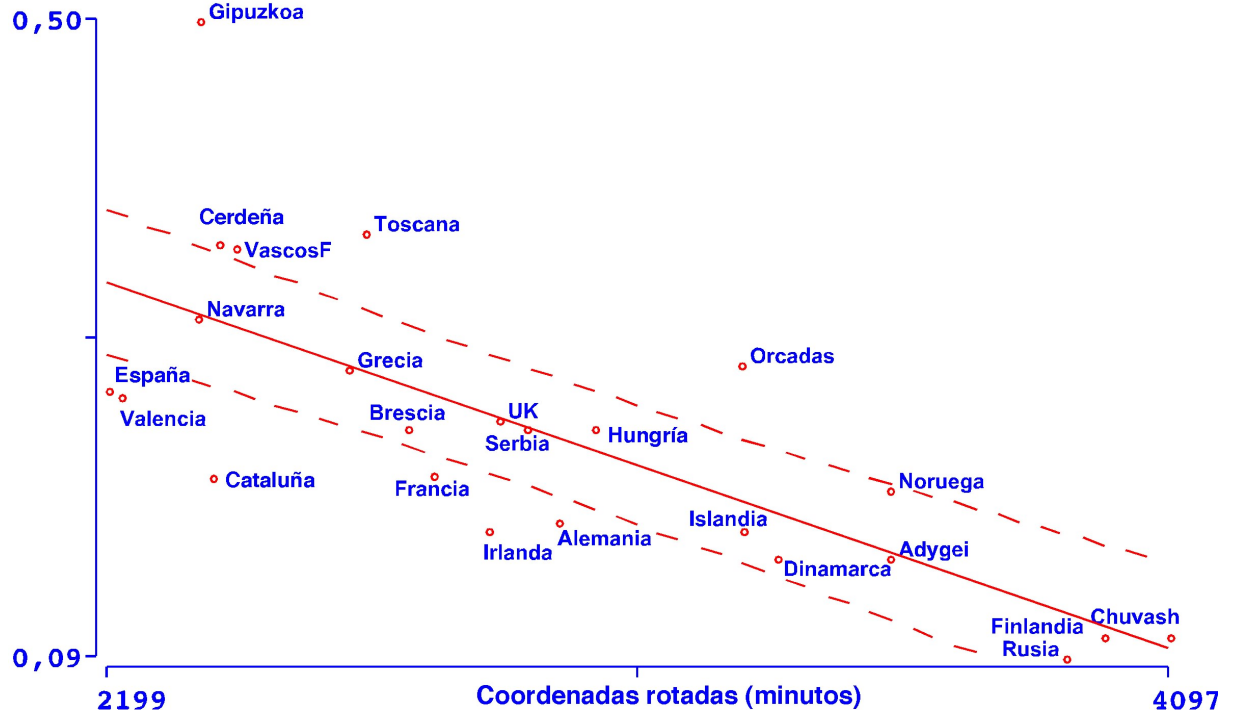


Figura R21. Recta de regresión con intervalos de confianza del 95% de la clina observada para H2 en Europa.



En un nuevo grupo de análisis a nivel de subhaplotipo (H1, H2 y H2D) se incluyeron las poblaciones de Residentes y Gitanos del País Vasco (**Figuras R22, R23 y R24**, respectivamente). En este caso, las clinas mantienen a grandes rasgos las mismas orientaciones, pero las significaciones de las regresiones empeoraron apreciablemente ( $R^2$ : 0,530, 0,300 y 0,490 respectivamente).

En lo referente al haplotipo *MAPT\*H1*, los Residentes del País Vasco (0,717) se ubican en una posición intermedia entre la población vasca de Navarra (0,689) y Valencia (0,742), muy cerca del valor de la frecuencia para el haplotipo H1 de la población española general (0,736).

Los Gitanos del País Vasco, en cambio, aparecen bastante alejados del resto de poblaciones del área que habitan, pero también del resto de poblaciones de la Península Ibérica y lejos de los valores esperados para una población ubicada en el extremo sudoeste de Europa. Es lícito suponer que esta especificidad puede deberse al origen asiático de la población romaní y a un cierto grado de aislamiento. Estas circunstancias se discutirán más adelante.

La distribución de frecuencias del haplotipo H2 sigue sin formar un patrón claro para su interpretación. Los Residentes del País Vasco (0,022) presentan un valor muy similar a vascos de Navarra (0,027) y a Valencia (0,026). En la población Gitana vasca, al igual que en Finlandia y en la muestra de España, no se ha detectado este haplotipo.

Siguiendo una tendencia inversa a la encontrada para el haplotipo H1, los gitanos vascos muestran una frecuencia del haplotipo H2D por debajo de la esperada para una población del sudoeste de Europa. Su valor (0,125) es menor de la mitad que el hallado para la población española general (0,264) y destaca por ello entre las poblaciones de su entorno inmediato.

Los Residentes vascos (0,261) presentan una frecuencia prácticamente igual a la general española, un poco más alta que la de Valencia (0,234) y algo menor que la de vascos de Navarra (0,283). Un término medio coherente con las características de esta muestra de descendientes de vascos y españoles de otras comunidades.

Frecuencias de H1'

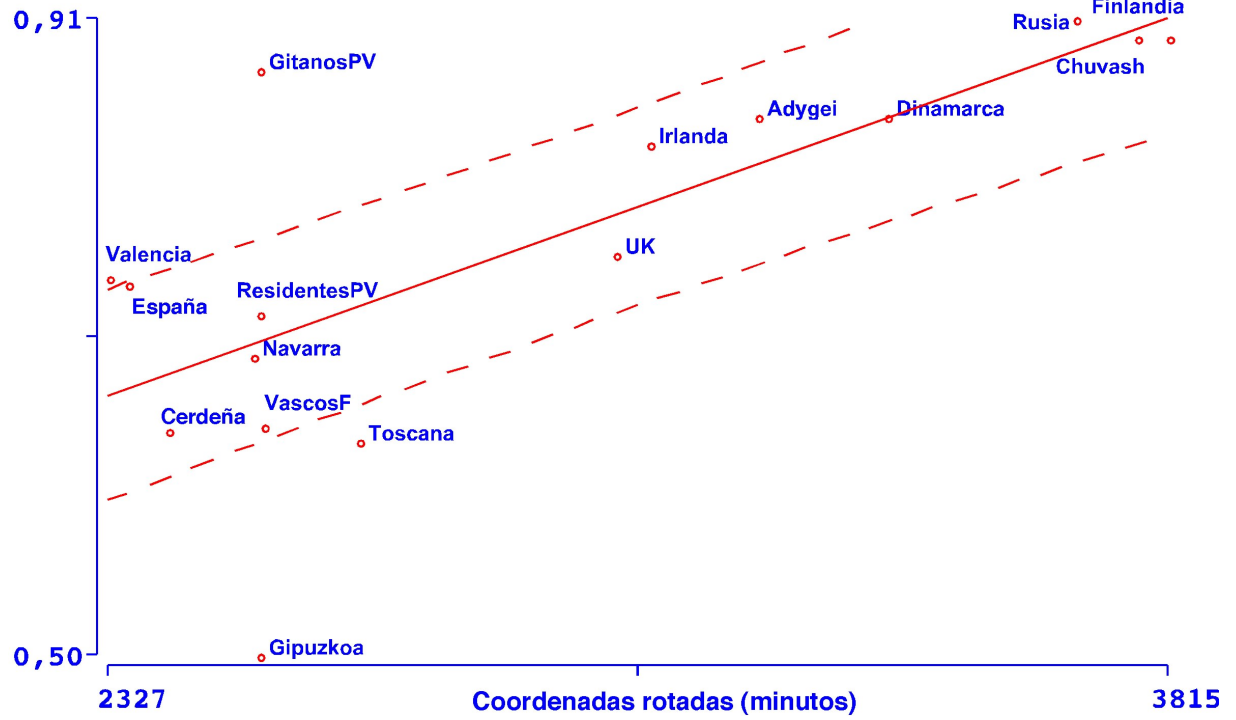


Figura R22. Recta de regresión con intervalos de confianza del 95% y orientación de la clina observada para H1' en Europa, incluyendo las poblaciones de residentes y gitanos del País Vasco.

Frecuencias de H2'

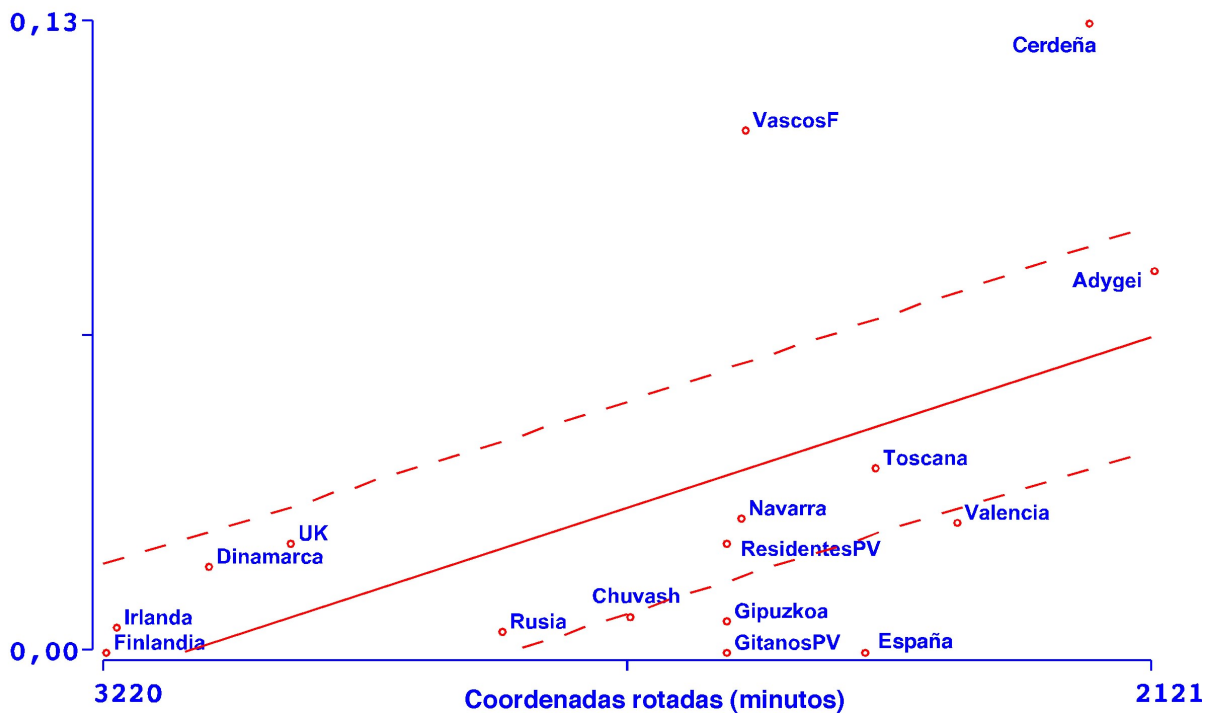
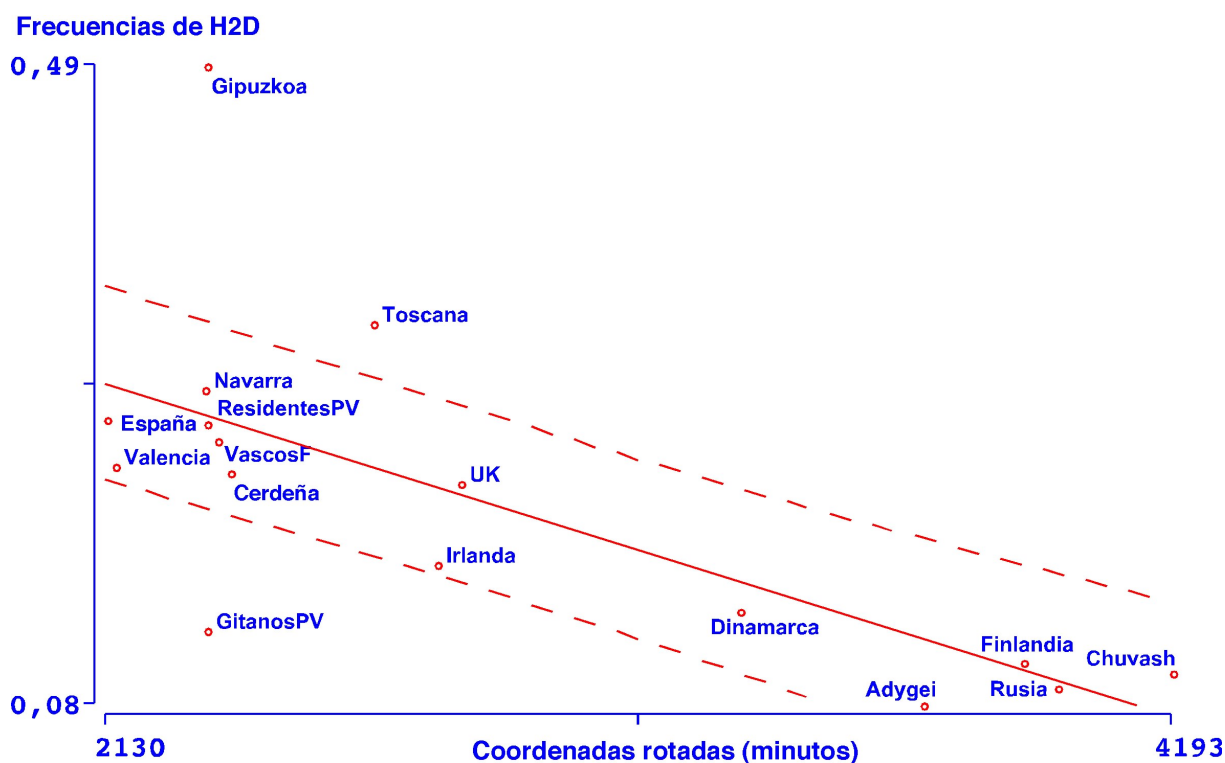


Figura R23. Recta de regresión con intervalos de confianza del 95% y orientación de la clina observada para H2' en Europa, incluyendo las poblaciones de residentes y gitanos del País Vasco.



**Figura R24.** Recta de regresión con intervalos de confianza del 95% y orientación de la clina observada para H2D en Europa, incluyendo las poblaciones de residentes y gitanos del País Vasco.

En las **figuras R25, R26 y R27** pueden observarse tres mapas sintéticos que representan el gradiente de frecuencias de H1', H2' y H2D en Europa, respectivamente. En ellos se aprecia, en líneas generales, que la frecuencia del haplotipo H1' aumenta en el sentido SO – NE, al contrario que las frecuencias del haplotipo H2D que aumenta en el sentido NE – SO y está principalmente representado en el tercio oeste del continente europeo, mientras que H2' presenta frecuencias menores que H2D y muestra una mayor distribución en la zona mediterránea de Europa, disminuyendo su presencia a medida que se aleja de esta región.

Estas tendencias coinciden con las detectadas en los análisis presentados previamente en este trabajo y hacen quizá más evidente la posición fundamental del País Vasco por sus valores extremos en los haplotipos H1' y H2D. Tomando como centro la región geográfica de influencia histórica vasca, la frecuencia del haplotipo H1' aumenta según se analizan poblaciones más lejanas de este punto mientras que la presencia del haplotipo H2D disminuye hasta apenas detectarse entre los grupos más distantes, Rusos (0,09) y Adygei (0,08).

En lo que atañe al haplotipo H2', la falta de un patrón regular evidente queda difuminada en el mapa sintético, dado que en este tipo de

representaciones las irregularidades quedan suavizadas, observándose una tendencia a valores más altos en la región mediterránea.

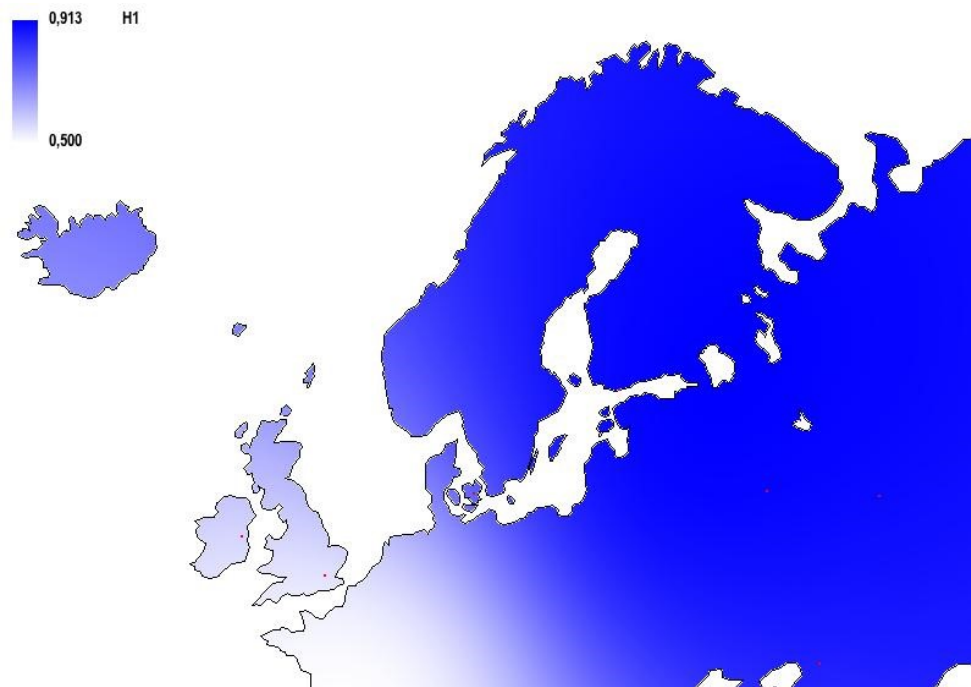


Figura R25. Mapa sintético que representa el gradiente de frecuencias de H1' en Europa.

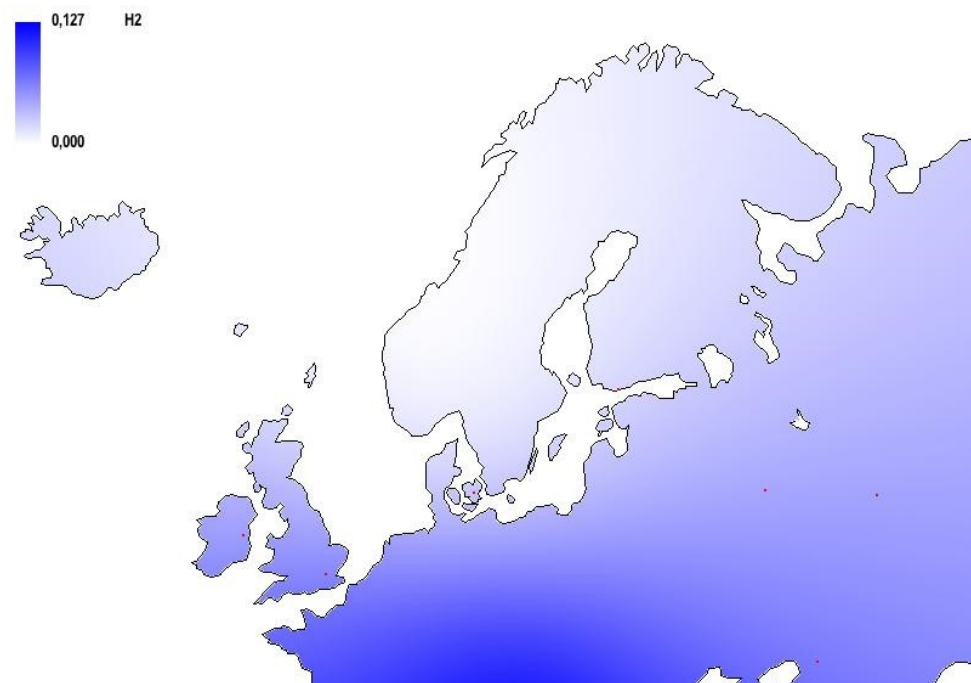


Figura R26. Mapa sintético que representa el gradiente de frecuencias de H2' en Europa.

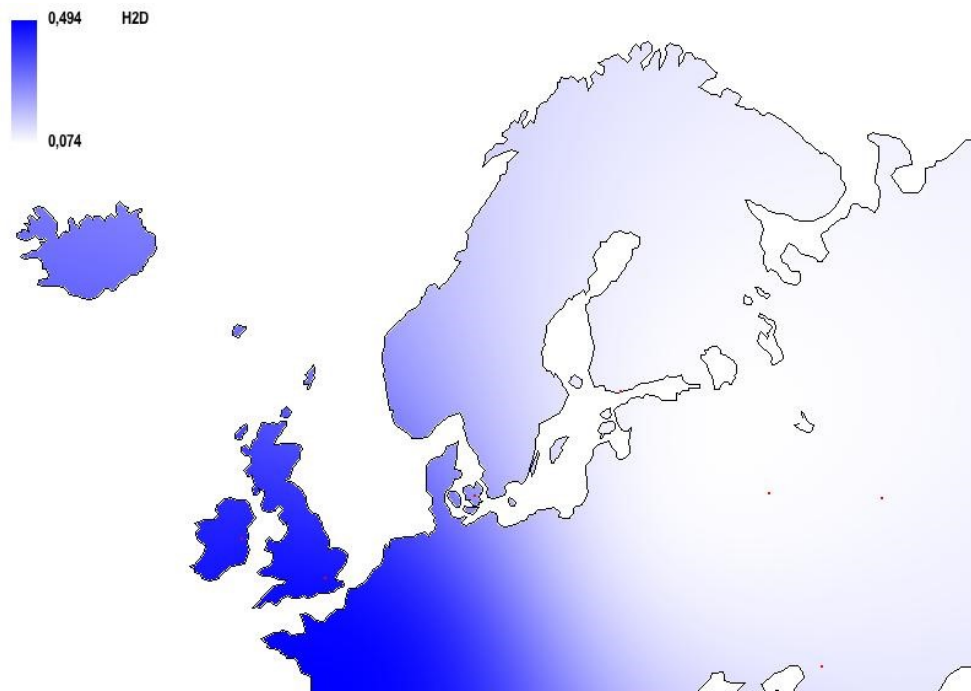


Figura R27. Mapa sintético que representa el gradiente de frecuencias de H2D en Europa.

Con el fin de profundizar en las causas que han podido determinar la distribución de frecuencias de los haplotipos MAPT en Europa, se ha realizado un análisis del centroide (Harpending y Ward, 1982). En la **figura R28** se observa un grupo de poblaciones que muestran una heterocigosidad mucho mayor de la esperada en relación al conjunto de poblaciones europeas, lo que implica un exceso de flujo génico. Este grupo incluye particularmente a los guipuzcoanos, pero también sardos, vascofranceses y toscanos y en general las poblaciones del sur del continente. Puesto que al menos una parte de estas poblaciones se han considerado tradicionalmente caracterizadas por un cierto grado de aislamiento, puede pensarse más bien en una interpretación basada en un origen diferenciado de una parte de su patrimonio genético.

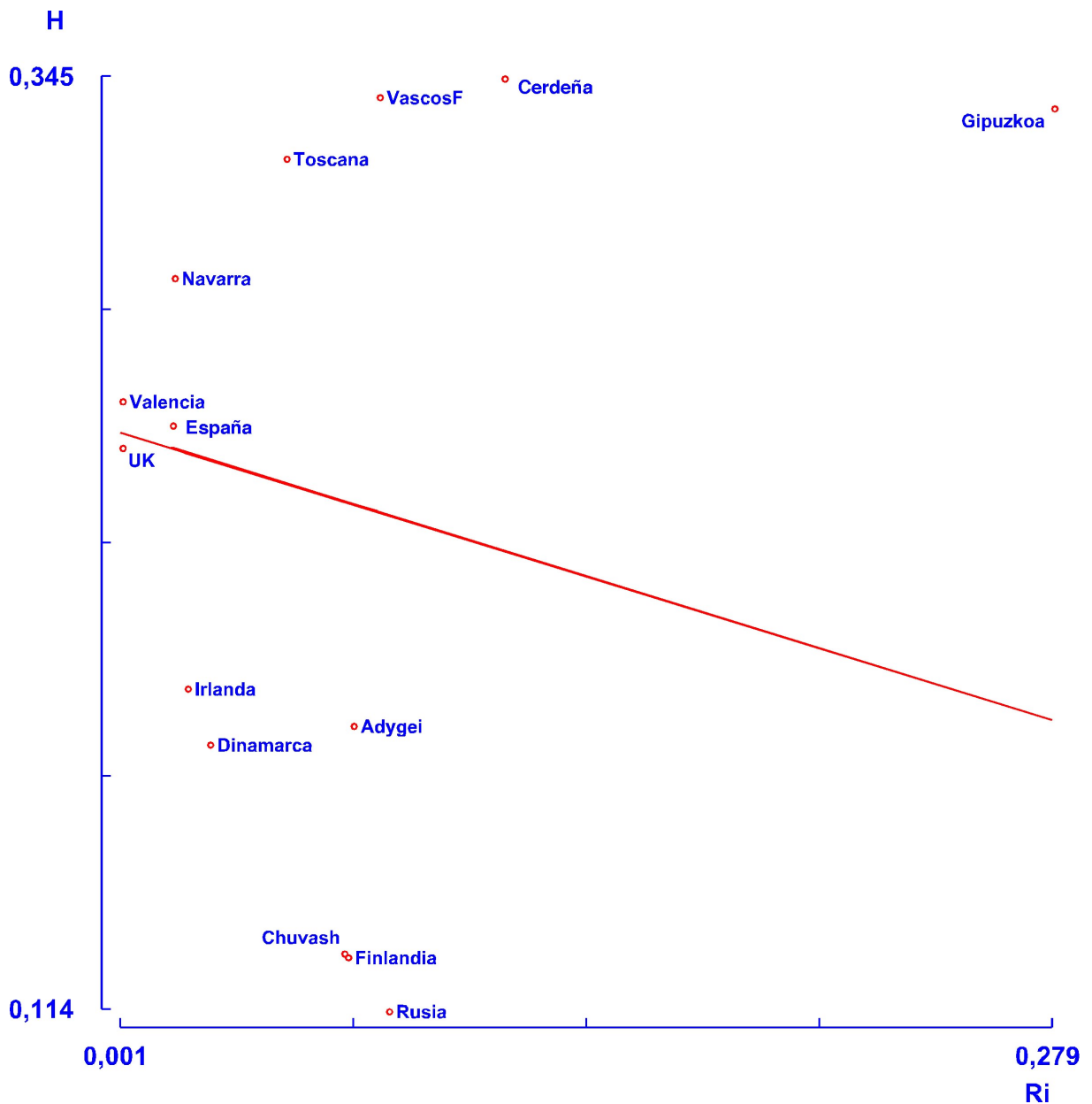
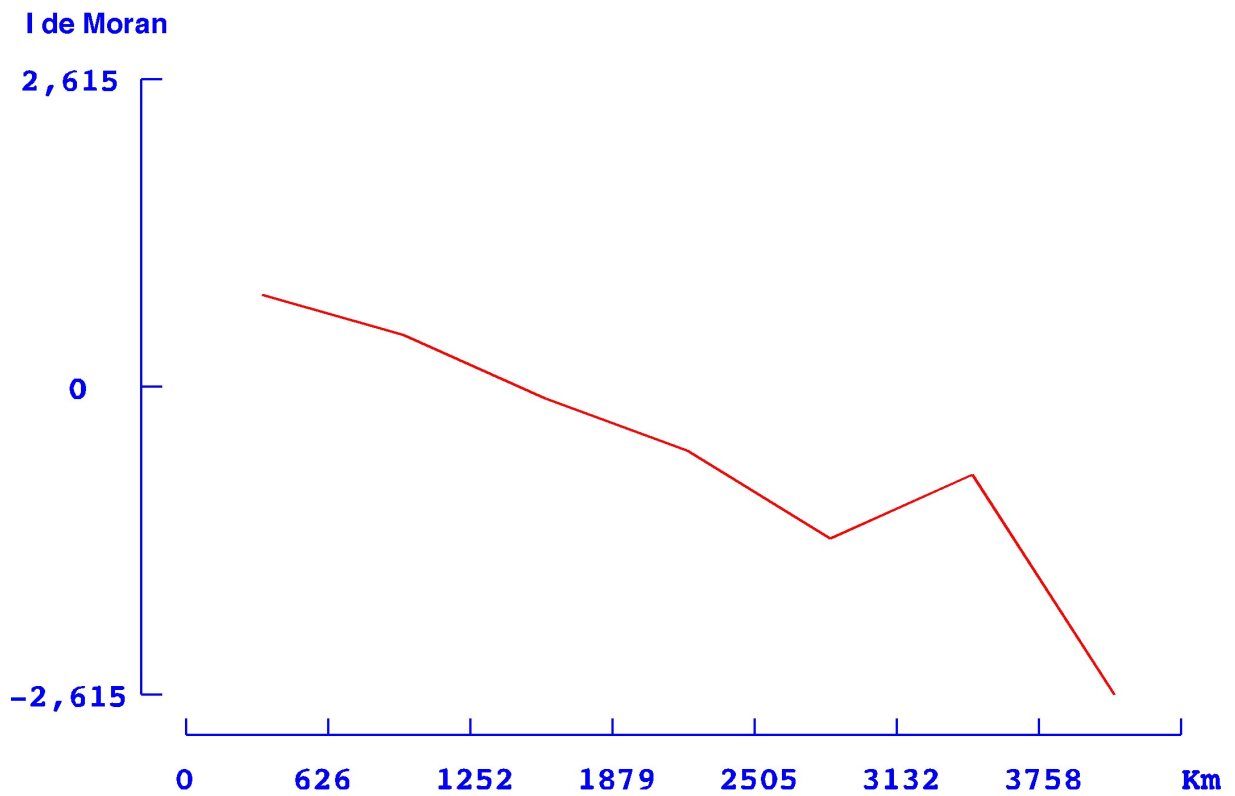


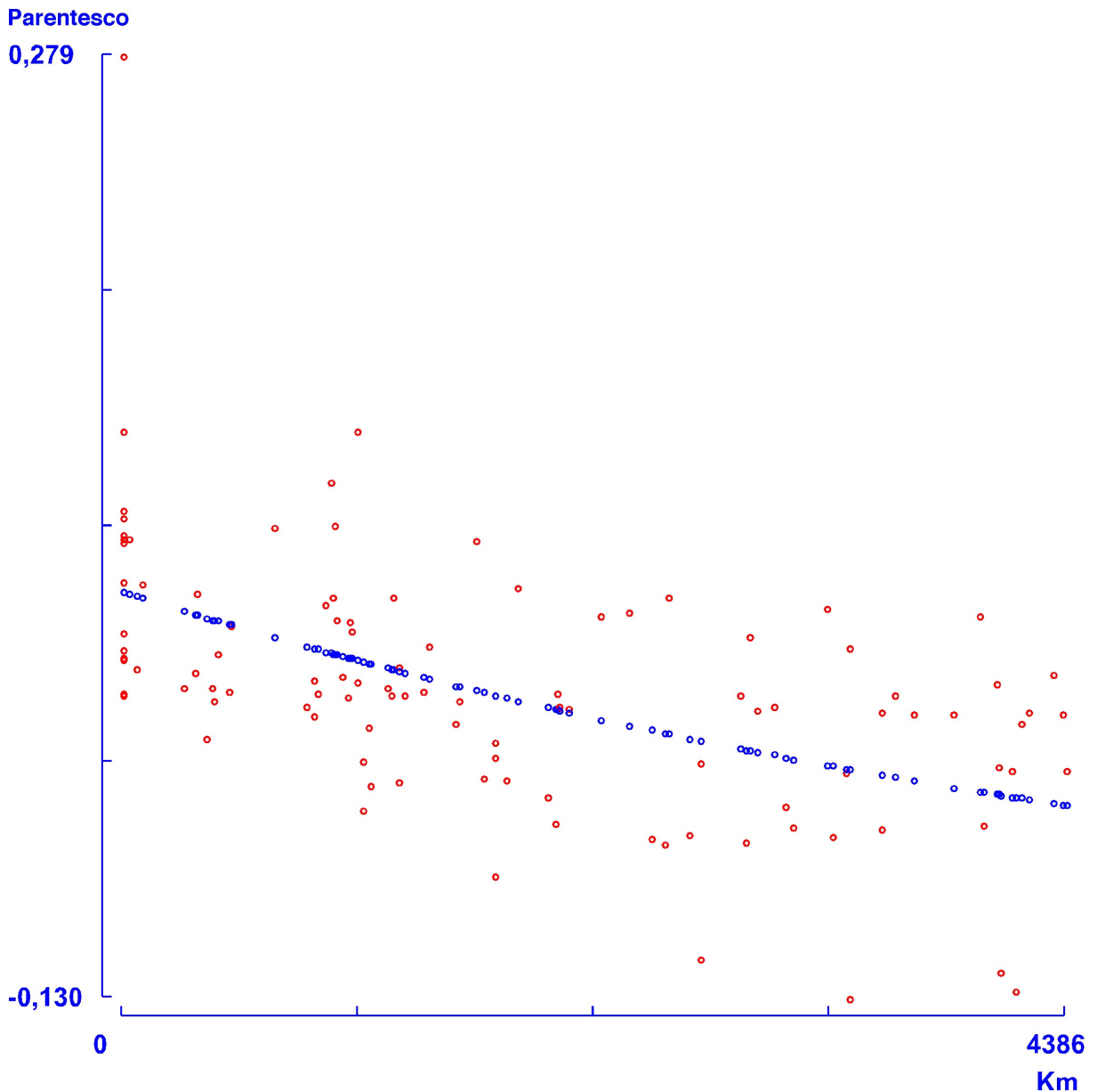
Figura R28. Análisis del centroide sobre frecuencias de haplotipos MAPT en poblaciones europeas.

Se ha realizado un análisis de autocorrelación espacial, cuyo resultado se muestra en la **figura R29**. Se observa una tendencia a la disminución del índice de Moran con la distancia. Aunque el análisis no muestra ningún valor significativo, el patrón descrito es similar al esperado de acuerdo al modelo de aislamiento por la distancia, que genera habitualmente una disminución del parentesco entre las poblaciones conforme aumenta la separación geográfica de las mismas. La ausencia de significación y la irregularidad observada en el patrón descendente del autocorrelograma podrían estar relacionadas con el pequeño número de comparaciones que se obtiene al incluir tan sólo 3 haplotipos y uno de ellos con una distribución poco definida (H2').

A continuación, se ha contrastado precisamente el modelo de aislamiento por la distancia. Se han obtenido unos valores de a:  $0,04632 \pm 0,00842$  y b:  $0,00017 \pm 0,00003$ . Como se observa en la **figura R30**, los valores observados se ajustan razonablemente bien a los esperados.



**Figura R29.** Análisis de autocorrelación espacial de los haplotipos MAPT en Europa.



**Figura R30.** Aislamiento por la distancia observado a partir de los haplotipos MAPT en Europa.

Con el objetivo de comparar la matriz de distancias geográficas con la matriz de distancias genéticas R de Harpending y Jenkins, se ha realizado un test de Mantel. La asociación ha resultado significativa ( $R: 0,3058$ ,  $p: 0,0299$ ), de modo que la posición geográfica de las poblaciones ha condicionado al menos en parte su patrimonio genético.

A partir de los resultados descritos en los últimos párrafos, puede afirmarse que existe en Europa un patrón de distribución espacial de los haplotipos MAPT.

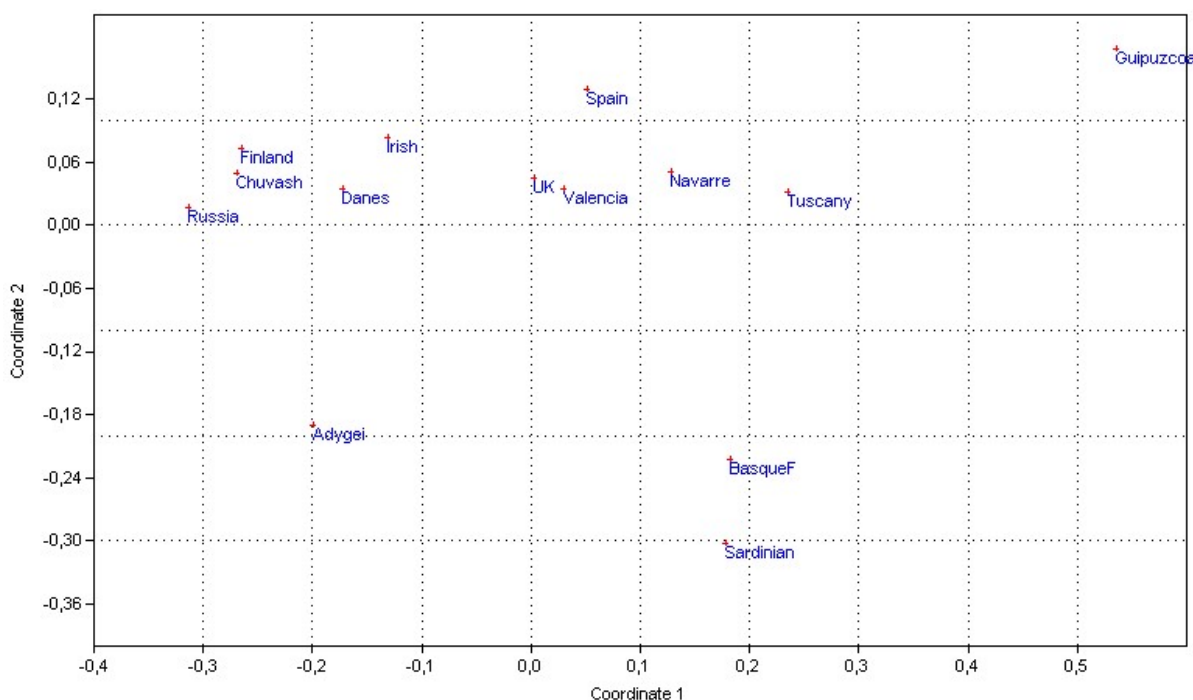
Se han realizado dos análisis de escalamiento multidimensional no métrico, cuyos resultados se muestran en las **figuras R31 y R32**. En el primero se



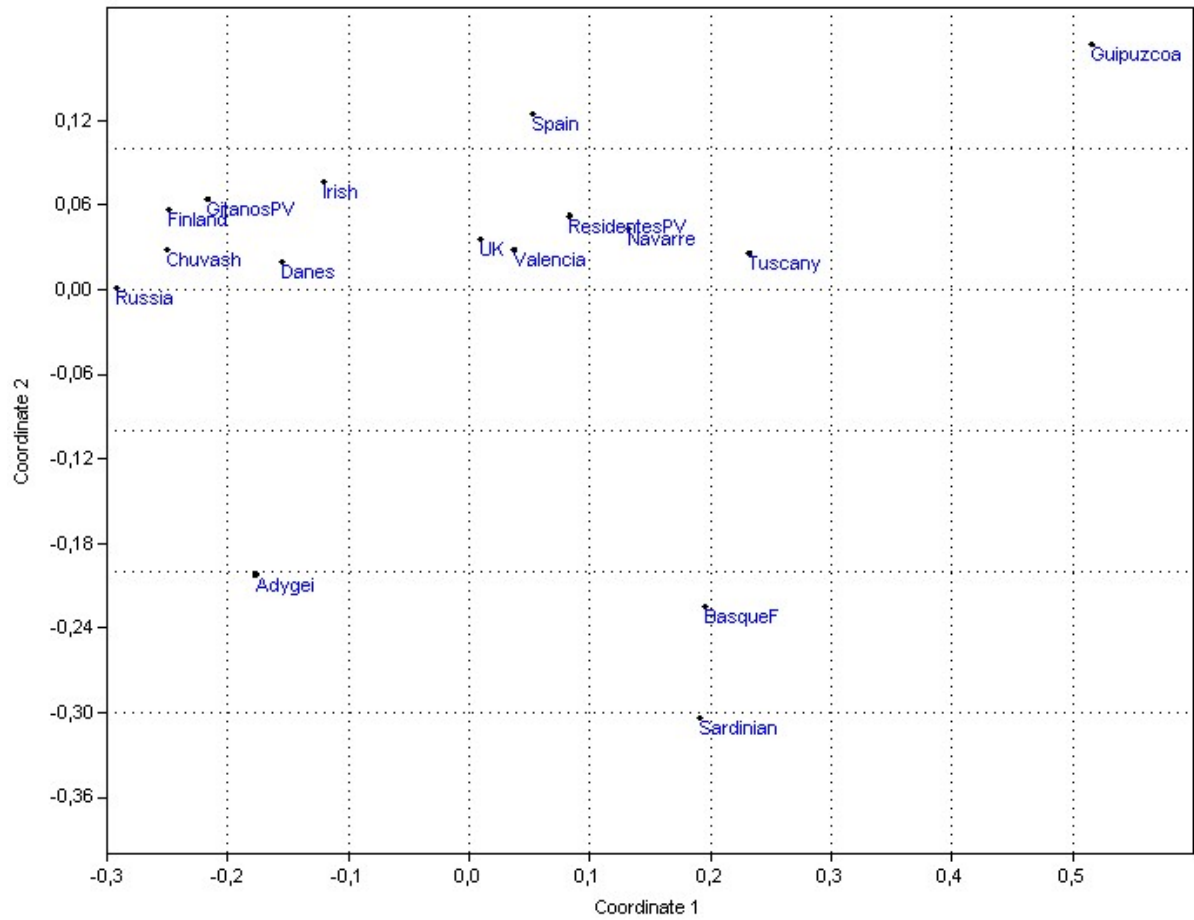
consideran aquellas poblaciones representativas de su región y en el segundo se añaden las muestras de Gitanos y Residentes del País Vasco.

Las representaciones bidimensionales de las matrices R de Harpending y Jenkins explican el 100% de la varianza con un coeficiente de estrés de 0. El MDS, a grandes rasgos, refleja la distribución geográfica de las poblaciones en Europa. Así, en general las poblaciones del norte del continente - con la excepción del Reino Unido - aparecen agrupadas en el cuadrante superior izquierdo, mientras que las poblaciones del sur de Europa se ordenan en el cuadrante superior derecho de la figura. Separadas de estos dos grupos se encuentran los Adygei, vascos de Francia, Sardos y vascos de Gipuzkoa. Parece evidente que la dimensión 1 del análisis se origina por los haplotipos H1' y H2D, en tanto que la 2 está determinada por las frecuencias de H2'.

La disposición prácticamente no varía cuando se añaden los Gitanos y los Residentes del País Vasco, de modo que los Gitanos se ubican entre las poblaciones del norte de Europa y los Residentes entre las poblaciones del sur.



**Figura R31.** Análisis MDS de la matriz de distancias R de Harpending y Jenkins a partir de frecuencias haplotípicas MAPT en Europa.



**Figura R32.** Análisis MDS de la matriz de distancias R de Harpending y Jenkins a partir de frecuencias haplotípicas MAPT en Europa. Se incluyen, respecto a la figura anterior, las muestras de Gitanos y Residentes del País Vasco.

## 2.2. África

En África se observan correlaciones escasamente significativas para los haplotipos H1' y H2D, mientras que no se ha detectado clina para el subhaplotipo H2. Los gradientes de frecuencias obtenidos muestran clinas de orientación nordeste-sudoeste para H1' y oeste-este para H2D (Figura R33).

A pesar de la escasa significancia de estas clinas, se puede apreciar un cierto patrón; los haplotipos H2 están prácticamente ausentes en las poblaciones del Oeste de África mientras que permanecen en los genotipos de las poblaciones del Este de África. Además, el haplotipo H2 se encuentra principalmente en poblaciones cazadoras-recolectoras como los Sengwer (0,083), Mbuti (0,077), Sandawe (0,054), Bakola (0,053), Biaka (0,044), Hadza (0,028) o los San (0,025), aunque no aparece entre los Boni (cazadores recolectores de Kenia) y en los Beja, una etnia de pastores de Sudán, presenta el valor más alto de este haplotipo en el continente, sólo superado a nivel mundial por Cerdeña (0,127).

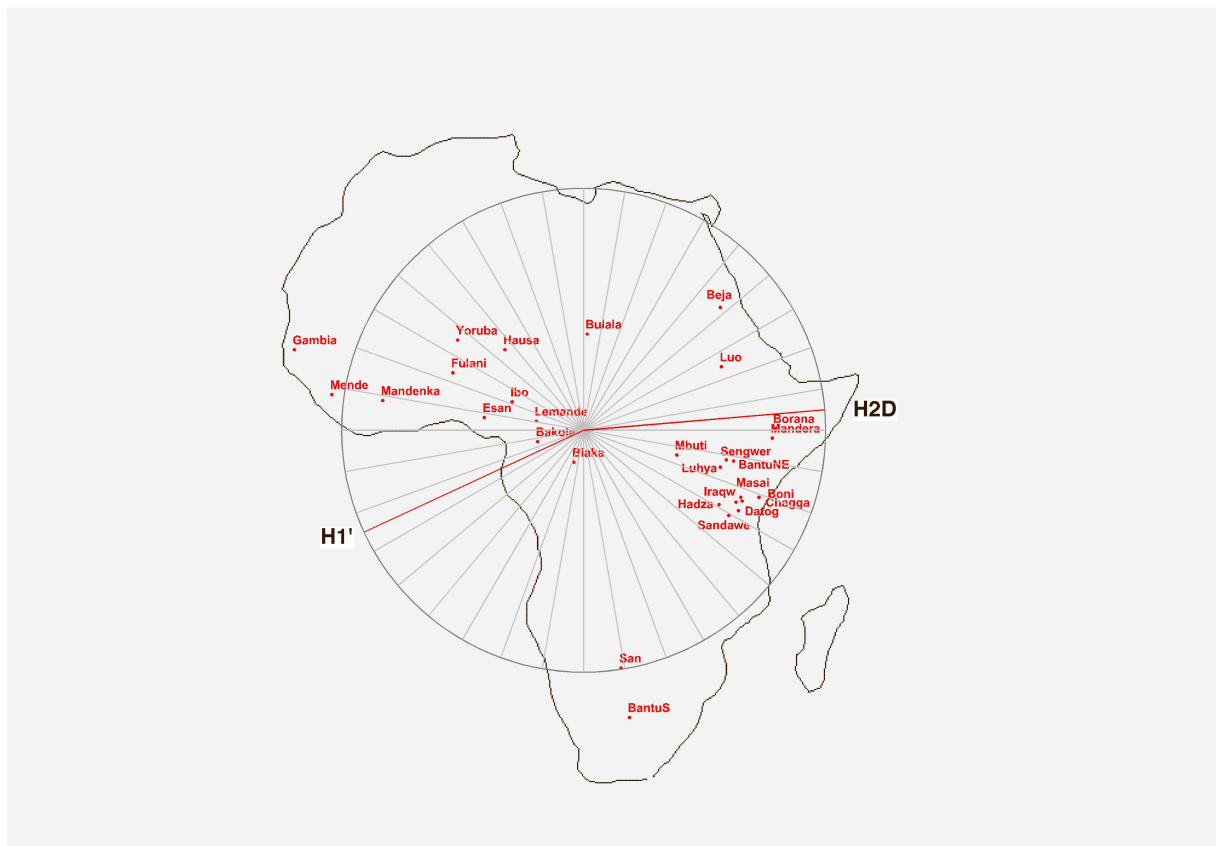


Figura R33. Orientación de las clinas correspondientes a los haplotipos MAPT en África.

La clina de H1', con una orientación de 243 grados, muestra sus frecuencias más bajas en la región centro-este del continente (Figura R34). Ésta aparece algo inclinada respecto al eje norte-sur dado que la frecuencia mínima para este haplotipo en África la presentan los Beja (0,861), un grupo de

pastores que habita el noreste de Sudán. Ausentes en este análisis debido a que sólo existen datos para los haplotipos principales (H1 y H2) y no para los subhaplotipos H1', H2y H2D, cabe mencionar a los Mozabitas del estudio de Donnelly *et al* (2010). Este grupo bereber de Argelia muestra una frecuencia de H1 muy similar a la de los Beja (0,867), siendo las dos más bajas del continente.

En la zona geográfica compuesta por el noreste de la República Democrática del Congo, Kenia y Tanzania se encuentran los grupos humanos con frecuencias de H1 más bajas de África, después de los ya comentados Beja y Mozabites, como son los Sandawe (0,913), los Datog y los Sengwer (0,917), los Boni (0,921), los Mbuti (0,923) o los Masai (0,927). En esta región también se encuentran grupos con frecuencias más altas, como los Iraqw (0,972), los Hadza (0,972) y los Chagga (0,989) e incluso con el haplotipo H1' fijado (Bantú NE, Luo y Manderá).

Las poblaciones que habitan el resto del continente presentan todas frecuencias de H1' mayores que 0,970, salvo los Bakola (0,947). En varios grupos el haplotipo H1' es el único que aparece (Bulala, Lemande, Bantús S, Esan y Fulani).

La significación estadística de la regresión es  $p: 0,0186$ , para un coeficiente de determinación  $R^2: 0,190$ .

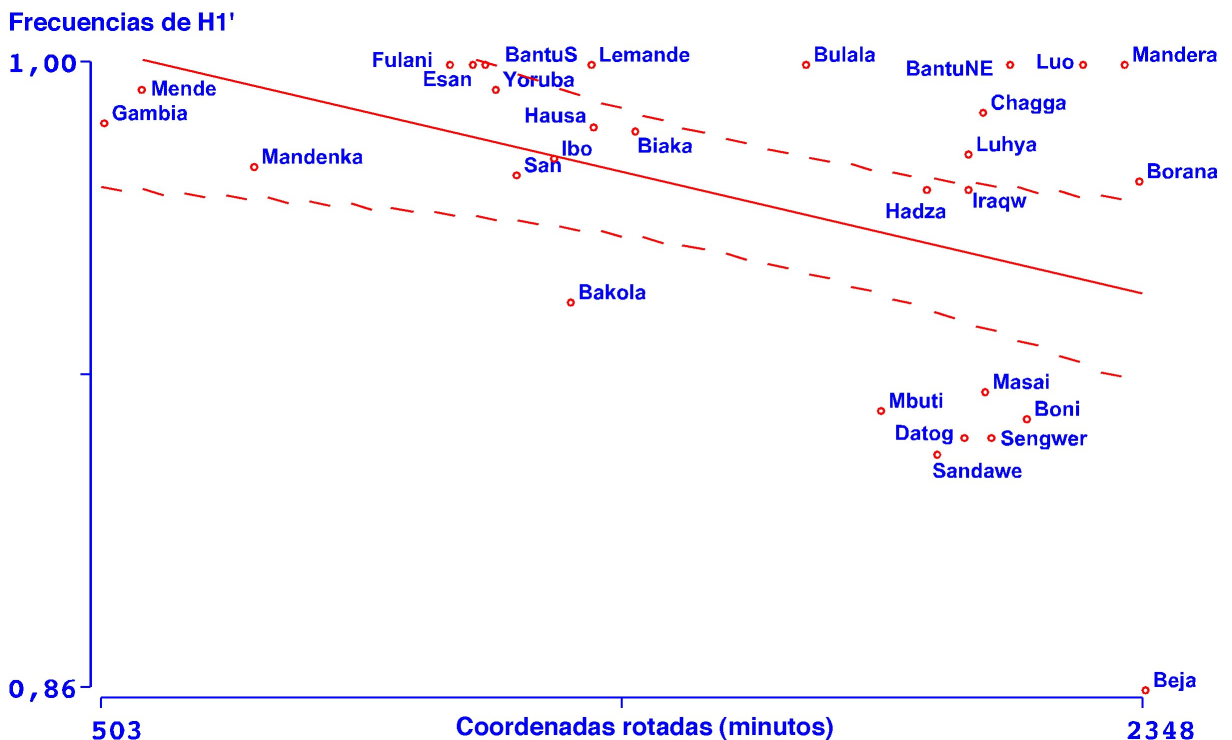


Figura R34. Recta de regresión con intervalos de confianza del 95% de la clina observada para H1' en África.

La clina de H2D, tiene una orientación de 85 grados, prácticamente oeste – este (Figura R35). Muestra sus frecuencias más bajas en el tercio oeste del continente, donde, salvo en los Mandeka (0,023), no alcanza valores polimórficos en ninguna población.

Las frecuencias máximas de este haplotipo en África se encuentran en dos poblaciones que habitan en sur de Kenia, los Boni (0,079), la única población cazadora-recolectora presente en el estudio de Steinberg *et al* (2012) y los Masai (0,049), pastores semi-nómadas que también ocupan el norte de Tanzania.

En torno a esta región se concentran otras de las poblaciones con mayores frecuencias del haplotipo MAPT\*H2D como los Sandawe (0,033), los Beja (0,028), los Datog (0,028), los Iraqw (0,028) y los Borana (0,026). Además de en las poblaciones donde el haplotipo MAPT H1 está fijado, existen otras en las que el subhaplotipo H2D no aparece, como los Hausa, los Mende, los Bakola, los San, los Mbuti y los Sengwer.

La significación estadística de la regresión es  $p: 0,0395$  y el coeficiente de determinación  $R^2: 0,150$ .

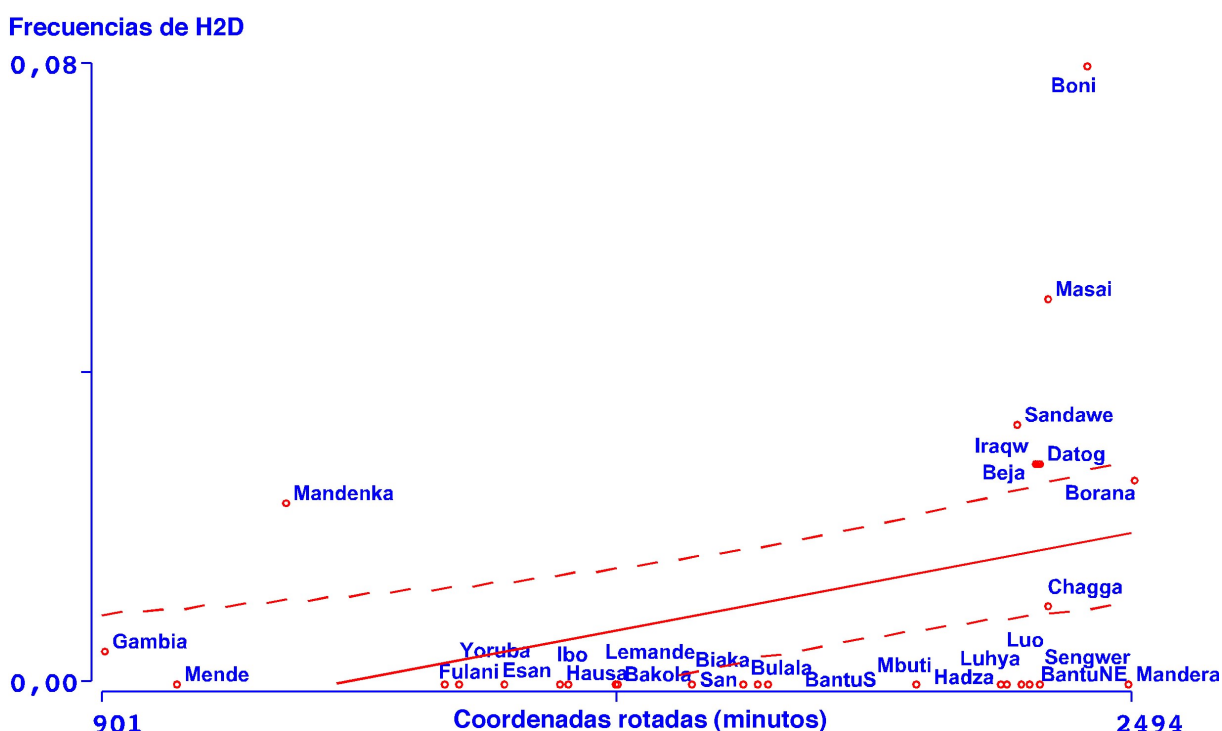


Figura R35. Recta de regresión con intervalos de confianza del 95% de la clina observada para H2D en África.

Como se ha mencionado anteriormente, el **haplotipo H2** aparece principalmente en grupos de cazadores-recolectores, aunque no únicamente en ellos. La distribución de frecuencias de este haplotipo es irregular en el continente, por lo que no se ha hallado una clina. De oeste a este y de norte a sur, en Gambia y entre los Mende y los Mandenka, el haplotipo H2 no aparece – en estos últimos – o se detecta a frecuencias muy bajas (0,009 en Gambia y 0,006 entre los Mende).

Si nos desplazamos más al centro, se mantiene esta tendencia, poblaciones en las que no aparece H2 (Yoruba, Bulala, Fulani, Esan y Lemande, donde el H1 es el único haplotipo detectado) o en las que aparece a frecuencias bajas (Hausa, 0,014; Ibo, 0,021). Al sur de estos grupos, los Bakola y los Biaka muestran frecuencias más altas, cercanas al 0,05 (0,053 y 0,044, respectivamente). Mucho más al sur, los San presentan una frecuencia de H2 de 0,025 y en los Bantús de Sudáfrica el haplotipo H1 está fijado, por lo que la frecuencia de H2 es 0.

En la región este del continente es donde se encuentran más poblaciones con frecuencias de H2 por encima de 0,05 (Mbuti, 0,077; Sengwer, 0,083; Datog, 0,056 y Sandawe, 0,054), incluidos los Beja, la segunda frecuencia de H2 más alta detectada en el mundo hasta ahora (0,111), si bien se intercalan con poblaciones en las que no aparece (Luo, Borana, Mandera, Bantú, Luhya, Iraqw y Boni) o aparece a frecuencias apenas perceptibles, como en los Chagga (0,001) o relativamente bajas (Masai, 0,025; Hadza, 0,028).

En las **figuras R36 y R37** pueden observarse dos mapas sintéticos que representan el gradiente de frecuencias de H1 y H2D en África, respectivamente. Dado que no se ha encontrado clina significativa para el haplotipo H2, no se ha realizado su mapa sintético.

En ambos mapas, resulta notable la presencia de un “centro de ausencia”. En el caso del mapa sintético que representa el gradiente de frecuencias de H1 en el continente africano (**figura R36**), se observa claramente la región centro-este como la zona con menores frecuencias de este haplotipo a partir de la cual aumentan los valores de H1. En el caso del mapa sintético que representa el gradiente de frecuencias de H2D en África (**figura R37**) es la región oeste, la que presenta menores frecuencias de los haplotipos invertidos.

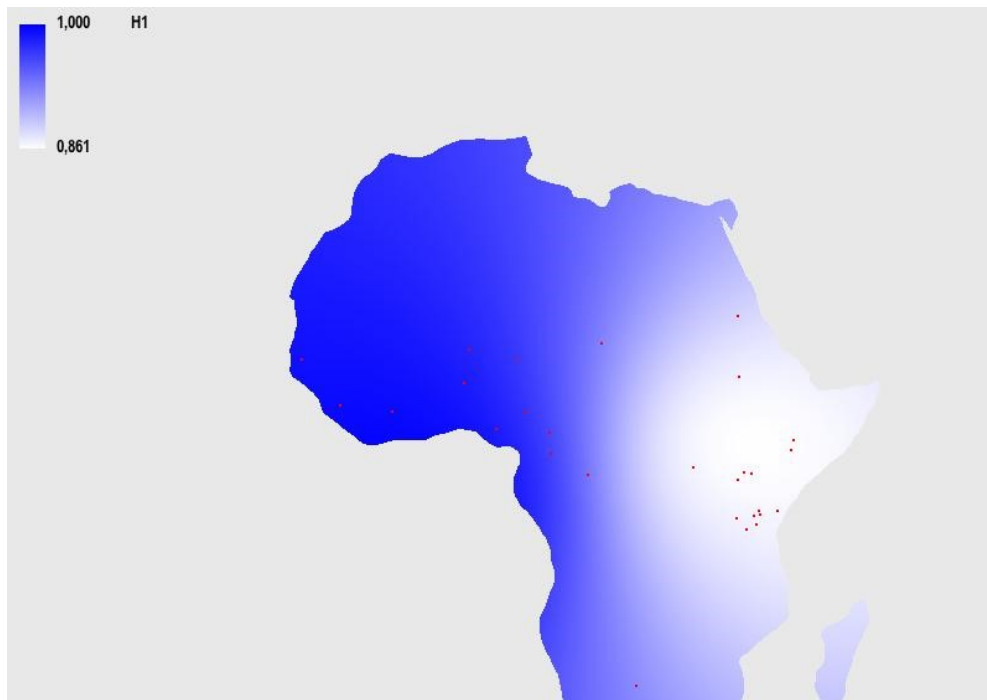


Figura R36. Mapa sintético que representa el gradiente de frecuencias de H1' en África.

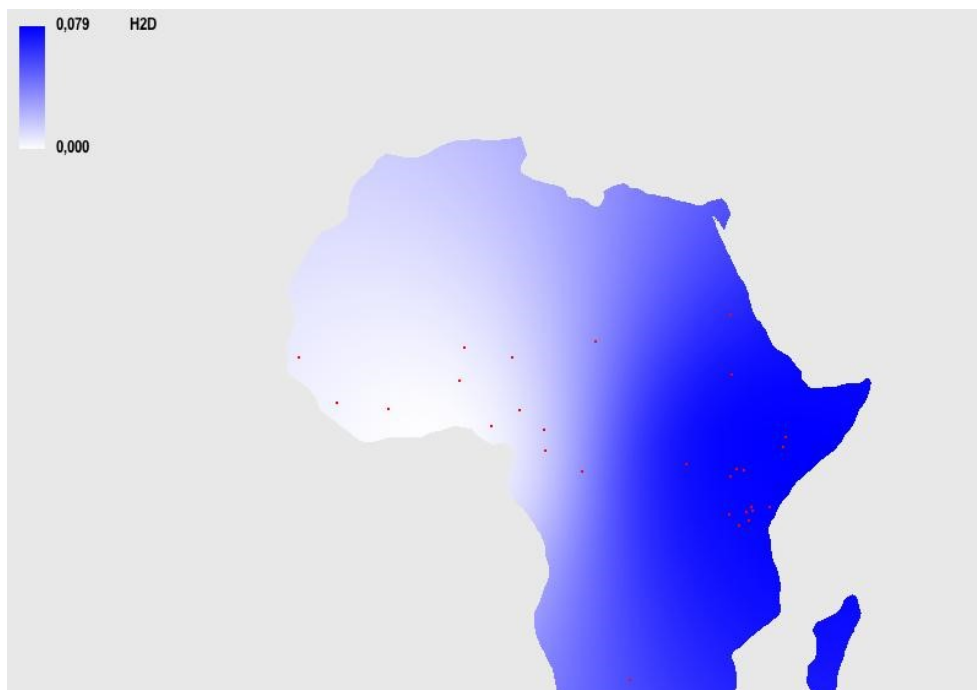
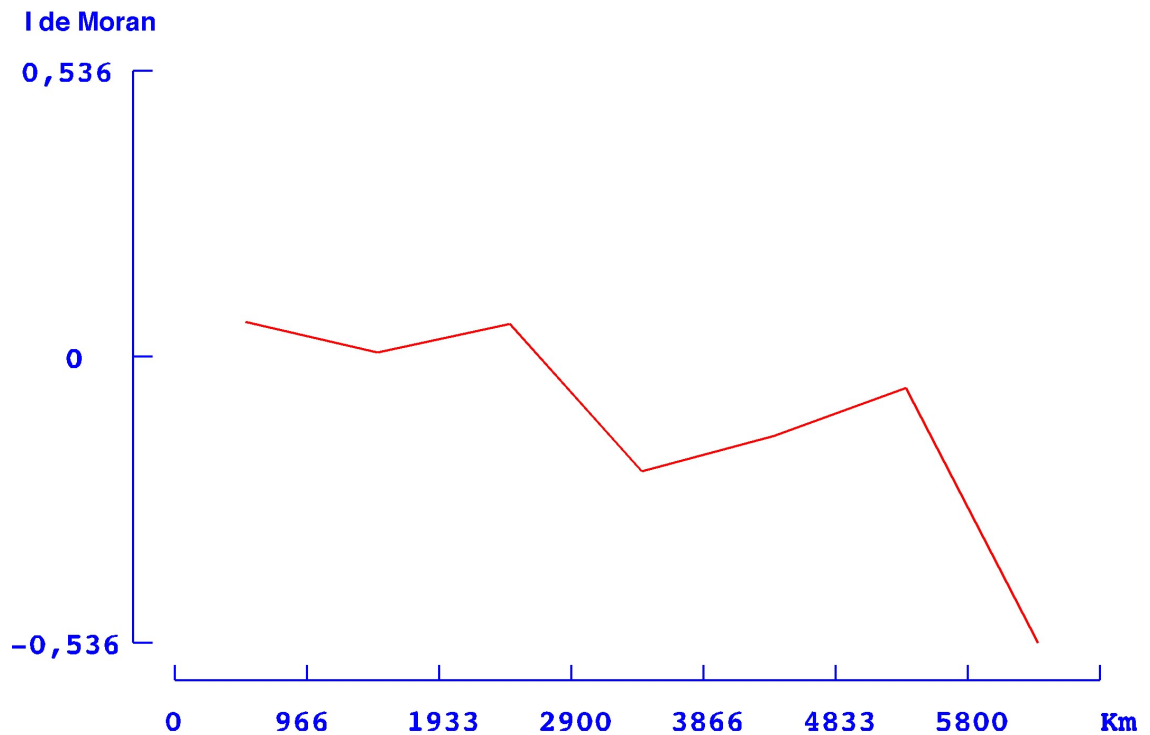


Figura R37. Mapa sintético que representa el gradiente de frecuencias de H2D en África.

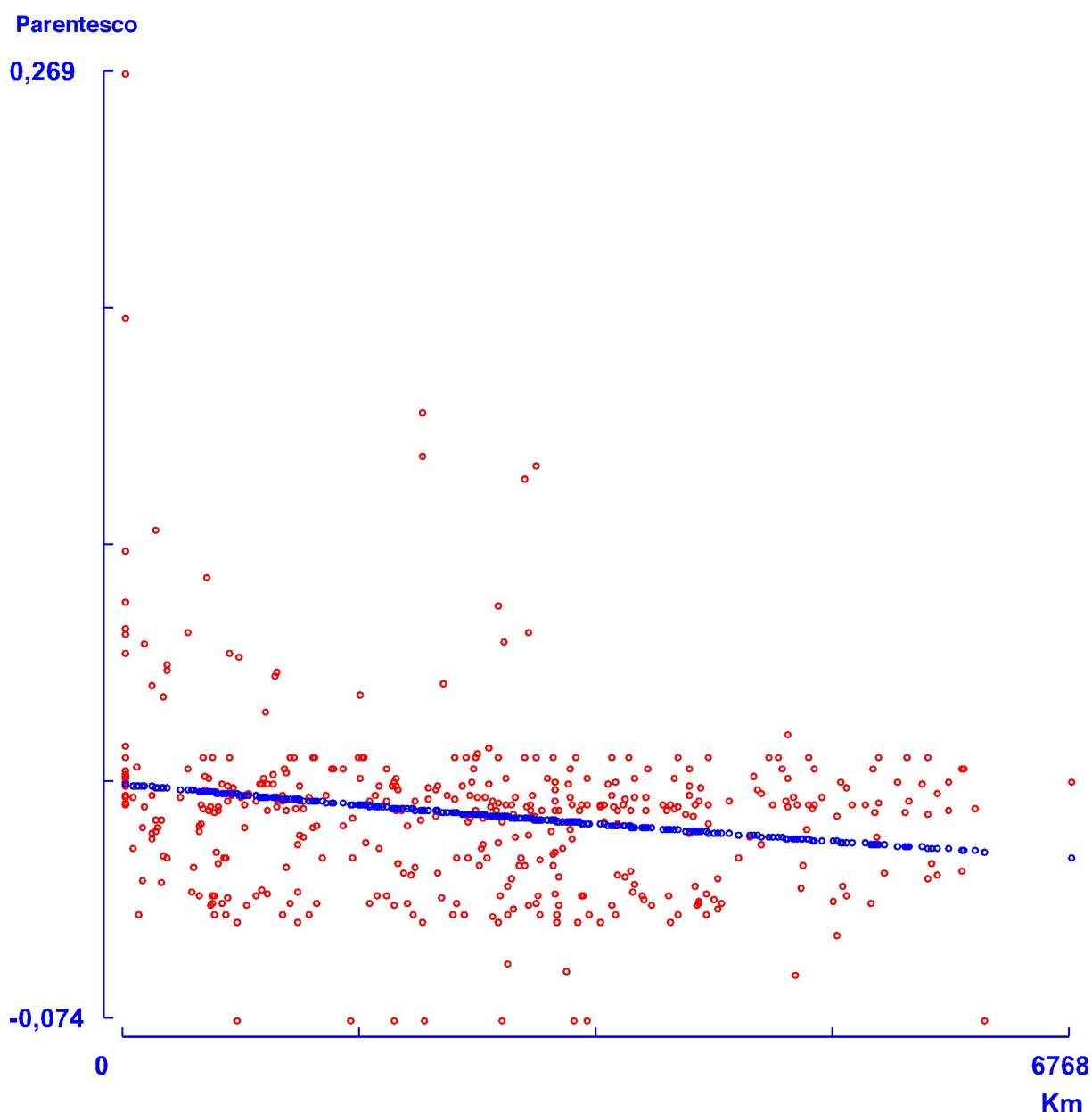
El análisis de autocorrelación espacial (Fig. R38) no sigue el patrón esperado de acuerdo al modelo de aislamiento por la distancia, ya que los valores del Índice de Moran no disminuyen con la distancia, sino que muestran irregularidades sistemáticas.

En la **figura R39** se muestran los resultados del análisis del modelo de aislamiento por la distancia. Se han obtenido unos valores de a:  $0,01170 \pm 0,00297$  y b:  $0,00005 \pm 0,00001$ . Estos valores muestran una tasa de parentesco local (a) y una tasa de declive del parentesco (b) menores que los obtenidos en Europa, lo que implica una tendencia menos acusada de ajuste al modelo de aislamiento por la distancia. De hecho, se observa una disposición de los valores en la gráfica R32 muy homogéneos para diferentes distancias.



**Figura R38.** Análisis de autocorrelación espacial de los haplotipos MAPT en África.





**Figura R39.** Aislamiento por la distancia observado a partir de los haplotipos MAPT en África.

El test de Mantel no mostró correlaciones significativas entre la matriz de distancias geográficas y la matriz de distancias  $R$  de Harpending y Jenkins ( $R: -0,0060$ ,  $p: 0,4321$ ), de modo que la posición geográfica de las poblaciones no parece ser un condicionante fundamental para el patrimonio genético relativo a los haplotipos MAPT en el continente africano.

Se ha realizado un análisis de escalamiento multidimensional no métrico cuyo resultado se muestra en la **figura R40**. Se observan en la mitad negativa del eje I y algo dispersas las poblaciones del este de África con valores de  $H1'$  inferiores a 0,930, quedando en el extremo los Beja, con el valor mínimo (0,861). En la otra mitad se agrupan todas las poblaciones con frecuencias de

Hísuperiores a este valor. Los Bakola (0,947), aparecen entre ambos grupos de poblaciones. El eje II distribuye las poblaciones en función de la frecuencia de H2D, desde el valor máximo de los Boni (0,079) en la parte superior, hasta las poblaciones con valores 0 en la parte inferior.

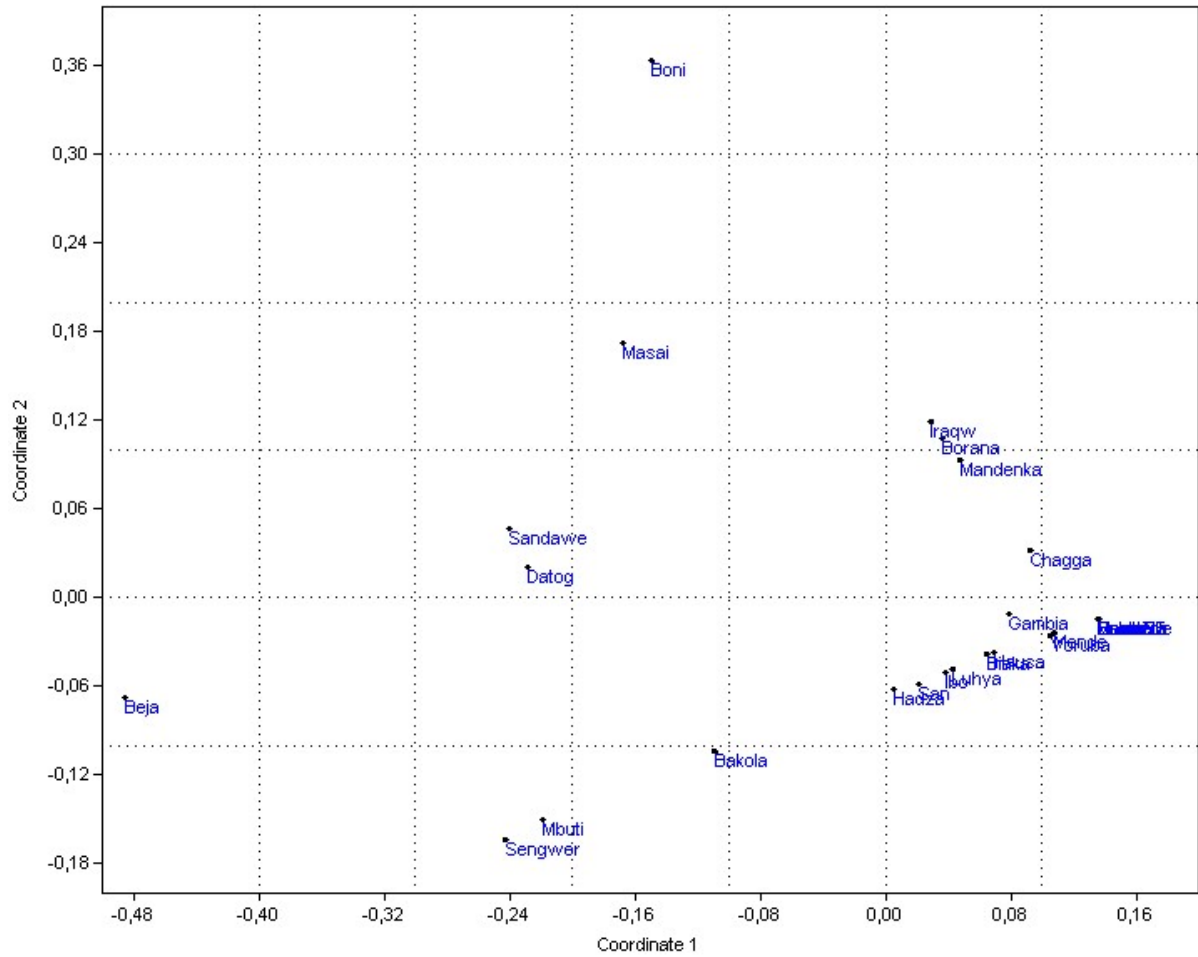
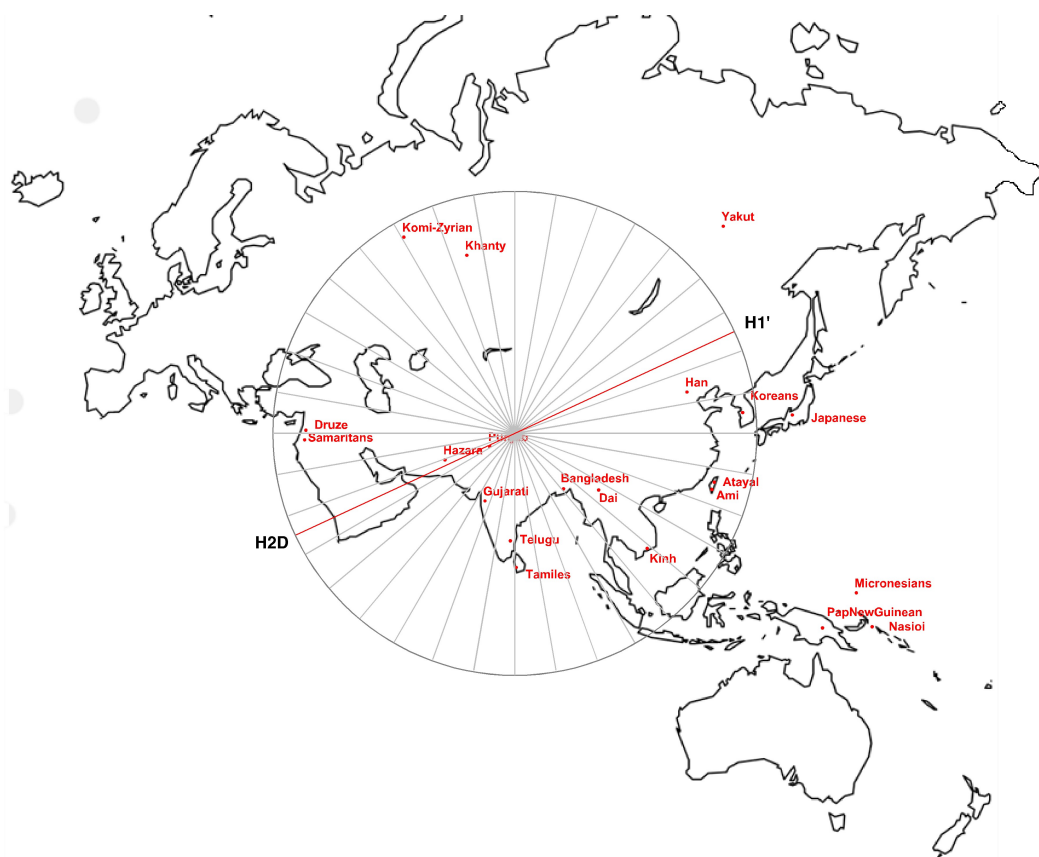


Figura R40. Análisis MDS de la matriz de distancias R de Harpending y Jenkins a partir de frecuencias haplotípicas MAPT en África.

### 2.3. Asia-Pacífico

En la región geográfica Asia – Pacífico se han obtenido gradientes de frecuencias para los haplotipos H1 y H2D de MAPT que muestran clinas de orientación suroeste – nordeste para H1' y nordeste – suroeste para H2D. No existe, por tanto, una acusada orientación norte-sur como en Europa (Fig. R41).

La inversión H2 es más común entre las poblaciones del oeste asiático, más cercanas a Europa, como samaritanos y drusos y de manera más modesta entre los Hazara, los Gujarati y los habitantes del Punjab. El resto de poblaciones analizadas muestran frecuencias de H1 superiores a 0,950, aunque los Telegu y los habitantes de Bangladesh quedan justo en este límite. Es en el este donde se encuentran las frecuencias máximas del haplotipo H1, incluyendo muchas poblaciones donde se ha fijado, como los Dai, los Kinh, las etnias indígenas de Taiwan Ami y Atayal, coreanos, Nasioi y Nueva Guinea.



**Figura R41.** Orientación de las clinas correspondientes a los haplotipos MAPT en Asia y el Pacífico.

La clina de H1', con una orientación de 65 grados y una significación estadística de la regresión de  $p: 0,0001$  ( $R^2: 0,550$ ) (fig.: R42), muestra su máxima frecuencia en las poblaciones situadas más al este, donde se ubican todas aquellas donde el haplotipo MAPT\*H1 se ha fijado, como se ha mencionado en el párrafo anterior. En esta región el valor mínimo de H1 lo

presentan los micronesios (0,97), quedando el resto de poblaciones con frecuencias por encima de este valor (Yakut, 0,980; Han y Japoneses, 0,990).

En la zona centro los valores decaen ligeramente entre las etnias rusas Khanty y Komi-Zyrian (0,970), los Tamiles (0,971), los Telegu (0,951) y los Bangladesíes (0,953). En esta misma región geográfica y central de la clina se encuentran tres poblaciones con frecuencias de H1 considerablemente más bajas que las comentadas hasta el momento. Los Hazara (0,930), los Gujarati (0,910) y los Punjabíes (0,901).

Los valores más bajos se hallan en los Drusos y Samaritanos, cuyas frecuencias del haplotipo H1 (0,690 y 0,640 respectivamente) están muy por debajo de los valores esperados para la región. Aunque son las poblaciones más cercanas a Europa, donde el haplotipo H1 muestra frecuencias mínimas a nivel mundial, se trata además de grupos con altas tasas de endogamia.

La clina para el subhaplotipo MAPT\*H2D (**fig. R43**), como en el caso de Europa, es prácticamente complementaria a la del haplotipo H1, con una orientación de 245 grados, noreste - sudeste. Los Hazara componen la única población que sin ser monomórfica para el haplotipo H1 presentan una frecuencia 0 del subhaplotipo H2D. Otras poblaciones en las que la frecuencia del suhaplotipo H2' es mayor que la del H2D son los Komi-Zyrian y los micronesios (ambos 0,020 para H2'y 0,010 para H2D).

Las poblaciones de noreste, Khanty, Yakut, Han y japoneses, muestran frecuencias muy bajas de H2D y no se detecta presencia del haplotipo invertido sin duplicación (H2).

El coeficiente de determinación de la regresión de esta clina es 0,470 y su significación estadística p: 0,0006.

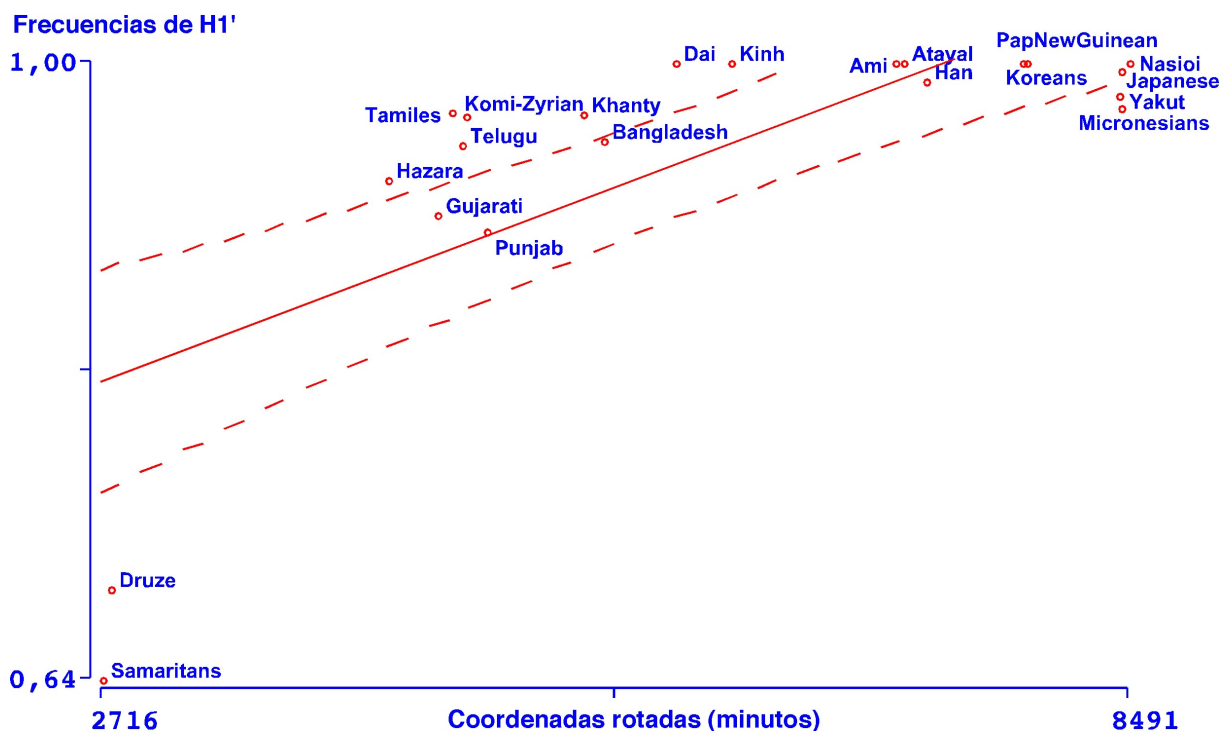


Figura R42. Recta de regresión con intervalos de confianza del 95% de la clina observada para H1' en Asia y el Pacífico.

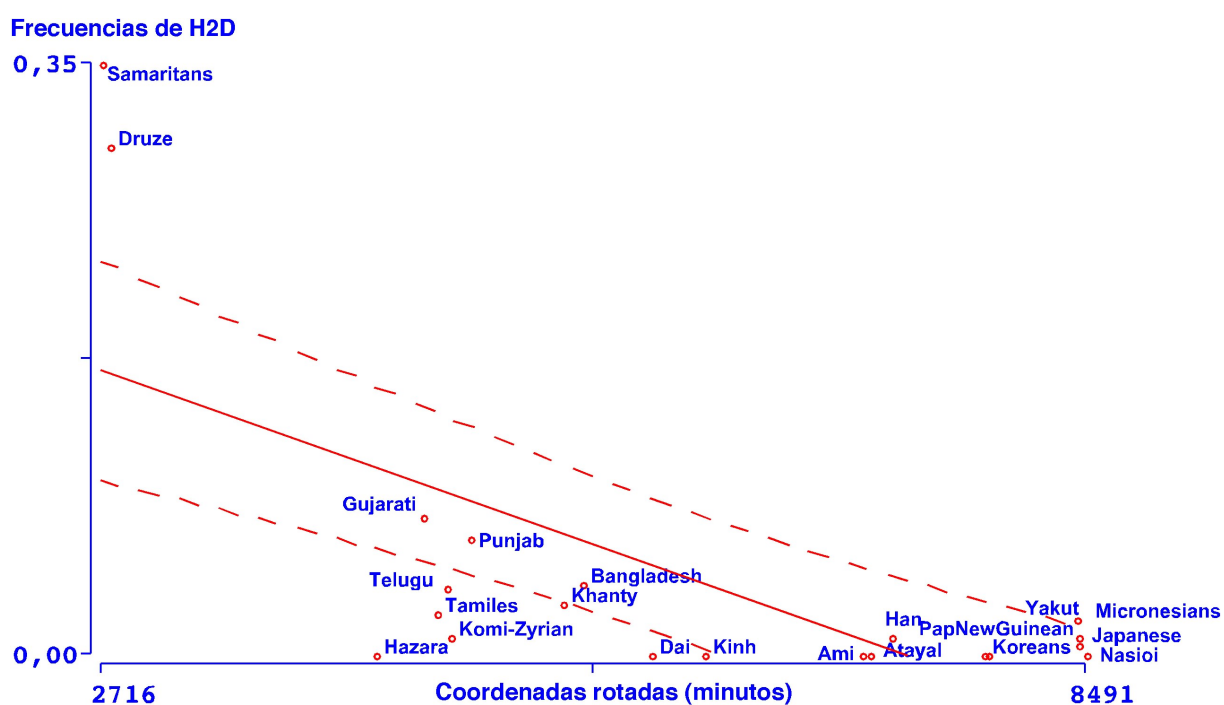
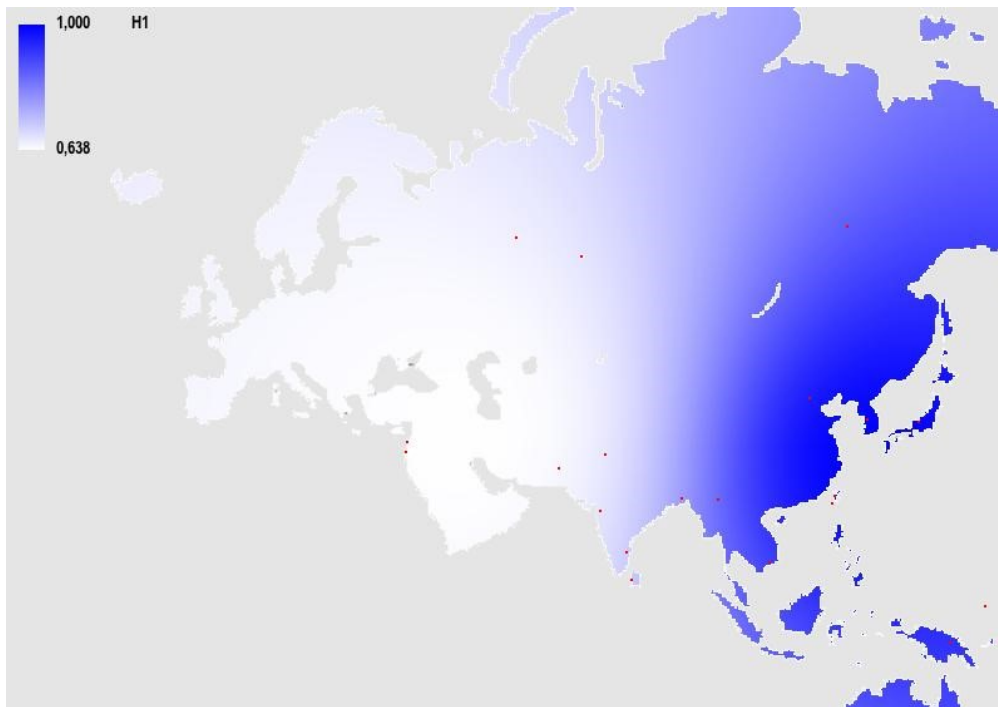


Figura R43. Recta de regresión con intervalos de confianza del 95% de la clina observada para H2D en Asia y el Pacífico.

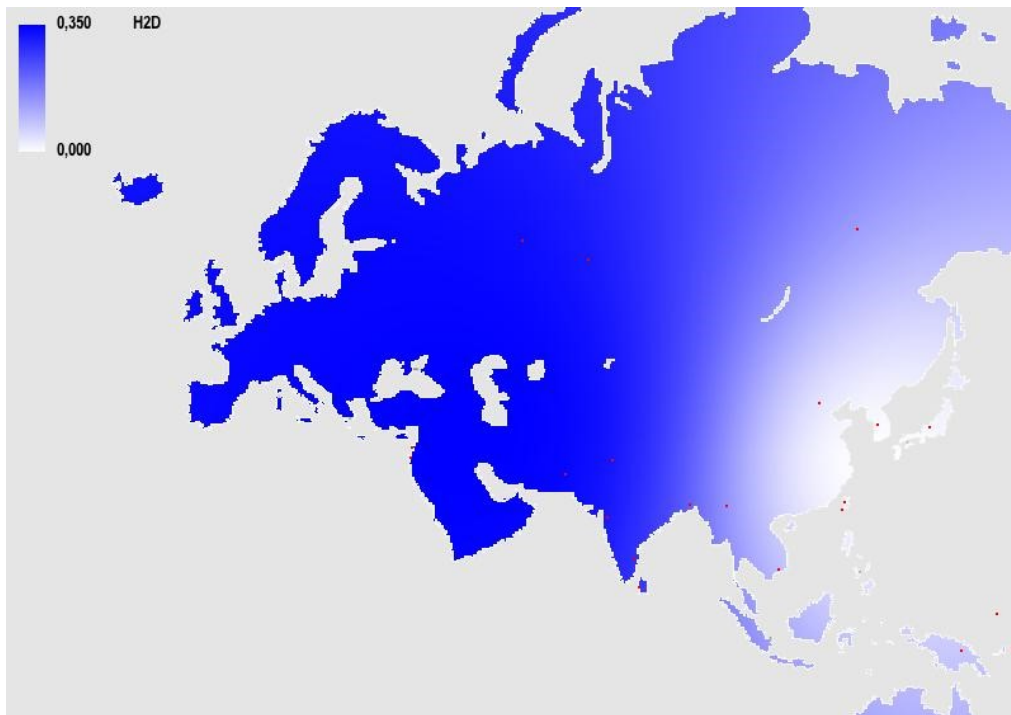
En las figuras R44 y R45 se han representado los gradientes de frecuencias de H1 y H2D respectivamente en la región Asia-Pacífico.

El mapa sintético del haplotipo H1 (**Fig. R44**) muestra el cambio abrupto en frecuencias entre Próximo Oriente por un lado y el resto del continente por otro. Entre las poblaciones con frecuencias mínimas de H1, Drusos (0,690) y Samaritanos (0,640) y las poblaciones con frecuencias moderadas como Punjab (0,901), Gurujati (0,910) y Hazara (0,930), no existen grupos con valores de frecuencias intermedias, que hagan de esta transición un cambio más gradual.

En el mapa sintético relativo al haplotipo H2D (**Fig. R45**) llama la atención un “núcleo de ausencia” del mismo en la zona sudeste, donde se encuentran todas las poblaciones con el haplotipo H1 fijado, además de los Han, Japón y Micronesia, que presentan una frecuencia de 0,01 para H2D. Las poblaciones más norteñas, Komi-Zyrian, Yakut y Khanty, también presentan unas frecuencias mínimas para este haplotipo (0,01, 0,02 y 0,03, respectivamente), aunque no queda tan patente en el mapa sintético. La frecuencia 0 de los Hazara para este haplotipo queda enmascarada por sus poblaciones vecinas que presentan frecuencias de entre 0,068 (Punjab) y 0,350 (Samaritanos).



**Figura R44.** Mapa sintético que representa el gradiente de frecuencias de H1' en Asia y el Pacífico.



**Figura R45.** Mapa sintético que representa el gradiente de frecuencias de H2D en Asia y el Pacífico.

En la **figura R46** se presenta el correlograma para los haplotipos MAPT en Asia. Aunque se puede observar una tendencia a la disminución del Índice de Moran con la distancia, se observan irregularidades. El análisis no muestra ningún valor significativo.

Se ha contrastado el modelo de aislamiento por la distancia (**Fig. 47**) y se han obtenido unos valores de a:  $0,10821 \pm 0,01949$  y b:  $0,00009 \pm 0,00001$ . Estos valores muestran una tasa de parentesco local (a) muy alta y una tasa de declive del parentesco (b) muy baja. Ello implica un escaso ajuste al modelo de aislamiento por la distancia y refleja la existencia de aislados genéticos que condicionan el valor del parentesco local.

Esto se confirma al realizar un análisis del centroide (Harpending y Ward, 1982). En la **figura R48** se observa que las poblaciones de Drusos y Samaritanos reflejan una heterocigosidad mucho mayor de la esperada en relación al conjunto de poblaciones asiáticas, lo que implica un exceso de flujo génico. Teniendo en cuenta que son aislados genéticos, la explicación más adecuada puede encontrarse en un origen diferenciado del resto de poblaciones incluidas en el análisis. De hecho, presentan frecuencias de H1' muy por debajo de las otras poblaciones asiáticas y frecuencias de H2D muy por encima.

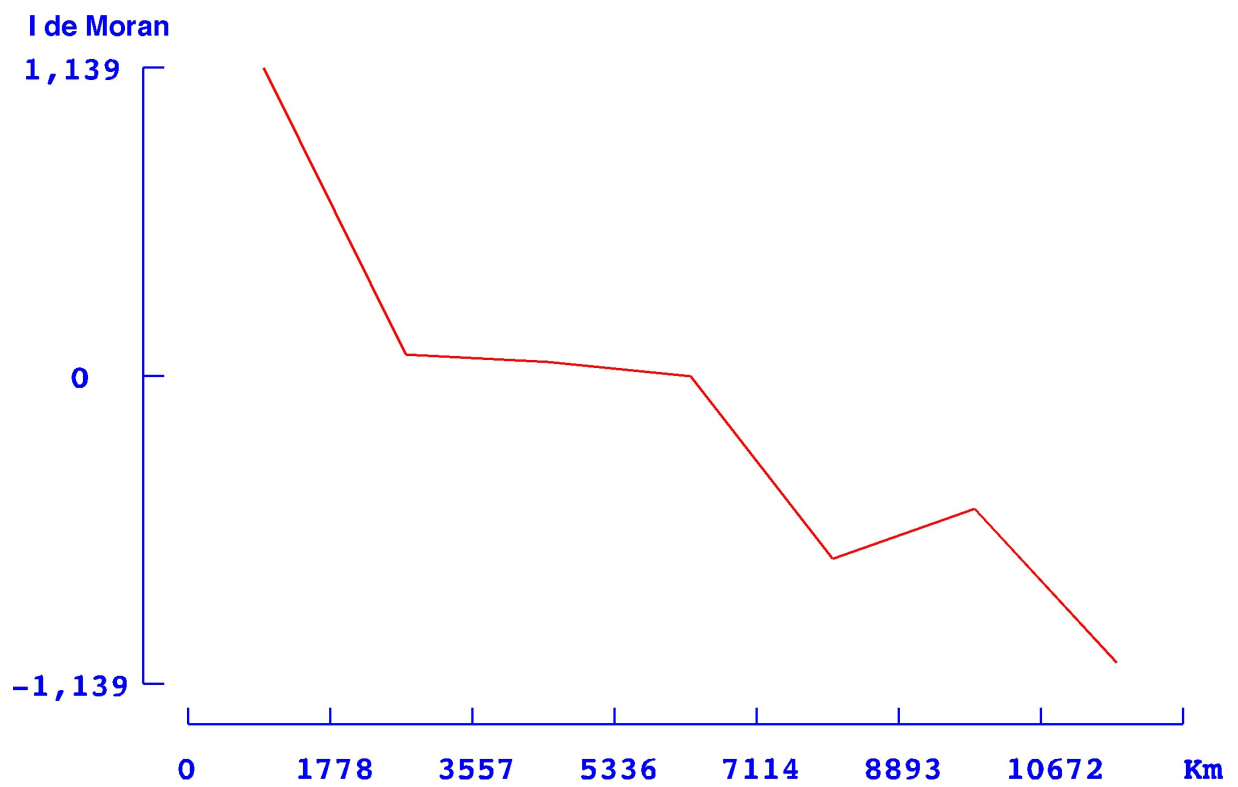


Figura R46. Análisis de autocorrelación espacial de los haplotipos MAPT en Asia y el Pacífico.



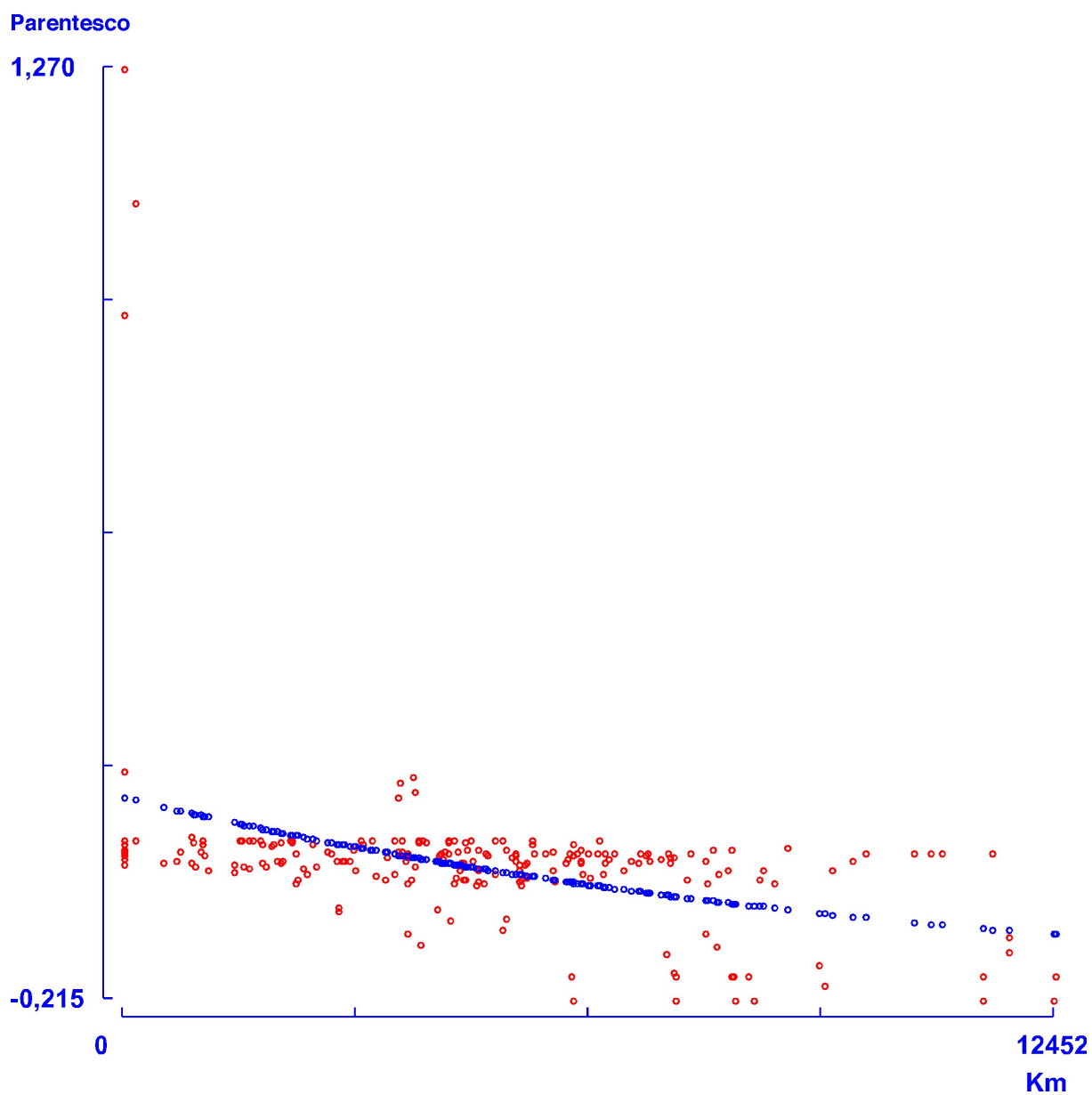


Figura R47. Aislamiento por la distancia observado a partir de los haplotipos MAPT en Asia y el Pacífico.

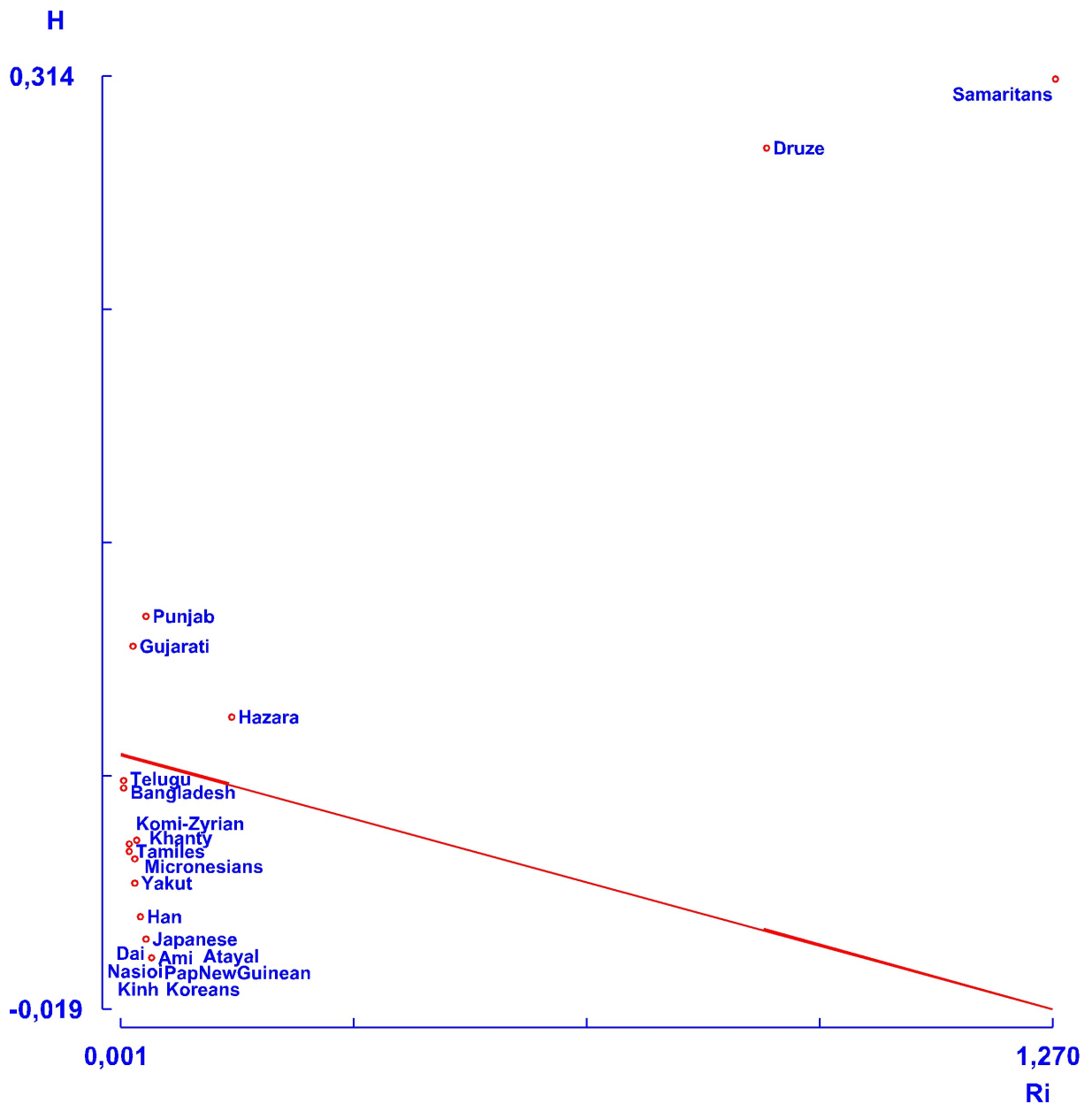
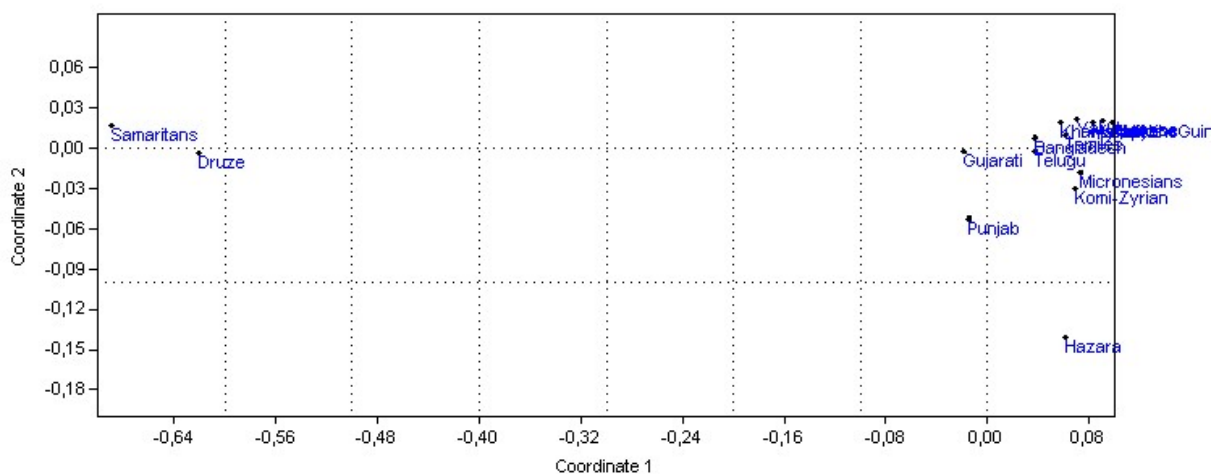


Figura R48. Análisis del centroide en poblaciones asiáticas: heterocigosidad observada frente a distancia del centroide. La línea representa la heterocigosidad predicha.

El test de Mantel mostró correlaciones significativas entre la matriz de distancias geográficas y las matrices de distancias R de Harpending y Jenkins (R: 0,3925, p:0,0010) y  $F_{ST}$  (R: 0,4370, p: 0,0002), indicando que, al menos en parte, el patrimonio genético de las poblaciones analizadas está condicionado por su posición geográfica.

En el análisis MDS realizado sobre las frecuencias haplotípicas de Asia y Oceanía (figura R49) se observa nuevamente la notable heterogeneidad de Drusos y Samaritanos respecto del resto de poblaciones asiáticas. Si el eje I refleja sobre todo las variaciones en H1', el eje II muestra fundamentalmente la influencia de H2', que presenta unos valores máximos en Hazara y Punjab, las poblaciones que ocupan su extremo inferior.



**Figura R49.** Análisis MDS de la matriz de distancias R de Harpending y Jenkins a partir de frecuencias haplotípicas MAPT en Asia y el Pacífico.

## 2.4. América

En este continente no se han encontrado clinas significativas para los haplotipos MAPT. De hecho, se observa una distribución irregular de frecuencias por la que aparecen en dos poblaciones próximas los valores máximos (Ticuna, 1,00) y mínimo (Quechua: 0,935) de H1', siempre dentro de una escasa heterogeneidad. Así, ninguna población alcanza el valor 0,05 para H2' o H2D. Puesto que no hay tendencias que puedan reflejar, no se han realizado mapas sintéticos. En todo caso, el análisis de autocorrelación espacial (**figura R51**) refleja perfectamente el patrón irregular de distribución de frecuencias en las poblaciones analizadas de este continente.

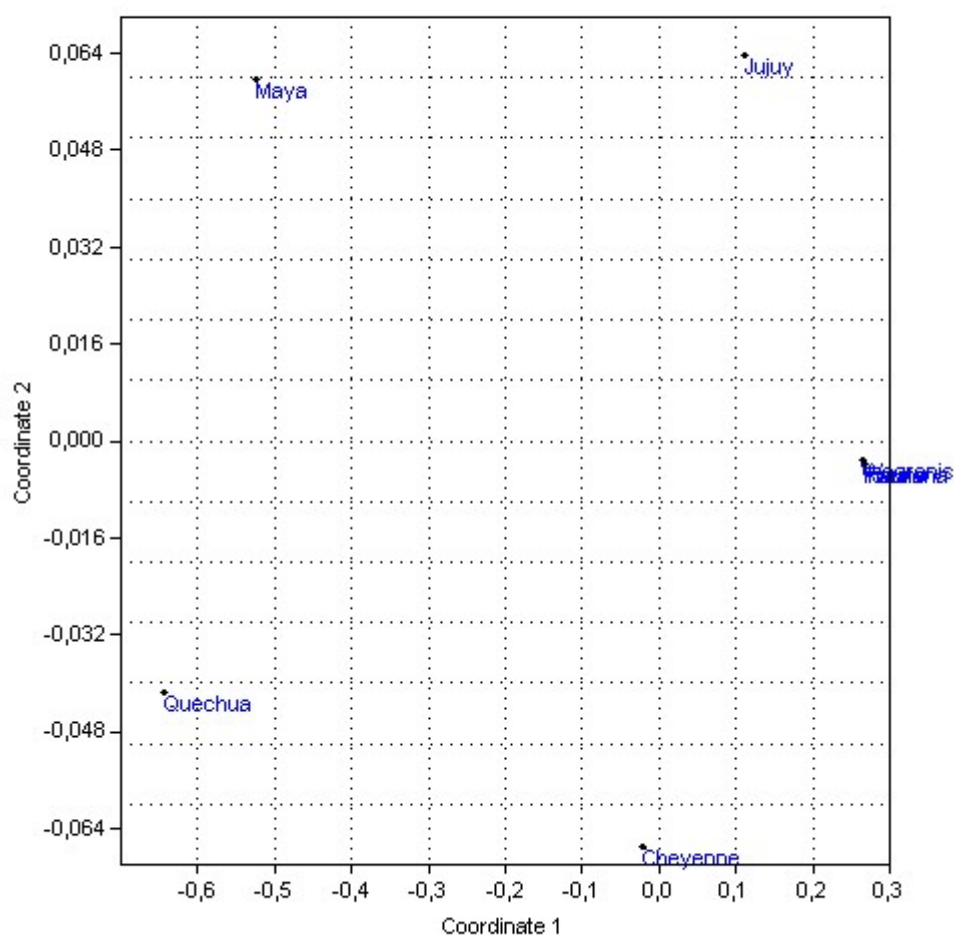


Figura R50. Poblaciones analizadas para los haplotipos MAPT en América.

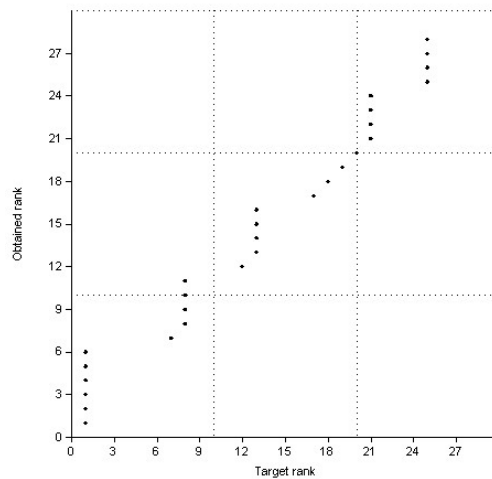


Figura R51. Análisis de autocorrelación espacial de los haplotipos MAPT en América.

Se ha realizado un análisis MDS sobre la matriz de distancias R de Harpending y Jenkins (**figura R52**), observándose una ordenación en el primer eje de acuerdo a las frecuencias de H1', con un valor mínimo en Quechuas y Mayas. El eje II, sin embargo, no refleja las variaciones de H2' o H2D, quedando las poblaciones ubicadas de un modo aparentemente aleatorio. Fruto de esta situación es el relativamente alto coeficiente de estrés (0,12) y el diagrama de Sephard de la **figura R53**.



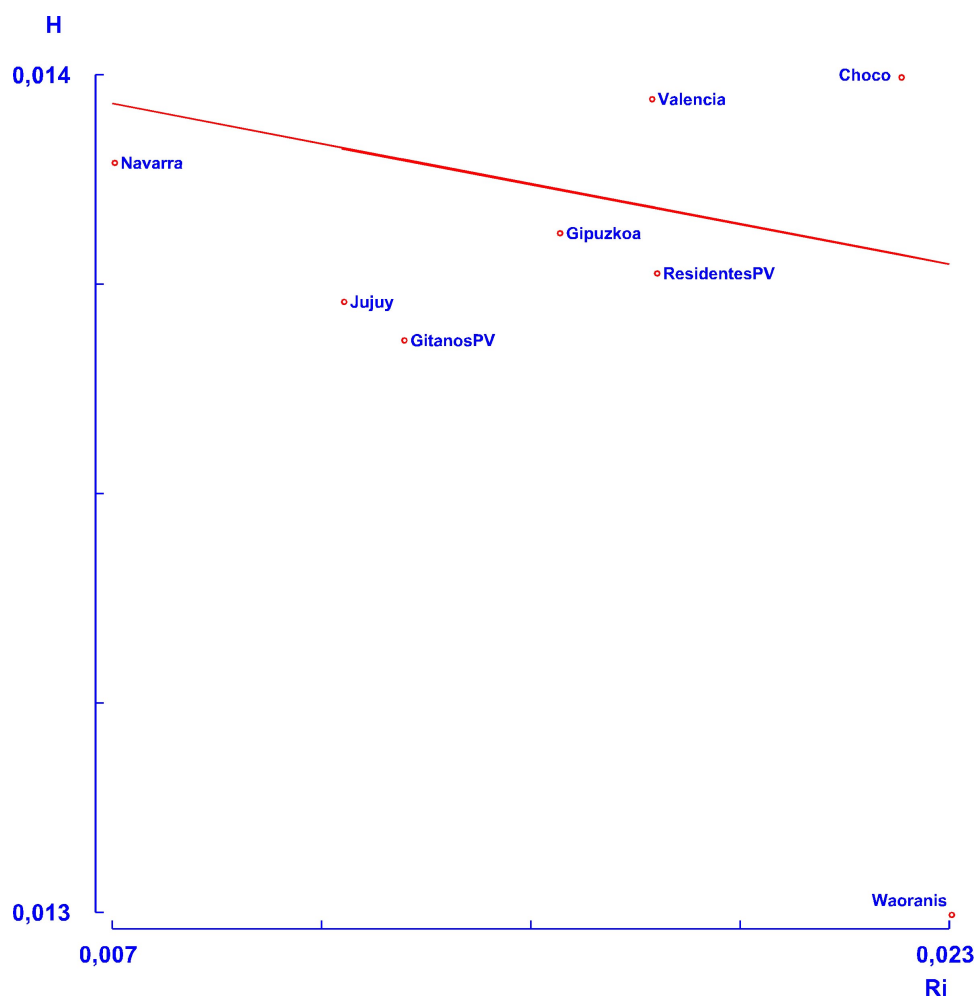
**Figura R52.** Análisis MDS de la matriz de distancias R de Harpending y Jenkins a partir de frecuencias haplotípicas MAPT en América.



**Figura R53.** Diagrama de Sephard correspondiente al análisis MDS de poblaciones americanas.

### Sección 3. Distribución geográfica de los haplotipos SNP/STRP.

Al disponerse de un número limitado de muestras con haplotipos SNP/STRP disponibles, únicamente las de este trabajo, son escasos los análisis estadísticos relevantes que han podido realizarse. Entre ellos, el método del centroide confirma el alto grado de aislamiento de la población Waorani, que aparece ubicada en el extremo inferior derecho y el notable flujo génico entre afroamericanos y amerindios de la población de afrocolomboanos de Chocó, que aparece en el extremo superior derecho (figura R54)



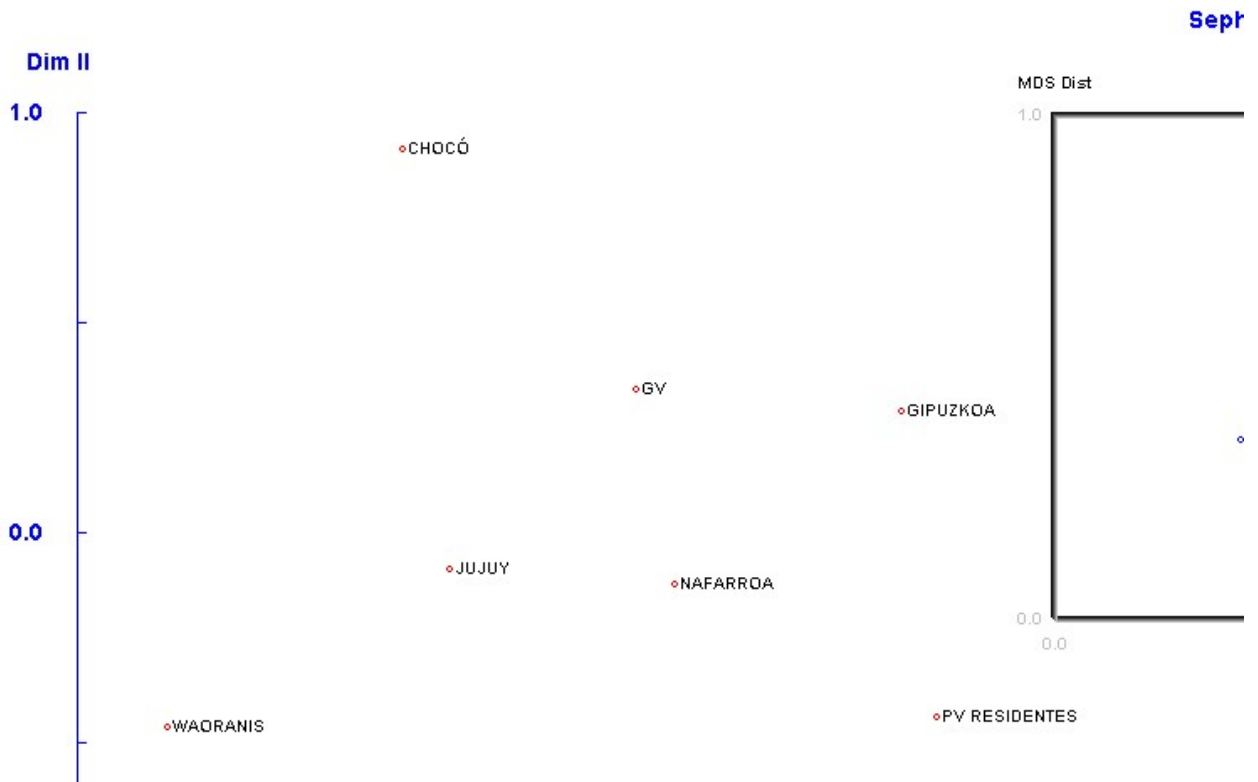
**Figura R54.** Análisis del centroide en poblaciones asiáticas: heterocigosidad observada frente a distancia del centroide. La línea representa la heterocigosidad predicha.

Los resultados del análisis de escalamiento multidimensional (MDS) aplicado a la matriz de distancias genéticas se muestran en la **figura R55**. No se ha tenido en cuenta el SNP rs10514879 ya que no se dispone de sus frecuencias en la población de Chocó. El coeficiente de estrés de la representación es del 5,6%, bastante bajo considerando que se trata de un análisis con poblaciones de diferentes continentes y varios marcadores de características heterogéneas.

Las poblaciones americanas se agrupan en el extremo negativo de la dimensión I, dispersas en el eje de la dimensión II, Jujuy y Waoranis en su mitad negativa y los afrodescendientes del Chocó, en el extremo positivo.

Las poblaciones europeas aparecen en la región positiva de la dimensión I. La población de gitanos vascos presenta una posición intermedia en este eje,

debido sin duda a la falta de datos de poblaciones asiáticas de referencia para los haplotipos extendidos analizados en este proyecto.



**Figura R55:** Gráfico de los resultados del análisis de escalamiento multidimensional (MDS) aplicado a la matriz de distancias genéticas R de Harpending y Jenkins para las ocho poblaciones del estudio. Las distancias genéticas fueron calculadas basadas en las frecuencias de seis marcadores (el Indel, los microsatélites MAPT07, MAPT08, MAPT09 y MAPT14 y el SNP rs199451).

#### Sección 4. Perspectiva histórica en el contexto geográfico de Europa y del Oeste de Asia.

##### 4.1. Características/ Datos de las muestras analizadas.

Las 33 muestras de pobladores prehistóricos analizadas para este proyecto quedan resumidas en la **tabla R17**, donde se especifica, además del nombre de la muestra, su frecuencia para la configuración H2, el número de individuos que la componen y la edad promedio de los mismos.



**Tabla R17.** Nombre de las poblaciones analizadas, frecuencia del haplotipo H2 hallado en las mismas, número de individuos que componen las muestras y edad promedio de los mismos.

Nombre población	Frecuencia H2	Núm. de individuos	Edad promedio
Alfold Linear Pottery	0,107	9	6750
Anatolia Neolithic	0,167	24	6137
Asia Bronze Age	0,136	44	3805
Asia Copper Age	0,143	7	4347
Asia Iron Age	0,294	17	5080
Asia Neolithic	0,125	8	4932
Balkans Bronze Age	0,167	6	3535
Balkans Chalcolithic	0,091	11	4998
Balkans Neolithic	0,333	15	6825
Beaker Britain	0,222	18	3912
Beaker Central Europe CzechRep.-Hungary-Poland	0,182	33	4025
Beaker Central Europe Germany-Netherlands	0,071	28	4128
Beaker Iberia-France-Italy	0,000	11	3887
Britain Neolithic	0,360	25	4831
Bronze Age Russia	0,071	14	4262
Central Europe Bronze Age	0,188	16	3787
Corded Ware Central Europe	0,000	7	4461
England CA EBA	0,200	15	3870
England Middle-Late Bronze Age	0,412	17	3190
Germany Czech Neolithic	0,133	12	5257
Hungary - Baden Copper Age	0,091	11	5335
Hungary Neolithic	0,107	28	6499
Iberia Chalcolithic	0,067	15	4462
Iberia-France Neolithic	0,083	12	6417
Iron Gates Balkans	0,000	25	8402
Krasnoyarsk MLBA	0,444	9	3472
Latvia Hunter-Gatherer	0,000	10	6807
LBK	0,136	22	6627
Sappali Tepe	0,167	12	3637
Scotland Bronze Age	0,300	10	3219
Sintashta	0,375	16	3649
Steppe	0,167	18	4145
Ukraine Neolithic	0,071	14	6807

En los más de cinco mil años de historia que abarcan las muestras, el rango de frecuencias de los haplotipos H2 varía desde valores próximos a 0,5 hasta valores de 0.

Las frecuencias más altas se encuentran en Krasnoyarsk, un yacimiento ruso de finales de la Edad de Bronce (0,444) y en los yacimientos de Inglaterra (*England Middle-Late Bronze Age*, 0,412), también de la segunda mitad de la Edad de Bronce, mientras que no se detecta presencia de la configuración H2 en la cultura de cerámica cordada de Europa central (*Corded Ware Central Europe*), la cultura del vaso campaniforme del Sur de Europa (*Beaker Iberia-France-Italy*) y los cazadores-recolectores de Letonia (*Latvia Hunter-Gatherer*) y de los Balcanes (*Iron Gates Balkans*).

#### 4.2. AMOVA.

Para estudiar cómo se estructura en el espacio y en el tiempo la heterogeneidad genética observada, se han realizado dos análisis moleculares de la varianza (AMOVA). Se han empleado todas las poblaciones de la **tabla R17**, agrupándolas según su posición geográfica y su edad promedio. Los resultados se muestran en la **tabla R18**.

En relación a la heterogeneidad de las regiones geográficas en que se agruparon las 33 poblaciones, el porcentaje de la varianza entre los seis grupos propuestos ( $F_{CT}$ ) ha sido del 2,72%, el porcentaje de varianza dentro de las poblaciones ( $F_{ST}$ ) del 90,09% y la varianza de las poblaciones dentro de los grupos propuestos ( $F_{SC}$ ) del 7,19%.

En cuanto a la variación temporal, se establecieron en 3 grupos los más de 5.000 años del rango temporal considerado. La varianza entre los citados grupos ( $F_{CT}$ ) fue de 2,29%, mientras que el porcentaje obtenido para la varianza dentro de las poblaciones ( $F_{ST}$ ) fue del 89,83%. La variación genética entre las poblaciones dentro de los grupos propuestos ( $F_{SC}$ ) fue del 7,88%.

**Tabla R18.** Análisis molecular de la varianza. Porcentaje de variación y significación (\* $p < 0,05$ , \*\* $p < 0,01$ ).

criterio	$F_{ST}$	$F_{SC}$	$F_{CT}$
Posición geográfica	90,09%**	7,19%**	2,72%*
Edad promedio	89,83%**	7,88%**	2,29%*

$F_{ST}$ : Variación genética entre individuos dentro de poblaciones.  $F_{SC}$ : Variación genética entre poblaciones dentro de grupos.  $F_{CT}$ : Variación genética entre grupos.

En ambos análisis, tanto entre zonas geográficas como entre grupos temporales, se ha detectado por tanto una heterogeneidad significativa.

### 4.3. Correlaciones parciales.

Dada la significancia hallada en los AMOVA, se ha realizado un análisis exploratorio con el objetivo de determinar la causa de las diferencias detectadas. Para ello se han estudiado las correlaciones parciales de varios pares de variables, suprimiendo la influencia del resto de factores considerados.

Las variables y los resultados se detallan en la **tabla R19**, donde queda patente una correlación negativa entre la frecuencia de H2 y la edad promedio de las poblaciones ( $p < 0,05$ ). Es decir, a mayor edad promedio de la muestra, menor es la frecuencia de H2 presentada por la misma. Al eliminar la influencia de otras variables, en cambio, los valores obtenidos para frecuencia-posición geográfica o frecuencia-cultura no muestran significación.

Estos resultados indican, en definitiva, que existen variaciones temporales estadísticamente significativas en las frecuencias de *MAPT\*H2* en el continente europeo.

**Tabla R19.** Resultados de la correlación parcial entre variables. En el triángulo inferior de la matriz se muestran los valores de la correlación. En el superior se muestran los valores de la significación. \* $p < 0,05$ .

	Frecuencia H2	Promedio edad	Posición geográfica	Cultura	Latitud	Longitud
Frecuencia H2		0,035*	0,505	0,188	0,705	0,199
Promedio edad	-0,392		0,763	0,064	0,986	0,620
Posición geográfica	-0,129	0,058		0,237	0,000	0,000
Cultura	-0,252	-0,349	-0,227		0,607	0,256
Latitud	0,074	0,003	-0,769	-0,100		0,093
Longitud	0,246	-0,096	0,617	0,218	0,317	

#### 4.4. Variabilidad geográfica y temporal. Mapas sintéticos.

Para facilitar la interpretación de las variaciones temporales del alelo H2 en Europa, se han realizado los mapas sintéticos que se muestran en las figuras R56, R57 y R58. En ellos se han representado las frecuencias de MAPT\*H2 durante los periodos 8500-6000 BP, 6000-4000 BP y 4000-2000 BP.

Durante el primer periodo (Fig. R56), los valores más altos de H2 se observan en los yacimientos neolíticos de los Balcanes (0,333) mientras que entre los cazadores-recolectores de los Balcanes y Letonia no se halla presencia de la configuración invertida.

En el segundo (Fig. R57), las frecuencias más altas de H2 se encuentran en las Islas Británicas (0,360), aunque cabe también destacar el grupo de yacimientos de la Edad de Hierro del centro de Asia, en concreto de Paquistán, bajo el nombre *Asia Iron Age* y con un valor muy cercano a 0,3 (0,294). En el yacimiento calcolítico de la Península Ibérica, en cambio, se observa la frecuencia más baja de H2 (0,067) y entre los miembros de la cultura de la cerámica cordada de Centroeuropa (*Corded Ware Central Europe*) esta configuración ni siquiera aparece.

Finalmente, en el tercer periodo (Fig. R58), los valores más altos de H2 se observan en los yacimientos de la Edad de Bronce del centro de Asia *Krasnovarks MLBA* (0,444), de nuevo en las Islas Británicas (0,412) y en los yacimientos entre el Mar Negro y Mar Caspio (Sintashta, 0,375). En cambio, en los yacimientos de la cultura campaniforme de la Península Ibérica, Francia e Italia la frecuencia de H2 observada es 0.

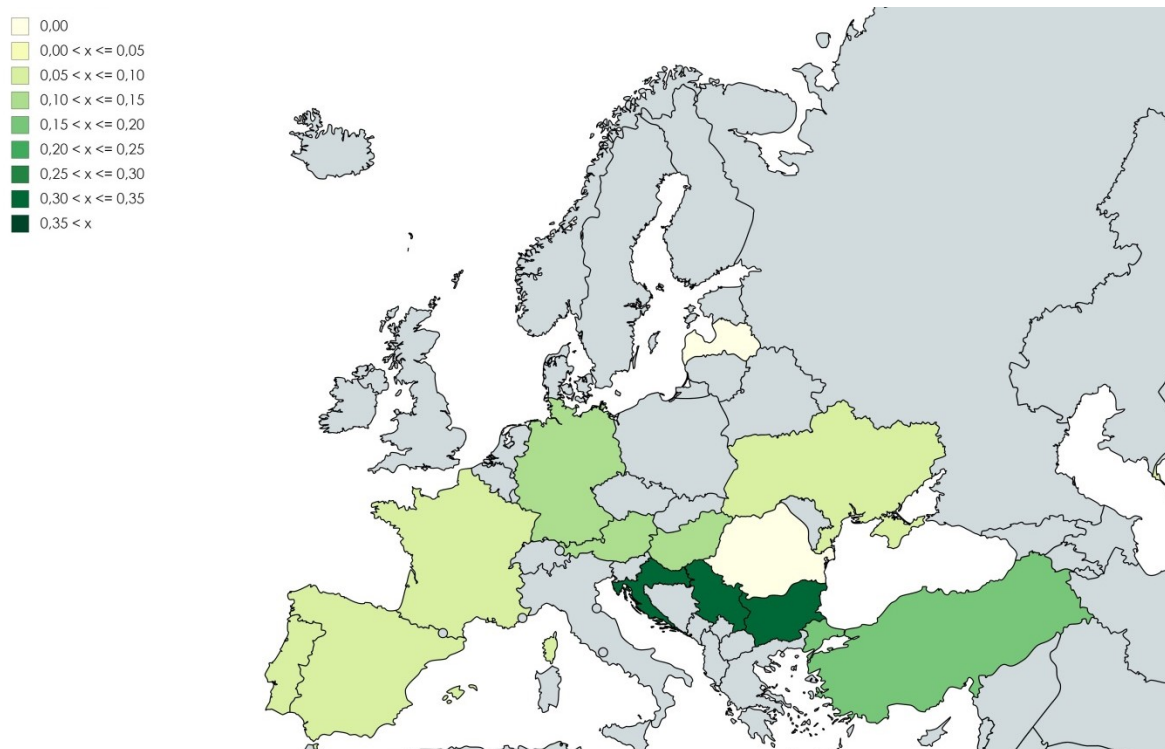


Figura R56. Representación de las frecuencias de MAPT H2 en Europa durante 8500-6000 BP.

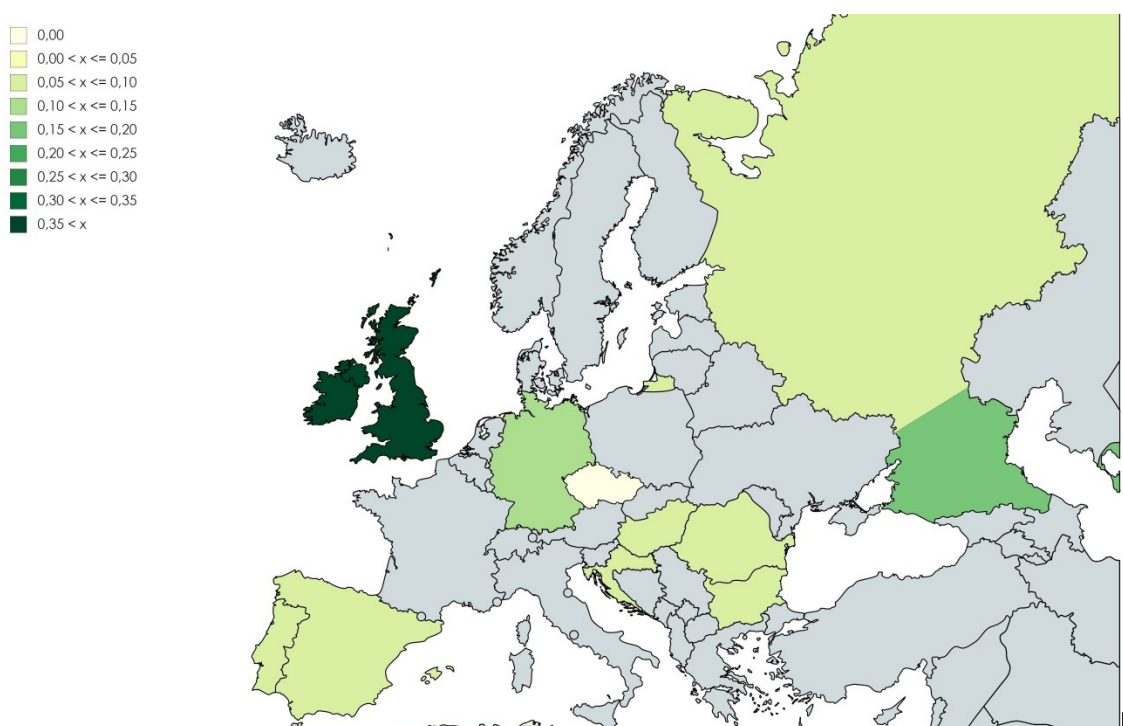


Figura R57. Representación de las frecuencias de MAPT H2 en Europa durante 6000-4000 BP.

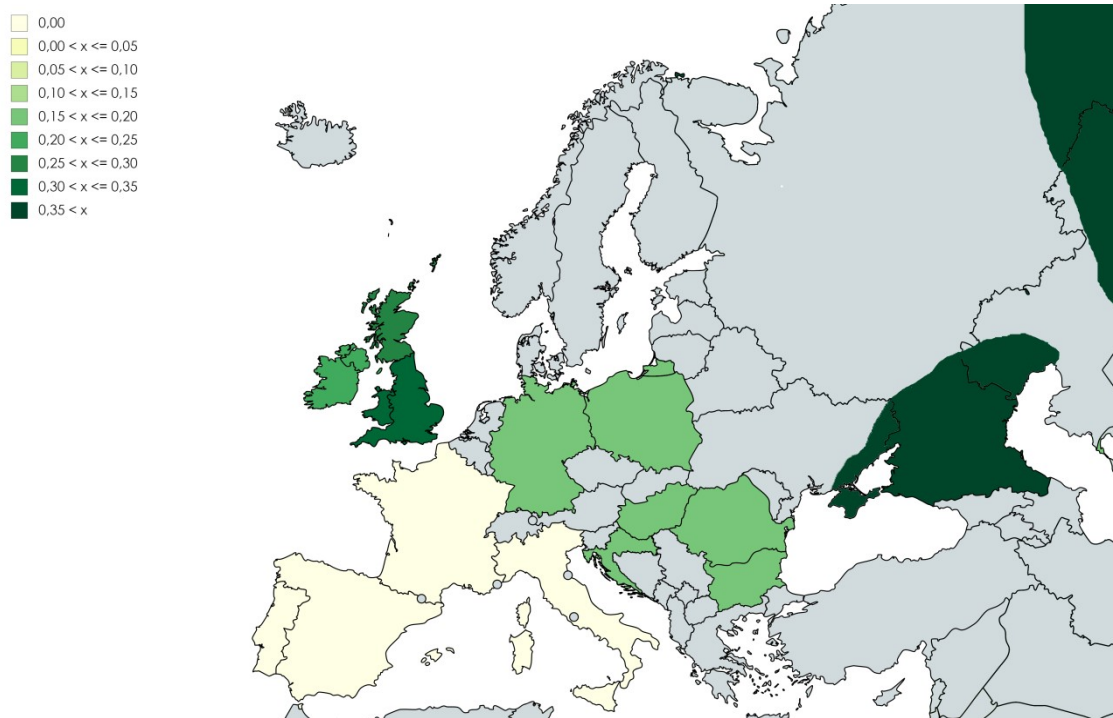


Figura R58. Representación de las frecuencias de MAPT H2 en Europa durante 4000-2000 BP.







# DISCUSIÓN

---



## DATACIÓN DEL ORIGEN DE LOS HAPLOTIPOS MAPT

La divergencia de los haplotipos H1 y H2 se ha fechado en hace aproximadamente tres millones de años por Stefansson et al, (2005). La antigüedad estimada por este mismo equipo para las variantes del linaje H1 (H1'-H1D) en función de las duplicaciones parciales del gen NSF es de aproximadamente 250.000 a 300.000 años, en consonancia con las edades de coalescencia de varios marcadores autosómicos en humanos (Excoffier, 2002) y semejantes a las obtenidas por Steinberg et al, (2012). Empleando otras regiones duplicadas, para este grupo el tiempo de coalescencia para H1 y H2 sería de 2,3 millones de años y para los subhaplotipos H1' y H1D de entre 224.000 años y 276.000 años. Este mismo grupo ha calculado la edad de coalescencia para los haplotipos H2' y H2D en poblaciones africanas en 117.000 – 155.000 años; finalmente, los haplotipos invertidos (H2) europeos y los africanos estarían separados por 37.000 – 59.000 años (Steinberg et al, 2012).

Donnelly y colaboradores (2010) plantean dos edades de coalescencia diferentes en función del haplotipo que sea considerado original. En el caso de ser el H1, como plantea su equipo, la fecha de coalescencia entre los cromosomas H1 y H2 sería de hace entre 16.400 y 32.800 años, mientras que en caso de que H2 fuera el cromosoma original del que derivó H1, esta fecha se retrasaría hasta hace entre 54.200 y 108.400 años. Las fechas presentadas en este trabajo son mucho más recientes que las aportadas por Stefansson et al, (2005) y Steinberg et al, (2012).

Considerando la hipótesis de Donnelly y colaboradores (2010) en la que se presenta el haplotipo H1 como original, las fechas de coalescencia no encajarían con una doble recombinación ocurrida entre ambos haplotipos principales previa a la expansión de los humanos anatómicamente modernos fuera de África descrita por Steinberg y colaboradores (2012), que parece explicar la región prácticamente idéntica que existe en el extremo 5' del gen CRHR1 en los haplotipos H1 y H2. Si tomamos las fechas de Donnelly y colaboradores (2010) como correctas, verosimilmente el haplotipo H2 tuvo que ser el original, pues la fecha de coalescencia propuesta en este escenario encaja con el resto de eventos detectados por otros grupos.

Cabe preguntarse por qué las fechas de Donnelly son tan distintas de las proporcionadas por otros equipos. Se debe tomar en consideración que los tiempos de coalescencia publicados se han obtenido aplicando distintas aproximaciones y que en general son una media obtenida a partir de los análisis en todo el intervalo. Adicionalmente, la caracterización detallada de los subhaplotipos de MAPT es

relativamente reciente (Steinberg et al, 2012; Boettger et al, 2012); dado que los haplotipos H1' y H1D y los H2' y H2D pueden recombinarse libremente, los segmentos analizados podrían representar secuencias formadas a partir de múltiples ancestros comunes. Puede pensarse también en la existencia de un evento de selección o de barrido reciente de los cromosomas de determinada configuración. Y finalmente, la mayoría de los trabajos en los que se caracterizan estos haplotipos emplean el indel rs72377086 y los SNPs rs10514879 y rs199451, que son marcadores indirectos del estado de la inversión de la región 17q21.31, no teniendo necesariamente que coincidir sus edades de coalescencia con las de los haplotipos que caracterizan.

Los datos recogidos hasta ahora en varias especies de primates, incluidos restos fósiles de individuos pertenecientes a distintas especies del género *Homo*, dan idea de una historia evolutiva muy compleja del locus 17q21.31, especialmente en el linaje que desembocó en humanos (Zody et al, 2008).

Muchos autores defienden que la orientación original de los haplotipos MAPT sería la H2 (Zody et al, 2008; Boettger et al, 2012; Steinberg et al, 2012). Otros aceptan que ésta era la configuración original en primates no humanos pero se inclinan por pensar que en alguna de las especies del género *Homo* previas a la aparición del *Homo sapiens* aparecieron los cromosomas H1, que se habrían fijado y que serían la orientación original en *Homo sapiens*, especie en la que volvería a darse una inversión que diera lugar de nuevo a un cromosoma H2 (Donnelly et al, 2010). Zody y colaboradores (2008), para quienes el haplotipo H2 es más antiguo tanto en primates no humanos como en humanos, consideran que tanto H1 como H2 derivarían de un haplotipo ancestral y Hardy y colaboradores (2005), que proponen esta misma idea, sugieren incluso algunos de los pasos que se dieron en la transformación de ese haplotipo ancestral similar al de los chimpancés actuales en el haplotipo humano H1 tras la mutación del gen de la Saitohina y el H2 tras la delección de más de 200 pares de bases en el intrón 9 del gen MAPT.

El número de primates no humanos analizados hasta el momento para esta región es escaso, no más de 15 en cada estudio, cuyo número también es escaso, tres hasta la fecha: Holzer et al, 2004; Zody et al, 2008 y Donnelly et al, 2010. Salvo el último, que considera sus resultados no concluyentes, los equipos de Holzer y Zody propusieron que el estado ancestral de la región es más similar al haplotipo H2 de humanos y que ha estado sujeto a inversiones recurrentes durante los últimos doce millones de años de evolución primate. En todas las especies la región 17q21.31 ha resultado ser polimórfica y en alguna la configuración equivalente al haplotipo H2 de

*Homo sapiens* ha resultado ser mayoritaria, al contrario que en humanos (Holzer et al, 2004; Zody et al, 2008).

Considerando la distancia temporal propuesta por Stefansson et al, (2005) entre el evento de duplicación en el linaje H2 y el antepasado más reciente de todos los haplotipos invertidos parecería que alguna presión evolutiva habría provocado una reducción de los haplotipos H2 que resultó en la reciente coalescencia de todos los haplotipos invertidos existentes hoy en día. La menor tasa de diversidad dentro de los propios subhaplotipos H2 al compararla con la de los subhaplotipos H1 (Steinberg et al, 2012; Boettger et al, 2012; Alves et al, 2015) apoyaría esta hipótesis.

## LUGAR DE ORIGEN DE LOS HAPLOTIPOS MAPT

El lugar en el que se produjo la inversión del haplotipo original también está en cuestión. La mayoría de los autores (Donnelly et al, 2010; Boettger et al, 2012; Steinberg et al, 2012) apuntan a África mientras que Donnelly et al, (2010) ofrecen además una alternativa en el suroeste de Asia.

Teniendo en cuenta las edades de origen y de coalescencia presentadas para las duplicaciones  $\beta$ -CNP205 (H1) y  $\alpha$ -CNP155 (H2), así, como de los haplotipos principales H1 y H2, parece claro que los haplotipos sin duplicaciones son los originales (como es lógico, ya que la estructura original generalmente es la más sencilla) y que ambas duplicaciones se generaron de manera independiente en ambos linajes. Además, dado que las duplicaciones son mayoritarias en Europa, cabe la posibilidad de que surgieran fuera de África.

Aceptando las dataciones más antiguas, si hubiesen ocurrido fuera de África, las duplicaciones del linaje H1 habría sido en *Homo heidelbergensis*, *Homo rhodesiensis* o incluso *Homo neanderthalensis* hace unos 250.000 años y las duplicaciones del linaje H2 habrían aparecido en *Homo erectus* o alguno de sus grupos descendientes. Teniendo en cuenta que la aparición de las duplicaciones fue un evento único, es decir, que sucedió una sola vez durante la historia evolutiva de la región 17q21.31, implicaría que la presencia en *Homo sapiens* de los cromosomas con duplicaciones se debería a la introgresión de los mismos desde otras especies del género *Homo*.

Si hubiesen ocurrido en África, podrían haberse perdido parcialmente como resultado de la deriva genética o de la selección; esto justificaría que la mayor diversidad de subhaplotipos H2 se encuentre en este continente, a pesar de su baja frecuencia (Alves et al, 2015). Esta idea se ve apoyada por la edad de los haplotipos H2 de los San, mayor que la de los haplotipos H2D africanos y los H2 europeos (Steinberg et al, 2012).

En consecuencia, parece que la secuencia de hechos más sencilla se da en el escenario en el que ambas duplicaciones ocurren en África y aumentan sus frecuencias en Eurasia por presión selectiva o, más probablemente, por acción de la deriva genética. Los grupos de *Homo sapiens* que se encontraban en África hace unos 70.000 años, portarían una representación de toda la diversidad que encontramos actualmente en África y en el resto del mundo y las frecuencias actuales observadas en todos los continentes podrían explicarse como resultado de las distintas migraciones y subsecuentes efectos fundadores.

Esta secuencia de hechos también sería coherente con la fecha de coalescencia propuesta por Steinberg y colaboradores para los haplotipos H2 africanos y los H2 europeos: hace entre 37.000 y 59.000 años.

Como se ha mencionado anteriormente, en los haplotipos MAPT actuales de ambos linajes existe la huella de una recombinación doble evidenciada por una región en el extremo 5' del gen CRHR1 prácticamente idéntica. Dado que esta es una característica común en todas las poblaciones analizadas por el equipo de Steinberg (2012) y que aparece en todos los grupos continentales, se ha de tratar de un evento ocurrido antes de la dispersión de los humanos anatómicamente modernos fuera de África.

### SELECCIÓN NATURAL EN LA REGIÓN MAPT

Uno de los factores que se ha valorado para explicar la distribución de frecuencias de los haplotipos MAPT en el sudoeste de Eurasia ha sido la acción de la selección natural, tanto fomentando el incremento del linaje H2 y el de los haplotipos con duplicaciones H1D y H2D como la disminución de la familia H1.

Como se ha mencionado en la introducción, Stefansson y colaboradores (2005) emplearon las tasas de reproducción en islandesas para tratar de demostrar que el

linaje H2 está siendo favorecido por la selección natural en Europa. En su estudio exponían que las mujeres con una copia de H2 tienen más hijos que aquellas que no la tienen, siendo la diferencia de menos de un hijo por mujer, y proponían un modelo de selección balanceadora en el que se favorecerían los heterocigotos.

Por otro lado, y como apuntan Donnelly y colaboradores (2010), la presencia del haplotipo H2 es más alta en otras regiones de Europa, como en la Península Ibérica, Grecia e Italia o incluso en algunas poblaciones de Próximo Oriente. Entre ellas pueden encontrarse ahora además las muestras de vascos y valencianos analizadas en este proyecto de investigación. Sin embargo, la recomendación de analizar otras poblaciones para constatar o no las conclusiones de Steffanson y colaboradores (2005) (Donnelly et al, 2010) no parece haber generado más estudios sobre el tema.

Se han recogido en todo caso observaciones que apuntan a que la presencia del haplotipo invertido H2 pueda estar relacionado con el aumento de la tasa de recombinación meiótica en mujeres (Fledel-Alon et al, 2011) y a su vez se acepta que una mayor tasa de recombinación en mujeres parece estar unida a una descendencia más numerosa (Kong et al, 2004; Coop et al, 2008).

En cualquier caso, Stefansson y colaboradores (2005) concluyen que el mecanismo mediante el que H2 afecta al número de hijos por mujer probablemente no sea la recombinación, es decir que el haplotipo invertido no tiene una implicación directa en la tasa de recombinación cromosómica. Por otra parte, en principio, la recombinación entre los linajes H1 y H2 no es posible y las mujeres en las que se da supuestamente una mayor tasa de descendencia son las heterocigotas.

Otras posibles evidencias de que el haplotipo H2 tiene ciertas ventajas selectivas proceden del estudio de Steinberg y colaboradores (2012). Al analizar 887 SNPs informativos comunes a H1 y H2 hallaron 4 que podrían alterar la estructura proteica del producto de uno de los genes que se ubica en esta región, el IMP5 (*Intermembrane Protease 5*). De los alelos presentes en H1, dos de ellos parecen relacionados con la enfermedad de Parkinson (Simon-Sanchez et al, 2009) y proponen que los alelos derivados que aparecen en H2 podrían ser sustituciones sujetas a selección positiva. Curiosamente los alelos derivados para los 4 SNPs aparecerían en H2 mientras que los originales se mantendrían como en chimpancés. Así pues, el haplotipo H2 habría cambiado su función mientras que el H1 habría mantenido su estructura más similar a chimpancés, apoyando la hipótesis de que la selección

natural habría actuado sobre H2. Boettger y colaboradores (2012) también consideran que la frecuencia de H2D en Europa y la baja diversidad relativa de este haplotipo detectada en varios estudios, incluyendo el suyo, puede ser indicador de una posible ventaja selectiva. Otros investigadores (Alves et al, 2015) han contestado contradiciendo a esta aparente falta de diversidad en el linaje H2, argumentando que hasta el momento de su trabajo la mayoría de las estimas habían sido realizadas a partir de polimorfismos comunes seleccionados de manera no aleatoria (Steffansson et al, 2005; Antonacci et al, 2009; Donnelly et al, 2010) o a partir de datos resecuenciados de un pequeño número de individuos (Steinberg et al, 2012), un sesgo que podría haber disminuido la diversidad general presente en el linaje H2.

Se ha sugerido también que la selección habría actuado favoreciendo en Europa los haplotipos con duplicaciones, tanto del linaje H1 como del H2. Al tener en cuenta los haplotipos duplicados frente a aquellos sin duplicaciones, la presencia de duplicaciones resulta mucho más discriminante que la inversión (Boettger et al, 2012). Como se ha evidenciado en la introducción, la mayor prevalencia de haplotipos duplicados se da en Europa (Steinberg et al, 2012; Boettger et al, 2012).

Según el estudio de Boettger y colaboradores, las duplicaciones del gen KANSL1 en los haplotipos H1D y H2D han alcanzado las altas frecuencias actuales en Europa de manera reciente, lo que indicaría un rápido incremento de estas frecuencias en su opinión debido a la selección natural actuando sobre estas duplicaciones en vez de sobre la inversión completa. Boettger y colaboradores (2012) sostienen que ambas duplicaciones podrían generar fenotipos similares que tuvieran cierta ventaja selectiva. De hecho, tanto la duplicación corta (alfa, correspondiente a los haplotipos invertidos) como la larga (beta, correspondiente a los haplotipos directos) son copias parciales de distintas longitudes (aproximadamente 155 kb y 205 kb respectivamente) del mismo gen, el KANSL1, en concreto de los exones ubicados en el extremo 5'.

Se ha especulado sobre si las duplicaciones podrían proteger contra las tauopatías y tener por tanto una ventaja selectiva frente a los haplotipos sin duplicaciones. Todos los haplotipos descritos como de riesgo para padecer estas enfermedades neurodegenerativas forman parte del linaje H1. Además, tanto en investigaciones relativas a la enfermedad de Alzheimer y la Parálisis Supranuclear Progresiva (Pittman et al, 2004; Hoglinger et al, 2011) como a la enfermedad de Parkinson (Skipper et al, 2004; Hoglinger et al, 2011; Setó-Salvia et al, 2011), las duplicaciones no aparecen o están infra-representadas.



Desafortunadamente, los individuos que portan los haplotipos considerados de riesgo para cada una de estas enfermedades suponen una fracción muy pequeña de la población y conforman una muestra de tamaño insuficiente para establecer una asociación significativa (Steinberg et al, 2012). A esta dificultad se añade que aún no se han identificado marcadores para diferenciar entre los haplotipos H1 con y sin duplicaciones y el bajo número disponible de individuos H2 hace que estas sugerencias continúen siendo meramente especulativas (Alves et al, 2015).

Tanto la selección positiva de los haplotipos H2 como de los haplotipos con duplicaciones parece contradictoria, ya que los subhaplotipos H2D son los únicos en los que puede darse el síndrome de la microdelección de la región 17q21.31, también conocido como síndrome Koolen-de Vries (Koolen et al, 2016), mediante recombinación alélica no homóloga facilitada por las repeticiones genómicas de bajo número de copias (Low Copy Repeats, LCRs) presentes en los mismos en orientación directa y que flanquean la región crítica para esta enfermedad (Sharp et al, 2006; Shaw-Smith et al, 2006; Koolen et al, 2006; Mehan et al, 2004; Lupski, 2007; Rao et al, 2010; Dubourg et al, 2011). Las poblaciones europeas, que son las que muestran mayor frecuencia de los haplotipos invertidos con duplicaciones, serían por tanto poblaciones de alto riesgo; de hecho un 97% de los casos del síndrome de la microdelección ocurre en europeos o descendientes de europeos (Cooper et al, 2011). Diferentes estudios estiman la prevalencia de esta enfermedad entre 1/16.000 (Koolen et al, 2008) y 1/55.000 (Vulto-van Silfhout et al, 2013; Coe et al, 2014) aunque no puede ser establecida con precisión debido al número limitado de casos identificados hasta ahora (Koolen et al, 2016).

Este hecho contrasta claramente con las frecuencias que se observan en Europa de H2D ya que debería estar sujeto a una selección purificadora que impidiera que hubieran llegado a las altas frecuencias que muestra en esta región.

Siguiendo este argumento, algo similar sucede con el haplotipo H1, relacionado con la susceptibilidad a padecer las enfermedades conocidas como tauopatías. Para que la selección natural pudiera actuar tendría que afectar a los individuos portadores antes de la edad de reproducción, pero en la mayoría de los casos las enfermedades neurodegenerativas se manifiestan pasada esta edad, aunque éste podría precisamente ser el motivo por el que unos haplotipos supuestamente perjudiciales pudieran haber alcanzado la distribución global que presentan y la prevalencia con la que aparecen en la mayoría de las poblaciones.

Los desórdenes causados por el mal funcionamiento de MAPT son altamente dependientes de los niveles de expresión de la proteína, por lo que el foco de interés se sitúa en la regulación epigenética de la expresión del gen MAPT (Koolen et al, 2012), el *splicing* alternativo de los exones y las modificaciones post transcripcionales de la proteína (Sündermann et al, 2016). Se han descrito diferencias en la expresión y en el *splicing* alternativo de MAPT entre los haplotipos H1 y H2 que afectan a las funciones celulares a diferentes niveles, pudiendo aumentar la susceptibilidad a determinadas enfermedades neurodegenerativas (Pascale et al, 2016).

Se han hallado desviaciones de lo esperado según el modelo de neutralidad en ambos haplotipos (Alves et al, 2015) pero dado que ocurren en ambas configuraciones Alves y colaboradores consideran que es indicativo de que la demografía ha tenido una mayor influencia que la selección natural en el modelado de la distribución de frecuencias MAPT europea.

Aunque con estos datos no puede desecharse definitivamente la acción de la selección natural sobre las frecuencias de los haplotipos MAPT en Europa, parece mucho más sencillo recurrir a un evento demográfico como la recolonización post glacial o la expansión del Neolítico desde el suroeste de Asia dentro de Europa como proponen Donnelly y colaboradores (2010) que aludir a la selección natural. Zody (2008), Steinberg (2012), y Alves (2015) también se decantan por que la llamativa distribución de frecuencias de los haplotipos MAPT sea resultado de una historia demográfica compleja.

### INTROGRESIÓN DESDE OTRAS ESPECIES DEL GÉNERO *HOMO*

¿Sería posible que el haplotipo minoritario H2 fuera un ejemplo de “herencia horizontal” entre distintas especies del género *Homo*?

Dada la distribución principalmente euroasiática de los haplotipos H2, una de las primeras propuestas sobre su origen fue la introgresión de los cromosomas invertidos desde *Homo neanderthalensis* (Hardy et al, 2005). Además de en la distribución, esta propuesta se apoyaba en el análisis del *slippage* de dinucleótidos llevados a cabo en investigaciones anteriores (Goldstein et al, 1995; Weber et al, 1993), que databan la aparición de los haplotipos H2 en *Homo sapiens* en hace aproximadamente entre 10.000 y 30.000 años. Dado que ambas especies de homínidos coexistieron en regiones de Eurasia hace entre 18.000 y 45.000 años (Stringer et al,

2001), Hardy y colaboradores propusieron, cuando aún no se conocía evidencia directa que apuntara a la hibridación entre ambas especies, descrita por primera vez tras la secuenciación del genoma neandertal (Green et al, 2010), que el haplotipo invertido habría surgido en Neandertales y que habría entrado a formar parte del genoma del *Homo sapiens* a través del mestizaje entre ambos grupos.

La secuenciación del genoma neandertal a partir de los restos de tres individuos (Green et al, 2010) ha llevado a algunos autores a afirmar que los Neandertales portaban el haplotipo H1 (Setó-Salvia et al, 2012). Esto no excluye que pudieran portar también el H2, aunque todavía falten evidencias directas de ello. Probablemente en Neandertales la región 17q21.31 fuera también polimórfica - dados los datos de coalescencia y origen tanto de los haplotipos principales como de las duplicaciones - y si presentaban frecuencias similares a las de *Homo sapiens*, la probabilidad de encontrar un homocigoto para H1 es más alta. Esta misma reflexión conduce de todos modos a que, si H2 era también menos frecuente en Neandertales, parece poco probable que *Homo sapiens* pudiera haberlo heredado de los primeros.

Adicionalmente, algunos autores dudan de que las coincidencias que se dan entre los genomas de humanos actuales y las de las especies de homínidos arcaicos - como Neandertales y Denisovanos - se deban a introgresiones desde éstos últimos a los primeros y consideran más plausible que estas afinidades sean consecuencia del mantenimiento de polimorfismos ancestrales comunes (Lowery et al, 2013).

En todo caso, como se ha mencionado, la edad de coalescencia para ambos linajes podría ser de unos tres millones de años (Steffanson et al, 2005), habiéndose propuesto que el haplotipo H2 se habría originado fuera de África y entrado de nuevo antes de la migración de los humanos anatómicamente modernos fuera de este continente hace unos 70.000 años. Esta hipótesis implicaría mestizaje entre humanos y otras especies de homínidos contemporáneos como *Homo heidelbergensis* u *Homo erectus* (Steffanson et al, 2005) en el continente africano.

En caso de que el mestizaje hubiera ocurrido, deberían encontrarse otras regiones con tiempos de coalescencia similares, especialmente en aquellas áreas del genoma con baja tasa de recombinación, tal y como predice la hipótesis del mestizaje (Steffanson et al, 2005). Como se ha expuesto previamente, una alternativa a la introgresión como explicación de la existencia de las afinidades genéticas observadas entre homínidos arcaicos como neandertales o denisovanos y humanos

contemporáneos es la existencia de polimorfismos comunes ancestrales que se mantuvieron en múltiples linajes.

## EFFECTOS FUNDADORES Y FLUJOS GÉNICOS EN EUROPA

Durante el desarrollo de este proyecto una de las cuestiones que ha surgido y se ha repetido más insistentemente es el motivo de la particular estructuración de las frecuencias de los haplotipos MAPT, cuyos primeros indicios se detectaron hace veinte años (Baker et al, 1999) y revelaron la estratificación de los haplotipos principales, corroborada en trabajos posteriores (Hardy et al, 2005; Stefansson et al, 2005; Donnelly et al, 2010; Steinberg et al, 2012, por citar algunos), entre los que se incluye este proyecto de tesis.

Como ha quedado descrito anteriormente, el haplotipo H2 se considera principalmente europeo por ser en este continente donde aparece de manera más regular y con frecuencias más altas, aunque moderadas, salvo una llamativa excepción (vascos del Goierri de Gipuzkoa). No obstante, la mayor diversidad de estos haplotipos se encuentra en África, por lo que se sospecha que su origen, como el de prácticamente toda la variabilidad genética humana, se ubica en aquel continente.

Tras el análisis detallado de los datos obtenidos en este trabajo, se ha encontrado una correlación entre las coordenadas geográficas y las frecuencias de los haplotipos MAPT en Europa, que forman un gradiente que se corresponde con dos clinas altamente significativas y prácticamente complementarias para los haplotipos H1 y H2D.

El análisis de la distribución de frecuencias mediante el método del centroide ha revelado que existe un grupo de poblaciones con una heterocigosidad claramente mayor de la esperada. Entre estas poblaciones, algunas se corresponden con grupos tradicionalmente considerados con cierto grado de aislamiento, como vascos y sardos (Cavalli-Sforza, 1994), por lo que no se puede justificar esta heterogeneidad aludiendo a una mayor intensidad de flujo génico. Atendiendo a la historia de estas poblaciones, más bien podría explicarse este fenómeno por un origen diferenciado de parte de su patrimonio genético.

Resulta llamativo que las poblaciones con frecuencias más altas de H2D están ubicadas en zonas que tuvieron un importante papel como refugios glaciales, las

penínsulas del sur de Europa (Iberia, Italia y los Balcanes) (Hewitt 1996, 2000; Taberlet et al, 1998).

¿Pudo tener esto algo que ver con la distribución actual de las frecuencias MAPT en Europa como ya se ha detectado con otros marcadores como por ejemplo con polimorfismos del ADNmt o del cromosoma Y? (Torrioni et al, 1998; 2001; Cardoso, 2008; Wiik, 2008) Algunos autores apuntan a que la región 17q21.31 era polimórfica en los grupos que salieron por primera vez de África, aunque los haplotipos invertidos (H2) debían de estar presentes a muy bajas frecuencias (Stefansson et al, 2005; Donnelly et al, 2010; Steinberg et al, 2012). Las condiciones climáticas extremas que se dieron durante el Último Máximo Glacial hace unos 30.000 años obligaron a las poblaciones que habitaban Europa – y que *a priori* podrían ser polimórficas para los haplotipos MAPT - a migrar hacia el sur y asentarse en alguna de las pocas zonas habitables, en los denominados refugios climáticos, lo que provocó un replegamiento y concentración de esas poblaciones además de una reducción del tamaño de las mismas y un fuerte aislamiento entre ellas.

Uno de estos refugios climáticos, al que se ha otorgado una especial importancia en la evolución del patrimonio genético europeo y ha sido considerado como el principal asentamiento de los grupos humanos que llegaron desde el norte de Europa durante la última glaciación, se encontraba en la costa franco-cantábrica y habría abarcado desde el suroeste de Francia hasta la parte oriental de la cornisa cantábrica (Torrioni et al, 2006). Como consecuencia, se consideró la posibilidad de que, debido a un efecto fundador y la acción de la deriva genética, las frecuencias del haplotipo invertido, especialmente las de aquellos con duplicaciones, aumentarían cuantiosamente en las poblaciones de estos refugios hasta alcanzar frecuencias tan altas como las mostradas en este trabajo.

En general, las poblaciones de seres vivos que habitan temporalmente en refugios climáticos tienden a diferenciarse debido a la deriva y/o a la selección natural en respuesta a las condiciones específicas de los refugios (Stewart et al, 2012); se ha observado que las variaciones ecológicas entre diferentes refugios pueden aumentar las diferencias entre las poblaciones aisladas (Hewitt 1996, 2000; Stewart, 2009).

Hace unos 16.000 años la última glaciación comenzó a remitir y las temperaturas a templarse. La retirada de los hielos acompañada de la consecuente expansión de las especies vegetales volvió a hacer habitables zonas más al norte.

Probablemente siguiendo los desplazamientos de los animales que cazaban como componente fundamental de su alimentación, algunos de los grupos humanos que hasta entonces habían sobrevivido en los refugios emprendieron una nueva migración hacia el norte y el este de Europa en el proceso conocido como recolonización postglacial (Richards et al., 1996; Torroni et al., 1998; Richards, 2003; Achilli et al., 2004, 2005; Pereira et al., 2005). Durante la misma, la dispersión de estos humanos desde los refugios glaciales hacia el norte y el este podría haber dejado como evidencia genética el gradiente de frecuencias del haplotipo invertido que disminuye en dirección nordeste. A medida que las poblaciones se alejaban del refugio franco cantábrico sucesivos efectos fundador podrían haber dado lugar a una disminución de las frecuencias de H2.

Esta tendencia que muestran las frecuencias de los haplotipos MAPT ha sido observada anteriormente tanto en la distribución geográfica de algunos haplogrupos del ADNmt (Torroni et al., 1998; 2001; Achilli et al., 2004; Alfonso-Sánchez et al., 2008b), como en las clinas latitudinales que se han identificado en numerosos marcadores genéticos (Sokal et al., 1989b; Peña et al., 2002; 2004; García-Obregón et al., 2010). Estos patrones se consideran en general una huella en los genomas europeos actuales de los procesos de deriva genética promovidos por los desplazamientos migratorios de pequeños grupos humanos. Los efectos del proceso de repoblamiento continental, que hizo que las variaciones genéticas acumuladas por las poblaciones durante su periodo de asentamiento en los refugios se diseminaran por toda Europa, han sido descritos también para diferentes haplogrupos de ADNmt y del cromosoma Y (Torroni et al., 1998; 2001; Cardoso, 2008; Wiik, 2008).

La distribución de frecuencias actuales de MAPT en las regiones estudiadas marca claramente un máximo para H2D en la región de la costa vasca, un centro a partir del cual, en todas direcciones, las frecuencias de este haplotipo disminuyen paulatinamente mientras aumenta la presencia de los haplotipos de orientación directa.

El análisis de la diversidad nucleotídica ( $\pi$ ) llevada a cabo por Steinberg y colaboradores (2012) resultó en un árbol cuya topología, junto con la falta de diversidad relativa de los haplotipos MAPT en Europa, sugiere un evento reciente de cuello de botella seguido de una expansión poblacional o de un barrido selectivo, lo que podría concordar con el escenario de recolonización postglacial desde el refugio franco cantábrico planteado en los párrafos anteriores,

Sin embargo, los datos de yacimientos prehistóricos analizados no apoyan la hipótesis de la dispersión de las altas frecuencias de los haplotipos invertidos desde el refugio franco cantábrico. Si bien es cierto que es difícil recuperar ADN de individuos paleolíticos y que no existen demasiadas evidencias individuales y aún menos poblacionales, la base de datos estudiada para este proyecto muestra una frecuencia muy baja, prácticamente inexistente del haplotipo invertido en Europa durante el Paleolítico. En las poblaciones indentificadas como “*Iron Gate Balkans*” y “*Latvia Hunter Gatherer*” no se ha detectado la presencia de los haplotipos invertidos y entre los individuos no asignados a poblaciones, sólo se han encontrado tres portadores de la configuración H2 (Kostenky, MA1 y un cazador-recolector de centro Europa). H2 ha aparecido por tanto en menos del 5% de las muestras del Paleolítico; en consecuencia, se ha establecido para este alelo una frecuencia promedio inferior a 0,1 durante el Paleolítico en Europa.

En cambio, atendiendo a los datos del Neolítico, se puede establecer un valor promedio de H2 más similar al actual europeo, igual o superior a 0,15, con variaciones notables.

Los primeros neolíticos con frecuencias altas aparecen en los Balcanes (*Balkans Neolithic*, 0,33, edad promedio 6825 BP) y más tarde en Anatolia (*Anatolia Neolithic*, 0,167, edad promedio 6167 BP), pero lo más probable es que este orden temporal se deba exclusivamente a la edad de los yacimientos estudiados y que el primer aumento de las frecuencias de los haplotipos invertidos ocurriera en la región de Anatolia, donde, con el inicio de la agricultura, se dio un primer efecto fundador “*in situ*”. El gran crecimiento de la población local, facilitado por la nueva abundancia de alimento, a partir de un número reducido de individuos, favoreció la sobrerrepresentación, en este caso, de los haplotipos invertidos en la zona.

Con la expansión de la agricultura, estas poblaciones también difundieron una frecuencia de H2 más alta de lo que se había visto hasta ese momento en Eurasia. Así, hace 7000 años – y de acuerdo también con la evidencia arqueológica (Krauß, 2018) – en los Balcanes se habría ocasionado un nuevo efecto fundador, tras el cual, las frecuencias de H2 habrían alcanzado el valor 0,33 en dicha región.

Se considera que en el continente europeo la expansión de la agricultura siguió dos rutas principales; una a lo largo de la cuenca del Danubio hasta el norte de Europa y otra costeano el Mediterráneo por el norte, vía Italia y sur de Francia, hasta la Península Ibérica (Hofmanová, 2015).

Así, durante la primera época de la difusión de la agricultura en Europa (Fig. R56), los grupos neolíticos ibéricos presentan diferencias genéticas con los agricultores del centro y norte de Europa en cuanto a los haplotipos MAPT; la frecuencia de la configuración invertida en la Península Ibérica es inferior a 0,1 mientras que en el centro y sudeste de Europa se han hallado frecuencias más altas, en torno a 0,15 o superiores. Las diferencias detectadas podrían deberse a un cuello de botella que provocó una disminución de la diversidad en las primeras poblaciones neolíticas ibéricas (Valdosera, 2018) y un descenso de la presencia del haplotipo invertido, pero también a un mayor grado de mezcla entre las poblaciones residentes de cazadores-recolectores y de agricultores y por tanto a un mayor flujo génico entre las poblaciones ibéricas que entre los grupos con distinta forma de vida / explotación de recursos que habitaban otras zonas de Europa (González-Fortes, 2017; Olalde, 2018).

Hace entre 6000 y 4000 años (Fig. R57), las frecuencias de MAPT presentan una distribución prácticamente inversa a la actual: en la mitad sur del continente durante aquella época se observan unos valores de H2 inferiores a los detectados en la mitad norte.

La frecuencia más alta en esta época en las Islas Británicas (*Britain Neolithic*, 0,36), muy por encima de los valores actuales, es muy llamativa. Aunque no puede establecerse su origen con certeza al no disponer de las frecuencias MAPT del periodo anterior, cabe pensar que podría deberse a algún proceso de efecto fundador asociado al poblamiento del territorio: Previo a la expansión neolítica, las poblaciones de cazadores-recolectores que habitaban las islas presentarían frecuencias muy bajas o inexistentes de los haplotipos invertidos. Los primeros agricultores llegados de Próximo Oriente remontando el Danubio, portarían unas frecuencias altas de H2 debido al efecto fundador original. Éstos se asentaron en las islas y aumentaron su número poblacional, provocando otro efecto fundador que intensificó el primero.

Otro valor destacable en este periodo es el observado en las culturas de la Edad de Hierro en Asia, en concreto en Paquistán (*Asia Iron Age*, 0,294), que siguiendo la línea de este trabajo, posiblemente puedan justificarse aludiendo a la influencia del centro de expansión de la agricultura en Próximo Oriente.

En la última época prehistórica analizada, entre los años 4000 y 2000 antes del presente (Fig. R58), el patrón de distribución de frecuencias de los haplotipos MAPT sigue siendo el opuesto al actual: los valores más elevados de H2 se



encuentran en el norte de Europa mientras que en el sur los valores son mínimos o nulos.

En este periodo las frecuencias más altas se hallan en las muestras del centro de Asia (Krasnoyarsk, 0,444) y en las estepas situadas entre el mar Negro y el mar Caspio (Sintashta, 0,375). Es en ésta última zona donde se asentaron los kurganes, considerados descendientes de los primeros agricultores neolíticos que partieron de Oriente Medio, donde, como se ha expuesto anteriormente, se dio el primer efecto fundador que elevó las frecuencias de H2 por encima de lo habitual hasta ese momento.

Hack y colaboradores (2015) han hallado evidencias de reemplazos (turnover) posteriores a la gran expansión neolítica asociada a la agricultura, incluyendo migraciones desde la estepa del Ponto-Caspio que sustituyeron el 75% del patrimonio genético europeo.

Sus datos apoyan la hipótesis de los kurganes o de estepa, que propone que las praderas del norte de los mares Negro y Caspio habrían sido al menos la fuente de alguno de los idiomas indo europeos, como parece indicar la migración masiva ocurrida hace unos 4.500 años y asociada a las culturas Yamnaya y la Corded ware, ambas identificadas previamente por quienes propusieron la hipótesis de estepa como probables vectores de la expansión de las lenguas indo europeas en Europa (Gimbutas M, 1956; Anthony DW, 2007; Ref. 9 de HAAK 2015; Chang et al, 2015).

Aunque no tan altas como en el periodo anterior, en las Islas Británicas, entre los individuos agrupados bajo la denominación “*Beaker Britain*”, se mantienen frecuencias de H2 relativamente altas, coincidentes con la media actual europea, en torno a 0,2. Pareciera que las frecuencias de los haplotipos invertidos se hubieran diluido, muy probablemente por simple deriva genética.

Un poco más tarde durante este mismo periodo, unos 3000 años antes del presente, en los grupos “*Scotland Bronze Age*” e “*English Middle Late Bronze Age*” se detectan de nuevo frecuencias de H2 por encima de la media europea, 0,3 y 0,4 respectivamente.

La huella génica en las Islas Británicas de movimientos pre-Romanos pero post-mesolíticos desde la Europa continental a Inglaterra, ha sido también sugerida para otros marcadores (Leslie et al, 2015).

Actualmente las frecuencias en las Islas Británicas, aunque, como en la mayor parte de Europa, son altas en comparación con las de otros lugares del mundo, son más modestas, encontrándose en la población británica una frecuencia de H2 de 0,2 y siendo la excepción en el norte de Europa las Islas Orcadas, que prácticamente alcanzan una frecuencia de 0,3 (Donnelly et al, 2010), indicando quizá la herencia de pobladores previos.

Las Islas Británicas resultan pues un perfecto ejemplo de cómo han cambiado las frecuencias de H2 en Europa a lo largo de la historia, principalmente afectadas por el azar.

Aunque se puede pensar en un único grupo portador de altas frecuencias de H2 desde Anatolia a los yacimientos europeos y asiáticos mencionados, la explicación más probable a la situación descrita por los datos de los restos neolíticos, con una frecuencia base de H2 en torno al 0,15 y varios picos concretos, es una serie de efectos fundadores en lugares y momentos puntuales, siendo el primero de ellos el ocurrido en Anatolia y fechado por este grupo entre el inicio de la agricultura hace unos 10000 años y el momento en el que se detecta esa primera alta frecuencia de H2, hace unos 7000 años.

Por tanto, parece bastante posible la entrada del haplotipo H2 en Europa a elevadas frecuencias desde Próximo Oriente junto con la expansión de la agricultura. Sin embargo, la distribución actual de las frecuencias de los haplotipos MAPT debe ser resultado de fenómenos más recientes, ya que, entre los años 8000 y 2000 antes del presente, los valores de las configuraciones de MAPT han oscilado notablemente, siendo el patrón observado en el último periodo exactamente inverso al actual.

Los valores de las frecuencias del **haplotipo H2'**, en cambio, no revelan una pauta clara y, añadido a esto la escasez de datos, su interpretación resulta más compleja.

H2' parece ser el más antiguo de los subhaplotipos invertidos y aparece siempre a frecuencias muy bajas. Varias líneas de evidencias indican que esta estructura menos habitual podría ser el estado ancestral del haplotipo H2 e incluso de todos los haplotipos MAPT que existen en la actualidad. Por un lado, H2' es más semejante al haplotipo H1 que H2D. Por otro lado, H2' ha sido identificado en poblaciones cazadoras-recolectoras del centro de África (incluyendo dos tribus de Pigmeos Mbuti y una de Pigmeos Biaka). Estas poblaciones podrían haber transmitido

el subhaplotipo H2', originado en poblaciones antecesoras hace entre 2 y 3 millones de años. H2' provee de un enlace potencial que explicaría cómo la inversión pudo haber ocurrido por un simple evento de recombinación homóloga no alélica (Boettger et al, 2012).

En Europa, las frecuencias más altas de este haplotipo se encuentran en Cerdeña (0,127), vascofranceses (0,105) y Adygei (0,077). En cambio, los grupos vascos analizados en este trabajo presentan frecuencias de H2' notablemente menores (Navarra 0,027, Gipuzkoa 0,006). Navarra es mucho más similar a Valencia (0,026), aunque en la población de España tomada de la bibliografía no se encontró este subhaplotipo. Las frecuencias de H2' en Europa son siempre superadas por las de H2D. En Europa son mucho más comunes los haplotipos con duplicaciones, representando el 60% de los haplotipos detectados (Boettger et al, 2012).

## FRECUENCIAS MAPT EN OTROS CONTINENTES

### ÁFRICA.

En **África** el subhaplotipo H2' presenta unos valores relativamente altos. Los valores máximos están en el este, con la excepción de los Bakola e incluyen una peculiar mezcla de grupos: algunas poblaciones de pigmeos (Mbuti y Bakola) pero no todas las analizadas, algunas poblaciones khoisán del este – aunque tampoco todas – como los Sandawe (y no los Hadza, emparentados con éstos últimos), algunos grupos afroasiáticos, los Beja y los Masai, así como los nilosaharianos Datog y Sengwer. H2D aparece en África con los valores más altos también al este, con los Mandenka como única excepción. Tomando en cuenta ambos haplotipos, la frecuencia de H2 más habitual en África, cuando está presente, es inferior a 0,1 (Steinberg et al, 2012).

Esta distribución irregular – tanto en cuanto a la distribución geográfica de las frecuencias, como al patrón de subsistencia de las poblaciones analizadas y a la familia lingüística a la que pertenecen las mismas - dificulta la interpretación de los resultados. Este grupo propone tres posibles escenarios para explicar el origen del patrón azaroso que se observa actualmente: (1) desde la frecuencia “base” de H2 en todo el mundo (0,10) la deriva ha aumentado ligeramente las frecuencias de los haplotipos invertidos en algunos grupos mientras que en otros la ha disminuido incluso hasta fijar H1. (2) El intercambio génico que acompañó la expansión poblacional y difusión de la agricultura procedente de Oriente Medio o (3) un flujo

génico inverso más reciente con poblaciones que habitaban la Península Arábiga, provocaron el aumento de la frecuencias de los haplotipos invertidos en aquellas regiones donde la mezcla fue mayor, ya que los grupos inmigrantes portaban proporciones de H2 significativamente mayores.

En África se han identificado hasta tres puntos donde la agricultura se pudo desarrollar de manera independiente: Etiopía, el Sahel y el Oeste de África, mientras que se considera que en el Valle del Río Nilo, la agricultura llegó con la migración de poblaciones de Próximo Oriente, hace aproximadamente. Dado que, como propone esta tesis, el efecto fundador original que elevó las frecuencias de los haplotipos invertidos por encima de la media se dio entre hace 10000 y 7000 años en Próximo Oriente, las poblaciones que llevaron la agricultura a la cuenca del Nilo pudieron haber portado también frecuencias altas de H2, comparadas con aquellas que existían en la zona. El intercambio génico entre los grupos autóctonos y los primeros agricultores habría dejado una huella de frecuencias de H2 ligeramente más altas en el entorno del Río Nilo que en el resto del continente africano. Esta huella, por antigua y por no estar sujeto el marcador a selección natural y ser por tanto más susceptible al efecto de la deriva, habría quedado desdibujada, resultando en el patrón irregular que observamos actualmente, en el cual, en su mayoría, las frecuencias más notables de H2 aparecen en la zona prevista pero no en todas las poblaciones que la habitan.

Parece, en cambio, que el modelo que más se adecua al patrón irregular observado sería la última propuesta, un intercambio genético más reciente, posterior al comienzo de las expansiones de los Kurganes, con poblaciones que habitaban la península arábiga.

Recientemente, se ha publicado un estudio (Pickrell et al, 2014) en el que se describe una proporción variable de mestizaje en el este y el sur de África con poblaciones del oeste de Asia. Así, las poblaciones Bantú, en las que no existen los haplotipos invertidos, presentan tasas inexistentes o muy bajas de mestizaje, inferiores al 2,5%. Entre los Hadza, sin embargo, se observó un porcentaje de *reflujo* genético del 6,4% y en los Masai y en los Sandawe el porcentaje estimado de mestizaje con euroasiáticos fue del 19% y del 16%. Si bien no coinciden exactamente las poblaciones incluidas en este trabajo con las de Pickrell et al. (2014), en el continente africano sí **se observa una clara asociación entre las frecuencias de H2 (H2' y H2D) y la intensidad del flujo génico desde el sudoeste de Eurasia.**

**Tabla D.1:** Estima de la proporción de herencia euroasiática occidental en poblaciones del este y sur de África (Pickrell et al, 2014).



**Table 1. Estimates of the proportion of West Eurasian ancestry in southern and eastern African populations**

Population	Language classification	West Eurasian ancestry
<b>Southern Africa</b>		
Nama	Khoe-Kwadi	0.11
Shua	Khoe-Kwadi	0.11
Hai  om	Khoe-Kwadi	0.11
Khwe	Khoe-Kwadi	0.11
Tshwa	Khoe-Kwadi	0.11
Naro	Khoe-Kwadi	0.11
G ui	Khoe-Kwadi	0.11
Taa_North	Tuu	0.11
G  ana	Khoe-Kwadi	0.11
!Xuun	Kx'a	0.11
#Hoan	Kx'a	0.11
Damara	Khoe-Kwadi	0.11
Kgalagadi	Bantu	0.11
Ju 'hoan_North	Kx'a	0.11
Taa_East	Tuu	0.11
Mbukushu	Bantu	0.11
Taa_West	Tuu	0.11
Himba	Bantu	0.11
Ju 'hoan_South	Kx'a	0.11
Tswana	Bantu	0.11
Wambo	Bantu	0.11
<b>Eastern Africa</b>		
Tygray	Semitic	0.11
Amhara	Semitic	0.11
Afar	Cushitic	0.11
Oromo	Cushitic	0.11
Somali	Cushitic	0.11

Según Pickrell et al. (2014), los procesos de flujo génico descritos habrían tenido en todos los casos una antigüedad menor de 4.500 años.

En África hay dos zonas importantes de mezcla con caucasoides: en el norte, a través de la Península Arábiga y del Mediterráneo, y en el este, donde existen numerosas pruebas de intercambios prehistóricos con los árabes (Cavalli-Sforza, 1996), como el reino árabe-etíope surgido hace unos 3000 años que propició migraciones en ambas direcciones y cuya primera capital fue Saba, en el sur de Arabia, y que después se trasladó a Aksum, en el norte de Etiopía.

En este escenario cabe preguntarse si es posible que las poblaciones caucasoides que en torno al 2500 a.C. migraron a África y se mezclaron con las

comunidades locales presentaran los haplotipos H2 en proporción suficiente como para haber dejado la huella del patrón observable actualmente.

Las fechas propuestas por Pickrell y colaboradores (2014) son posteriores al comienzo de la expansión de los Kurganes, que sucedió hace entre 5000 y 6000 años (Haak et al, 2015).

Los Kurganes, comunidades asentadas en el noreste del Cáucaso y del mar Caspio, son uno de los grupos candidatos a haber contribuido a la expansión de los idiomas indoeuropeos - la llamada **hipótesis de estepa** y propuesta por Marija Gimbutas en 1956 - y se les considera descendientes de los primeros agricultores neolíticos que partieron de Oriente Medio, una de las regiones de origen de la agricultura, en todas direcciones debido a la saturación poblacional. La ecorregión de la estepa que cubre el centro de Eurasia no era especialmente favorable a la agricultura así que desarrollaron una forma de vida nómada basada en el pastoreo y la cría del caballo, abundante en la región, que les habría proporcionado leche, carne, transporte y, más tarde, poder militar (Cavalli-Sforza, 1996, 2010).

Probablemente protagonizaron la primera expansión por Asia hace unos 5000 años (3000 aC), que atravesó Turkmenistán, Afganistán, Irán, Pakistán y llegó a la India, contribuyendo a la desaparición de la civilización del valle del Indo alrededor de 1500 a.C. En las regiones de Persia e India, las lenguas indoeuropeas habladas por los pastores nómadas sustituyeron a las lenguas dravídicas que allí se hablaban (Cavalli-Sforza, 1994). En esta misma época pudieron haberse desplazado también hacia África, a través de Irán y de la Península Arábiga.

## ASIA.

\*Se analiza en un apartado independiente la información concerniente a la población gitana dadas sus peculiares características y especial interés.

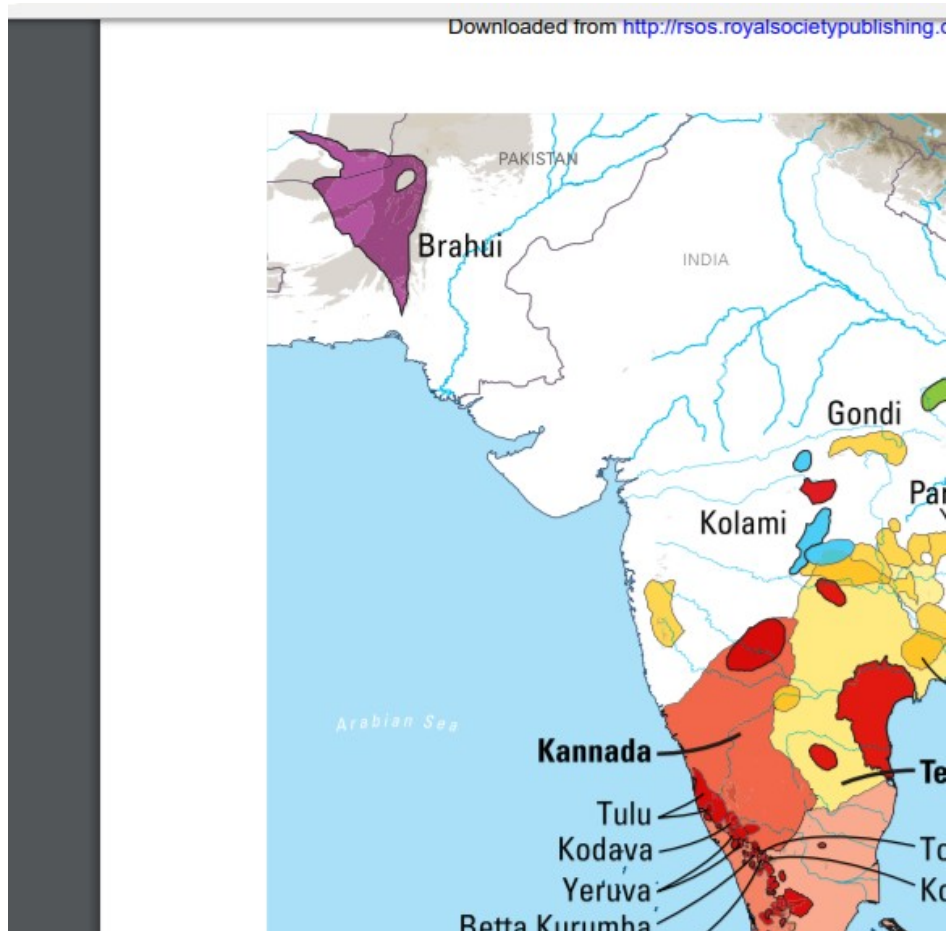
Tal y como ha quedado descrito en el apartado de resultados, en Asia se ha encontrado una clina SO → NE, sentido en el que decrece la frecuencia de los haplotipos invertidos, lo cual resulta coherente con una expansión poblacional desde Oriente Próximo, donde debido a un efecto fundador, los haplotipos H2 aumentaron su frecuencia significativamente; son las comunidades más cercanas al punto original de difusión de las configuraciones H2 quienes presentan mayores frecuencias de estos

haplotipos: los Samaritanos (0.36) y los Drusos (0.31). Ambos grupos se ubican en la Península Arábiga y están caracterizados por tratarse de comunidades relativamente pequeñas. En el caso de la población samaritana, ésta ha sufrido reducciones drásticas del número de individuos debido a guerras (Bonne B, 1966; Cazes MH, 1984) y en el caso de los Drusos, se trata de una comunidad con prácticas endógamas estrictas (Kirrish F, 1986; Nevo S, 1988).

En el resto de poblaciones asiáticas, la prevalencia de los haplotipos H2 es mucho menor, llegando a fijarse la configuración H1 en aquellos grupos ubicados más al este.

Sí resulta notable que, exceptuando a Samaritanos y a Drusos, las poblaciones con frecuencias apreciables de H2 - superiores a 0.05 - parecen restringidas al área de influencia de los idiomas indoeuropeos, en contraposición con las áreas que han conservado las lenguas drávidas o dravídicas.

Las lenguas dravídicas se hablan hoy en día sobre todo en el centro y sur de la India aunque también en algunas tribus del norte del país y de Pakistán. Antiguamente la influencia de estas lenguas se extendía desde la frontera occidental de Irán hasta el subcontinente indio, pasando por Pakistán (Cavalli-Sforza, 1994; Kolipakam V, 2018).



**Fig. D.1:** Mapa de las lenguas dravídicas de la India, Pakistán, Afganistán y Nepal, (Kolipakam V, 2018).

Aceptando la hipótesis propuesta en el apartado anterior, que explica el patrón irregular de la distribución de frecuencias de los haplotipos MAPT en África fundamentándose en que los pueblos de la estepa al norte de los mares Negro y Caspio portaban los haplotipos invertidos por mestizaje con grupos descendientes de aquellos que difundieron la agricultura desde Próximo Oriente, ésta debería tenerse en cuenta también para la interpretación de los resultados obtenidos para Asia, ya que esas mismas comunidades de pastores nómadas tuvieron una fuerte influencia en el continente.

Como se ha explicado anteriormente, los Kurganes migraron hace unos 5000 o 5500 años hacia el oeste, hasta Europa central y septentrional y también hacia el sureste de Asia a través de Asia central hasta Persia, Pakistán y la India, coincidiendo con el periodo en el que se desarrollan las primeras armas de bronce. Aunque la discusión sobre el origen de la familia lingüística indoeuropea no está cerrada, sí se asocia de manera generalizada a estas comunidades con los idiomas indoeuropeos.



Así pues, la expansión de los Kurganes hacia el sur de Asia, no sólo dejó la huella de la actual distribución de las familias lingüísticas en la zona, sino también una huella genética representada en este caso por los haplotipos H2, justificando las frecuencias que se encuentran actualmente en la India.

## AMÉRICA.

En América (N y S) el rango de valores de H2D es relativamente amplio, pero si se eliminan las poblaciones con alto grado de mestizaje, los valores restantes son siempre muy bajos (Tablas R9 a R14). En las poblaciones americanas sin ascendencia europea no hay frecuencias apreciables del haplotipo invertido, lo que concuerda con el hecho de que debió haberse perdido por efecto fundador antes de que las poblaciones humanas llegaran a este continente, ya que en el este y el norte de Asia la familia H2 es prácticamente inexistente, así como las duplicaciones  $\alpha$  y  $\beta$ . Probablemente las bajas frecuencias de los haplotipos invertidos no se conservaran mucho tiempo más tras la salida de África y su presencia en Eurasia se deba, como se ha dicho anteriormente, a una serie de efectos fundadores fruto de la expansión poblacional consecuencia de la presión demográfica tras el inicio de la agricultura en Próximo Oriente.

De esta forma, la propuesta de este trabajo es que la presencia de haplotipos invertidos en América se debe de manera exclusiva a la introducción de los mismos por europeos\*, ya que, hasta el momento, las únicas poblaciones que presentan configuraciones del linaje H2 son aquellas en las que se identifica un cierto grado de mestizaje. En África en cambio, considerando la distribución irregular detectada en el continente para los haplotipos MAPT, la presencia de haplotipos invertidos y su distribución puede ser en parte debida al origen de estos haplotipos en este continente, que en pequeñas proporciones siguieron manteniéndose en algunas poblaciones y en parte a un nuevo aporte como consecuencia del flujo genético de vuelta a África hace menos de 4.500 años desde Eurasia

## GITANOS

Al comparar las frecuencias de los haplotipos MAPT en los tres grupos del País Vasco incluidos en este trabajo (población residente con ascendencia europea, vascos autóctonos del Goierri de Gipuzkoa y población Roma afincada en esta región de la

Península Ibérica), se ha observado una proporción significativamente distinta de la inversión 17q21.31 en el pueblo gitano frente a las poblaciones vascas - tanto residentes como autóctonas - con las que comparten territorio: los Roma presentan una proporción significativamente más alta del haplotipo H1, resultados similares a los obtenidos en un estudio similar entre gitanos Olah y residentes no Roma de Hungría (Almos et al, 2008).

Cuando se analizan estas frecuencias en el contexto del continente euroasiático, mediante los test de AMOVA, el análisis de la regresión lineal y el test de Mantel, todos los datos apuntan de manera consistente a un patrón espacial de la diversidad de MAPT caracterizado por una fuerte asociación entre las coordenadas geográficas y las frecuencias de los haplotipos tau.

Sin embargo, ambas muestras de Roma, los Olah del trabajo de Almos y colaboradores (2008) y los gitanos vascos analizados en este proyecto, permanecen alejadas de la tendencia principal representada por la línea de regresión porque, en lugar de compartir los rangos de variación correspondientes a sus lugares de residencia, manifiestan frecuencias del haplotipo H2 que encajan con el rango de variación del sur de Asia, en concreto del subcontinente indio, cuyas frecuencias oscilan entre cero en los Thoti, Patanes / Pashtunes y los Kachari y 0,122 en los Mohanna (Donnelly et al, 2010).

Los grupos de la región norte de la India y de Pakistán muestran frecuencias de H2 comparables a las de las poblaciones Roma asentadas en territorio europeo (Hungría y País Vasco), con la única excepción de los ya mencionados Pashtunes. Esta similitud podría deberse a un rastro genético del origen geográfico de la diáspora Romaní.

Trabajos basados en una variedad de marcadores genéticos como SNPs (Mendizabal et al., 2012; Moorjani et al., 2013), antígenos leucocitarios humanos (HLA) (Inotai et al., 2015) o polimorfismos heredados por vía uniparental como los haplotipos de microsatélites del cromosoma Y (Y-STR) (Regueiro et al., 2012) y los haplogrupos de ADN mitocondrial (mtDNA) (Martínez-Cruz et al., 2016) han situado de manera consistente el origen de la población proto-Romaní en la región noroeste de la India.

En concreto se han propuesto las poblaciones indias actuales de Kashmir y Punjab como las candidatas más probables a ser el origen de los Roma. La estrecha relación genética detectada en los análisis de la inversión polimórfica en la región 17q21.31 en este proyecto entre los Romaní europeos y las poblaciones del norte de la India ofrece un fuerte apoyo a esta hipótesis, que también resulta compatible con

los descubrimientos realizados en estudios socioculturales y lingüísticos anteriores (Fraser, 1992).

Otro hallazgo común de los estudios arriba citados es la gran heterogeneidad genética presentada por los gitanos europeos. Sus autores sugieren largos episodios de deriva genética asociados a efectos fundadores históricos y cuellos de botella periódicos durante su diáspora a lo largo y ancho de Europa (Inotai et al., 2015; Martínez-Cruz et al., 2016). En alguno de estos trabajos se han encontrado niveles variables de flujo génico desde o hacia las poblaciones no gitanas receptoras de estas migraciones, lo cual puede haber reforzado la heterogeneidad entre los grupos Roma. En general, la dirección del flujo génico, tanto vía materna como paterna, sucede con una tendencia muy marcada desde las poblaciones receptoras hacia las poblaciones Roma y no al contrario, lo que sugiere diferentes grados de asimilación de las parejas mixtas entre los Roma y las poblaciones autóctonas (Mendizabal et al., 2012; Martínez-Cruz et al., 2016).

Cabe también destacar que varias investigaciones previas han descrito una divergencia geográfica en cuanto a la contribución genética europea a cada grupo Romaní local, un fenómeno atribuido al mestizaje diferencial con las poblaciones receptoras europeas (Martínez-Cruz et al., 2016) o, alternativamente, a las diferentes prácticas referidas a la endogamia de cada grupo Roma tras el efecto fundador inicial (Regueiro et al., 2012).

En esta línea, los análisis de los polimorfismos MAPT llevados a cabo durante este proyecto no han podido aportar evidencias sobre la heterogeneidad genética reportada entre grupos Romanís debido a efectos de la deriva o a un mestizaje diferencial con las poblaciones receptoras europeas. Esta aparente discrepancia puede justificarse, al menos, mediante dos argumentos principales. Primero, el número de muestras Romanís incluidas en el presente estudio es limitado. La escasez de datos existente para las configuraciones alternativas de la región 17q21.31 en este grupo étnico no ha permitido el análisis exhaustivo de la estructuración de los haplotipos MAPT en los gitanos europeos. Segundo, y quizás más importante, las características intrínsecas de la estructura del polimorfismo analizado - un bloque extremadamente grande (~1.5Mb) de desequilibrio de ligamiento - provoca que la probabilidad de que la región MAPT esté afectada por mutaciones recurrentes o por una reversión sea prácticamente nula y, en consecuencia, que este marcador genético sea más útil para detectar relaciones genéticas ancestrales que otros marcadores. Esta estabilidad evolutiva podría también ser responsable de la estrecha afinidad genética detectada

entre las poblaciones Romaní europeas y la hipotética población proto-Romaní del noroeste de la India y de las zonas adyacentes de Pakistán.

Como se ha mencionado anteriormente, la heterocigosidad es visiblemente menor en los Roma del País Vasco que en la población general de esta zona. El déficit significativo de heterocigosidad (equilibrio Hardy-Weinberg) detectado en la muestra de gitanos para el SNP rs10514879 podría explicarse aludiendo a los efectos de la deriva genética que sufren las sociedades con estructuras cerradas, como es el caso, y su reticencia a formar uniones interétnicas, un fenómeno reflejado en las altas tasas de endogamia y consanguinidad (Almos et al, 2008).

Las costumbres profundamente arraigadas de algunos grupos pueden convertirse en un factor limitante fundamental para el flujo génico y la mezcla de poblaciones al evitar la integración de inmigrantes en las poblaciones receptoras y al aumentar la endogamia, provocando como principal consecuencia el alejamiento de la panmixia (Alfonso-Sánchez et al. 2001).

En un sondeo sobre la prevalencia de síndromes de anomalías congénitas en una población de gitanos españoles, se estimó que la proporción de enlaces consanguíneos en el grupo Romaní era de 16 a 19,5 veces la registrada en poblaciones no romanís (Martínez-Frías y Bermejo, 1992). Asimismo, estos autores asociaron al alto nivel de endogamia de los Roma la mayor proporción de homocigotos para condiciones recesivas que se daba en la descendencia y en concordancia con un mayor ratio de síndromes recesivos, concretamente, siete veces mayor en la comunidad gitana.

Estos resultados permiten corroborar que distintos marcadores genéticos revelan diferentes capas de la evolución de las poblaciones humanas, dependiendo de las características específicas de cada loci genético en términos de nivel de polimorfismo, tasa de mutación y probabilidad de reversión de la mutación, entre otras (García-Obregón et al., 2007).

El caso concreto de la inversión MAPT es el de un marcador robusto, conservador y estable, empleable para los estudios evolutivos relacionados con la antigüedad genética de determinado grupo humano. El análisis de este polimorfismo ha aportado evidencias que permiten confirmar la estrecha relación genética entre las comunidades gitanas europeas y las poblaciones humanas que habitan actualmente la región noroeste de la India, apoyando por tanto la hipótesis de un origen en esta zona de la población que dio lugar a los Roma.

Por tanto, los estudios de la diversidad genética de la región genómica 17q21.31 podrían complementar de manera eficaz la información genética generada por los análisis de marcadores uniparentales para explorar las raíces genéticas de una comunidad humana en concreto. Por otra parte, incluso cuando se ha comprobado que el polimorfismo de la inversión es un marcador conservador y estable desde un punto de vista evolutivo, el estudio de las frecuencias de sus configuraciones ha permitido evaluar el efecto de las fuerzas evolutivas estocásticas, como la deriva genética y el flujo génico, en el contexto euroasiático, teniendo en cuenta que el análisis de la distribución espacial de los haplotipos H2 ha desvelado una asociación significativa entre la geografía y la genética representada por una clina este-oeste euroasiática.

Un valor añadido de este trabajo es la contribución de datos sobre una sociedad humana altamente endógama. La historia genética del pueblo gitano representa un caso de estudio típico para entender como una demografía compleja pueda impactar en el modelado genético de una población humana. A la luz de los descubrimientos aportados en este trabajo, puede deducirse que a pesar de la prolongada coexistencia entre los Roma residentes en País Vasco y el resto de las comunidades de ascendencia europea que habitan esta región, los Romani han preservado su contexto genético similar a las poblaciones ancestrales hipotéticas del subcontinente indio y con poca influencia del acervo genético europeo, lo que concuerda con los descubrimientos realizados en torno a los gitanos Olah de Hungría (Almos et al, 2008). Este pueblo parece por tanto refractario al mestizaje o bien los individuos que se mezclaron salieron del entorno gitano. Tomando en cuenta la falta de barreras geográficas entre los gitanos residentes en País Vasco y la población vasca, los datos observados en este trabajo apoyan la noción de que las costumbres socioculturales pueden ser barreras efectivas contra la panmixia. La profundamente arraigada etnicidad y las restricciones socioculturales de la población gitana del País Vasco podrían haber actuado de barrera a los emparejamientos al azar a través de una continua preferencia por los matrimonios intragrupal, de manera que el acervo genético nativo con un bajo porcentaje de mezcla se habría mantenido. Los resultados de este estudio constituyen un buen ejemplo de cómo la etnopsicología y los factores socioculturales podrían haber impactado notablemente en el modelado de las poblaciones humanas.



# CONCLUSIONES

---





1. La existencia de los gradientes o clinas latitudinales halladas en Europa para los haplotipos de MAPT no se debe al proceso de la recolonización postglacial de este continente desde los refugios del suroeste del continente, dado que los datos de ADN analizados en poblaciones Paleolíticas no soportan un origen tan antiguo de este patrón. Por tanto, se ha concluido que los gradientes detectados en el patrimonio genético actual deben ser motivo de fenómenos más recientes, de no más de 2000 años de antigüedad, ya que en ese período la distribución de frecuencias de MAPT era inversa a la actual.
2. Por el mismo motivo, se descarta que el origen de las altas frecuencias de MAPT\*H2 en Europa se deban a una introgresión desde neandertales u otras especies de homínidos.
3. De acuerdo con los datos estudiados, se concluye que la hipótesis más verosímil para el origen de las altas frecuencias de H2 en Europa y algunas poblaciones puntuales de Asia es aquella que propone Oriente Próximo como fuente de un primer efecto fundador que eleva las frecuencias de H2 por encima de la media del Paleolítico y que explica las altas frecuencias halladas a lo largo de la historia por medio de una serie de efectos fundadores posteriores ocurridos durante las migraciones asociadas a la expansión de la agricultura desde el foco mencionado.
4. Existe una clara asociación en poblaciones africanas entre las frecuencias de H2 (H2' y H2D) y la intensidad del flujo génico desde el sudoeste de Eurasia.
5. La presencia de H2 en América se debe probablemente a los fenómenos de mestizaje con europeos ocurridos en el continente a partir del siglo XV.
6. Las clinas actuales halladas en Asia resultan coherentes con una expansión poblacional desde Oriente Próximo, donde debido a un efecto fundador, los haplotipos H2 habrían aumentado sus frecuencias significativamente. Además, las poblaciones con frecuencias superiores a 0,05 de H2 – exceptuando Samaritanos y Drusos – parecen restringidas al área de influencia de los idiomas indoeuropeos, sugiriendo que la expansión de los kurganes, además de la huella de la actual distribución de familias lingüísticas en la zona, dejó también una huella genética detectable en este caso por las frecuencias de H2.
7. Por otra parte, incluso cuando se ha comprobado que el polimorfismo de la inversión es un marcador conservador y estable desde un punto de vista evolutivo, el estudio

de las frecuencias de sus configuraciones ha permitido evaluar el efecto de las fuerzas evolutivas estocásticas, como la deriva genética y el flujo génico, en el contexto euroasiático, teniendo en cuenta que el análisis de la distribución espacial de los haplotipos H2 ha desvelado una asociación significativa entre la geografía y la genética representada por una clina este-oeste euroasiática.

8. A la luz de los descubrimientos aportados en este trabajo, puede deducirse que a pesar de la prolongada coexistencia entre los Roma residentes en País Vasco y el resto de las comunidades de ascendencia europea que habitan esta región, los Romani han preservado su contexto genético similar a las poblaciones ancestrales hipotéticas del subcontinente indio y con poca influencia del acervo genético europeo.
9. Tomando en cuenta la falta de barreras geográficas entre los gitanos residentes en el País Vasco y la población vasca, los datos observados en este trabajo apoyan la noción de que las costumbres socioculturales pueden ser barreras efectivas contra la panmixia. Los resultados de este estudio constituyen un buen ejemplo de cómo la etnopsicología y los factores socioculturales podrían haber impactado notablemente en el modelado de las poblaciones humanas.

## REFERENCIAS

---



- ab Hortis SA (1994). Cigáni v Uhorsku 1775 / Zigeuner in Ungarn 1775
- Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ. 2008. *The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies*. PLoS One 12, 3 e1764.
- Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti AS, Bandelt HJ, Semino O & Torroni A (2005) *Saami and Berbers-an unexpected mitochondrial DNA link*. Am J Hum Genet, 76:883-886
- 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 (2004) *The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool*. Am J Hum Genet, 75:910-918
- Adams SM, Bosch E, Balaesque PL, Ballereau SJ, Lee AC, Arroyo E, López-Parra AM, Aler M, Grifo MS, Brion M, Carracedo A, Lavinha J, Martínez-Jarreta B, Quintana-Murci L, Picornell A, Ramon M, Skorecki K, Behar DM, Calafell F & Jobling MA (2008) The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. Am J Hum Genet, 83:725-736
- Agencia Estatal de Meteorología de España, Aemet, xxxx. Disponible en <http://www.aemet.es/es/serviciosclimaticos/datosclimatologicos>
- Aguirre, A., A. Vicario, L.I. Mazon, A. Estomba, M. Martinez de Pancorbo, V. Arrieta-Pico, F. Perez-Elortondo, and C.M. Lostao. 1991. Are the Basques a single and unique population? Am. J. Hum. Genet. 49: 450-458.
- Andolfatto, P., Depaulis, F. & Navarro, A. Inversion polymorphisms and nucleotide variability in Drosophila. Genet. Res. 77, 1-8 (2001).
- Alfonso-Sánchez MA, Aresti U, Peña JA & Calderón R (2005) Inbreeding levels and consanguinity structure in the Basque province of Guipúzcoa (1862-1980). Am J Phys Anthropol, 127:240-252
- Alfonso-Sánchez MA, Peña JA, Aresti U & Calderón R (2001) An insight into recent consanguinity within the Basque area in Spain. Effects of autochthony, industrialization and demographic changes. Ann Hum Biol, 28:505-521
- Alfonso-Sanchez MA, Cardoso S, Martinez-Bouzas C, Pena JA, Herrera RJ, Castro A, Fernandez-Fernandez I, De Pancorbo MM. 2008. Mitochondrial DNA haplogroup diversity in Basques: a reassessment based on HVI and HVII polymorphisms. Am J Hum Biol. 20:154-164
- Almagro Gorbea M. 2005. Etnogénesis del País Vasco: de los antiguos mitos a la investigación actual. Munibe Antropol Arkeol. 57:345-364.
- Álmos PZ, Horváth S, Czibula Á, I Raskó, Sipos B, Bihari P, Béres J, Juhász A, Janka z and Kálmán J (2008) H1 tau haplotype-related genomic variation at *17q21.3* as an Asian heritage of the European Gypsy population *Heredity* (2008) 101, 416-419
- Alonso S, Flores C, Cabrera V, Alonso A, Martin P, Albarran C, Izagirre N, de la Rúa C, Garcia O. 2005. The place of the Basques in the European Y-chromosome diversity landscape. Eur J Hum Genet. 13:1293-1302
- Alves JM, Lima AC, Pais IA, Amir N, Celestino R, Piras G, Monne M, Comas D, Heutink P, Chikhi L, Amorim A, and M. Lopes AM (2015) *Reassessing the Evolutionary History of the 17q21 Inversion Polymorphism*. Genome Biol Evol. 7(12): 3239-3248
- Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF. 2000. *The ancestry of Brazilian mtDNA lineages*. Am J Hum Genet 67, 444-461.

- Alzualde A, Izagirre N, Alonso S, Alonso A, de la Rúa C. 2005. Temporal mitochondrial DNA variation in the Basque country: influence of post-Neolithic events. *Ann Hum Genet.* 69:665–679.
- Ammerman, A.J. and L.L. Cavalli-Sforza. 1984. Neolithic transition and population genetics in Europe. Princeton University Press, Princeton, NJ.
- Andreadis A, Brown WM, Kosik KS (1992). *Structure and novel exons of the human tau gene.* *Biochemistry.* 31(43):10626-33.
- Anthony, D.W (2007) *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ. Press).
- Antonacci F, et al. 2009. Characterization of six human disease-associated inversion polymorphisms. *Hum Mol Genet.* 18(14):2555–2566
- Arenas M, Ray N, Currat M, Excoffier L. (2012). *Consequences of range contractions and range shifts on molecular diversity.* *Mol Biol Evol.* 29:207–218.
- Armitage SJ, et al. (2011) The southern route “out of Africa”: Evidence for an early expansion of modern humans into Arabia. *Science* 331(6016):453–456.
- Arrizabalaga A (2005) Las primeras ocupaciones humanas en el Pirineo Occidental y Montes Vascos. Un estado de la cuestión en 2005. *Munibe (Antropología-Arkeología)*, 57:53-70
- Arrizabalaga A (2009) The Middle to Upper Paleolithic Transition on the Basque Crossroads: Main Sites, Key Issues. *Mitteilungen der Gesellschaft für Urgeschichte*, 18:39-70
- Assal S, Susanszky E, Czeizel A (1991). High consanguinity rate in Hungarian gypsy communities. *Acta Paediatr Hung* 31: 299–304.
- Baeta Bafalluy M, 2011. Estudio de la variabilidad del genoma mitocondrial y de marcadores sexuales en grupos étnicos de Ecuador. Tesis doctoral. Universidad de Zaragoza.
- Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, Hardy J, Lynch T, Bigio E, Hutton M. 1999. *Association of an extended haplotype in the tau gene with progressive supranuclear palsy.* *Human Molecular Genetics*, Vol. 8, No. 4 711–715
- Balaresque, P., Bowden, G.R., Adams, S.M., Leung, H.Y., King, T.E., Rosser, Z.H., Goodwin, J., Moisan, J.P., Richard, C., Millward, A., et al. (2010). A predominantly neolithic origin for European paternal lineages. *PLoS Biol.* 8, e1000285.
- Ballatore C, Lee VM, Trojanowski JQ (2007). *Tau-mediated neurodegeneration in Alzheimer's disease and related disorders.* *Nat Rev Neurosci.* 8(9):663-72.
- Bamshad, M. & Wooding, S.P. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* 4, 99–111 (2003).
- Banks WE, d'Errico F, Peterson AT, Vanhaeren M, Kageyama M, Sepulchre P, Ramstein G, Jost A, Lunt D. 2008. *Human ecological niches and territories during the LGM in Europe derived from an application of eco-cultural niche modeling.* *J Archaeol Sci.*35:481–491
- Barbujani G and Goldstein DB (2004) Africans and Asians abroad: Genetic Diversity in Europe *Annu. Rev. Genomics Hum. Genet.* 5:119–150
- Barbujani G, Bertorelle G & Chikhi L (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J Hum Genet*, 62:488-492

- Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL (1997) An apportionment of human DNA diversity. *Proceedings of the National Academy of Sciences USA* 94: 4516–4519.
- Barbujani G. 1997. DNA variation and language affinities. *Am J Hum Genet* 61: 1011-1014
- Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences USA* 87: 1816–1819.
- Barciela V, Blasco R, Cuartero F, Peris JF, Hortelano L, Sañudo P (2013) Cova del Bolomor. 25 años en busca de un tiempo perdido. *Museu de Prehistòria de València - Diputació de València*. ISBN: 978-84-7795-676-1
- Barker G, et al. (2007) The ‘human revolution’ in lowland tropical Southeast Asia: The antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J Hum Evol* 52(3):243–261
- Bauduer F, Feingold J, Lacombe D. 2005. The Basques: review of population genetics and Mendelian disorders. *Hum Biol.* 77:619–637.
- Behar DM, Harmant C, Manry J, van Oven M, Haak W, et al. (2012) The Basque Paradigm: Genetic Evidence of a Maternal Continuity in the Franco- Cantabrian Region since Pre-Neolithic Times. *Am J Hum Genet* 90: 486–493.
- Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. *Nature* 466, 238–242.
- Balaresque P, Bowden GR, Adams SM, Leung HY, King TE, Rosser ZH, Goodwin J, Moisan JP, Richard C, Millward A, Demaine AG, Barbujani G, Previderè C, Wilson IJ, Tyler-Smith C, Jobling MA (2010) *A predominantly neolithic origin for European paternal lineages*. *PLoS Biol.* 8(1).
- Beckman JS, Weber JL. 1992. *Survey of human and rat microsatellites*. *Genomics* 12, 627-631.
- Bennett KD (1997). *Evolution and Ecology: The Pace of Life*. Cambridge Univ. Press, Cambridge
- Bermúdez de Castro, José María; Martínón Torres, María; Gómez Robles, Aida; Prado-Simón, Leyre; Martín Francés, Laura; Lapresa, María; Olejniczak, Anthony y Carbonell, Eudald (2011) «Early Pleistocene human mandible from Sima del Elefante (TE) cave site in Sierra de Atapuerca (Spain): A comparative morphological study». *Journal of Human Evolution*, 61(1): 12-25C
- Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D. 1995. Human mitochondrial-DNA variation and the origin of Basques. *Ann Hum Genet.* 59:63–81.
- Bertranpetit, J., and Cavalli-Sforza, L.L. (1991). A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann. Hum. Genet.* 55, 51–67.
- Bodner M, Perego UA, Huber G, Fendt L, Röck AW, Zimmermann B, Olivieri A, Gómez-Carballa A, Lancioni H, Angerhofer N, Bobillo MC, Corach D, Woodward SR, Salas A, Achilli A, Torroni A, Bandelt HJ, Parson W. 2012. *Rapid coastal spread of First Americans: Novel insights from South America’s Southern Cone mitochondrial genomes*. *Genome Res* 22, 811–820
- Boettger LM, Handsaker RE, Zody MC, McCarroll SA 2012. *Structural haplotypes and recent evolution of the human 17q21.31 region*. *Nature genetics* 44, 881 – 885.
- Bonnichsen R, Lepper BT, Stanford D, Waters MR (2005). *Paleoamerican Origins: Beyond Clovis*. Center for the Study of the First Americans, College Station, TX; pp. 183–195

- Bonne B. "Genes and phenotypes in the Samaritan isolate". *American Journal of Physical Anthropology* 24:1-19. (1966)
- Bortolini MC, Da Silva WA Junior W, De Guerra DC, Remonato G, Mirandola R, Hutz MH, Weimer TA, Silva MC, Zago MA, Salzano FM. 1999. African-derived South American populations: A history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. *Am J Hum Biol* 11, 551-563.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73, 524-539.
- Bortolini MC, Silva-Junior WA, Weimer TA, Zago MA, Castro de Guerra D, Schneider MPC, Layrisse Z, Mendez Castellano H, Salzano FM. 1998. Protein and hypervariable tandem repeat diversity in eight African- derived South American populations: Inferred relationships do not coincide. *Hum Biol* 70, 443-461.
- Bortolini MC, Weimer TA, Franco MH, Salzano FM, Layrisse Z, Schneider H, Schneider MP, Harada ML. 1992. Genetic studies in three South American black populations. *Gene Geogr* 6, 1-16.
- Bortolini MC, Weimer TA, Salzano FM, Callegari-Jacques SM, Schneider H, Layrisse Z, Bonatto SL. 1995. Evolutionary relationships between black South American and African populations. *Hum Biol* 67, 547- 559.
- Bortolini MC, Weimer TA, Salzano FM, Moura LB, Silva MCBO. 1997. Genetic structure of two urban afroBrazilian populations. *Int J Anthropol* 12: 5-16.
- Bosch MD, et al. (2015) New chronology for Ksâr 'Akil (Lebanon) supports Levantine route of modern human dispersal into Europe. *Proc Natl Acad Sci USA* 112(25): 7683-7688.
- Bouckaert R, et al. (2012) Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097):957-960.
- Bramblett GT, Goedert M, Jakes R, Merrick SE, Trojanowski JQ, Lee VMY. Abnormal tau phosphorylation at Ser396 in Alzheimer's disease recapitulates development and contributes to reduced microtubule binding. *Neuron* 1993; 10: 1089-99.
- Brandt G, Haak W, Adler CJ, Roth C, Szécsényi-Nagy A, Karimnia S, Möller-Rieker S, Meller H, Ganslmeier R, Friederich S, Dresely V, Nicklisch N, Pickrell JK, Sirocko F, Reich D, Cooper A, Alt KW & The Genographic Consortium. (2013) *Ancient DNA reveals key stages in the formation of central european mitochondrial genetic diversity*. *Science*, 342:257-261
- Briand, F., ed. The Messinian Salinity Crisis from Mega-Deposits to Microbiology – A Consensus Report (CIESM Workshop Monographs 30, CIESM, 2008)
- Brown P, Sutikna T, Morwood MJ, Soejono RP, Jatmiko, Saptomo EW, Due RA (2004). *A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia*. *Nature*. 431(7012):1055-61.
- Busby GBJ, Hellenthal G, Montinaro F, Tofanelli S, Bulayeva K, Rudan I, Zemunik T, Hayward C, Toncheva D, Karachanak-Yankova S, Nesheva D, Anagnostou P, Cali F, Brisighelli F, Romano V, Gerard Lefranc G, Buresi C, Chibani JB, Haj-Khelil A, Denden S, Ploski R, Krajewski P, Hervig T, Moen T, Herrera RJ, Wilson JF, Simon Myers S and Capelli C. 2015 The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. *Current Biology* 25, 2518-2526.



- C De-la-Rua, J Altuna, M Hervella, L Kinsley, R Grün (2014) Direct U-series analysis of the Lezetxiki humerus reveals a Middle Pleistocene age for human remains in the Basque Country (northern Iberia). *Journal of human evolution* 93, 109-119
- Cabanes MD, Ferrer R (1980) *Llibre del Repartiment del Regne de València*. EDK Anúbar Zaragoza, Spain
- Caffrey TM, Wade-Martins R 2007. *Functional MAPT haplotypes: Bridging the gap between genotype and neuropathology*. *Neurobiology of Disease* 27, 1 – 10.
- Calafell, F., and Bertranpetit, J. (1994). Mountains and genes: Population history of the Pyrenees. *Hum. Biol.* 66, 823–842.
- Calafell, F., and Bertranpetit, J. (1994). Principal component analysis of gene frequencies and the origin of Basques. *Am. J. Phys. Anthropol.* 93, 201–215.
- Calderón R, Pérez-Miranda AM, Fuciarelli M, Scano G, Carrión M, Alfonso-Sánchez MA, Peña JA, Ambrosio B, De Stefano G. 2006. Genetic polymorphisms in autochthonous Basques from northern Navarre. *Anthropol Anz* 64:173–187.
- Calderón, R., C. Vidales, J.A. Peña, A. Pérez-Miranda, and J.M. Dogoujon. 1998. Immunoglobulin allotypes (GM and KM) in Basques from Spain: Approach to the origin of the Basque population. *Hum. Biol.* 70: 667–698
- Callaway E. 2015. *South America settled in one go*. *Nature* 20, 598-599
- Callén E, Casado JA, Tischkowitz MD, Bueren JA, Creus A, Marcos R, Dası A, Estella JM, Muñoz A, Ortega JJ, de Winter J, Joenje H, Schindler D, Hanenberg H, Hodgson SV, Mathew CG, and Surrallés J (2005) A common founder mutation in FANCA underlies the world's highest prevalence of Fanconi anemia in Gypsy families from Spain *BLOOD*, 1 VOLUME 105, NUMBER 5
- Campell GL. 1998. *Concise compendium of the world's language*. London: Routledge
- Cardoso S (2008) *Diversidad del genoma mitocondrial en poblaciones autóctonas de la Cornisa Cantábrica: Huellas de la recolonización postglacial en Europa*. Universidad del País Vasco UPV/EHU, Leioa.
- Cardoso S, Valverde L, Alfonso-Sánchez MA, Palencia-Madrid L, Elcoroaristizabal X, et al. (2013) *The Expanded mtDNA Phylogeny of the FrancoCantabrian Region Upholds the Pre-Neolithic Genetic Substrate of Basques*. *PLoS ONE* 8(7)
- Cardoso S, Alfonso-Sánchez MA, Valverde L, Odriozola A, Pérez-Miranda AM, Peña JA and de Pancorbo MM. (2011) *The Maternal Legacy of Basques in Northern Navarre: New Insights Into the Mitochondrial DNA Diversity of the Franco-Cantabrian Area*. *AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY* 145:480–488
- Carolin SA, Cobb KM, Adkins JF, Clark B, Conroy JL, Lejau S, Melang J, Tuen AA. 2013. *Varied response of western Pacific hydrology to climate forcing over the last glacial period*. *Science* 340, 1564 – 1566.
- Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, RuizLinares A. 2003. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum Genet* 112, 534-541.
- Cavalli-Sforza, L.L. (1996). *Genes, pueblos y lenguas*. Ed. Crítica. Barcelona.
- Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. *The history and geography of human genes*. Princeton University Press, Princeton, NJ.

- Cavalli-Sforza L (1988) The basque population and ancient migrations in Europe MUNIBE (Antropología y Arqueología) 6 (Suppl):129-137.
- Carroll SB (2003). *Genetics and the making of Homo sapiens*. Nature. 422 (6934):849-57.
- Castro de Guerra D, Arvelo H, Rodríguez-Larralde A, Salzano FM. 1996. Genetic study in Panaquire, a Venezuelan population. Hum Her 46, 323-328.
- Cazes MH, Bonne B. "Genetic evolution of the Samaritans". *Journal of Biosocial Science* **16**:177-87. (1984)
- Chang W, Cathcart C, Hall D, Garrett A (2015) Ancestry-constrained phylogenetic analysis supports the Indo-European Steppe hypothesis. *Language* 91(1):194-244.
- Charlesworth WR (2005). Amish and Gypsy children: socialization within cohesive, strategizing groups. In: Burgess RL (Ed.), *Evolutionary Perspectives on Human Development*, Second ed, p. 464.
- Chikhi L, Nichols RA, Barbujani G, Beaumont MA. 2002. Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A*. 99:11008-11013.
- Chowdhury, R., Bois, P.R., Feingold, E., Sherman, S.L. & Cheung, V.G. (2009) Genetic analysis of variation in human meiotic recombination. *PLoS Genet*. 5 (9).
- Clark JD, et al. (2003) Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* 423(6941):747-752.
- Coe BP, Witherspoon K, Rosenfeld JA et al (2014) *Refining analyses of copy number variation identifies specific genes associated with developmental delay*. *Nat Genet* 46: 1063-1071.
- Colby DW, Prusiner SB. *Prions*. Cold Spring Harb Perspect Biol 2011; 3 (1).
- Collins R (1989) *The Arab Conquest of Spain*. EDK Blackwell, Oxford UK
- Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bertranpetit J. 1998. HLA evidence for the lack of genetic heterogeneity in Basques. *Ann Hum Genet*. 62:123-132
- Comas D, Mateu E, Calafell F, Perez-Lezaun A, Bosch E, Martinez- Arias R, Bertranpetit J. 1998. HLA class I and class II DNA typing and the origin of Basques. *Tissue Antigens* 51:30-40.
- Comrie B, Matthews S, Polinsky M. 2003. *The atlas of languages. The origin and development of languages throughout the world*. New York: Facts on File.
- Conrad, C., Vianna, C., Schultz, C., Thal, D. R., Ghebremedhin, E., Lenz, J., Braak, H., and Davies, P. (2004) Molecular evolution and genetics of the Saitohin gene and tau haplotype in Alzheimer's disease and argyrophilic grain disease *J. Neurochem*. **89**, 179-188
- Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. (2008) *High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans*. *Science*. 319:1395-8.
- Cooper GM, et al. A copy number variation morbidity map of developmental delay (2011). *Nat Genet*. 43:838-46.
- Corsi M, Crepaldi C, Lodovici MS, Bocagni P, Vasilescu C (2008). Ethnic minority and Roma women in Europe: a case for gender equality? European Commission. Directorate-General for Employment, Social Affairs and Equal Opportunities
- Creanza, N. et al. 2015. A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences of the United States of America* 112, 1265-1272

- Crosby AW. 2003. The Columbian exchange: biological and cultural consequences of 1492. Westport, CT, USA: Greenwood Press. Crosby AW. 2004. Ecological imperialism: the biological expansion of Europe. New York, USA: Cambridge University Press.
- Cuartero González F, Hortelano Piqueras L y Blasco R Museu de Prehistòria de València Cova del Bolomor: 25 años en busca de un tiempo perdido. 2013. ISBN 978-84-7795-676-1
- Curat M & Excoffier L (2005) *The effect of the Neolithic expansion on European molecular diversity*. Proc Biol Sci, 272:679-688
- DANE. Censo General Colombia. 2005. Departamento Administrativo Nacional de Estadística. 499 pp. (<http://www.dane.gov.co>).
- DANE (20067) Colombia: Una nación multicultural. [https://www.dane.gov.co/files/censo2005/etnia/sys/colombia\\_nacion.pdf](https://www.dane.gov.co/files/censo2005/etnia/sys/colombia_nacion.pdf)
- Da Silva WA Jr, Bortolini MC, Meyer D, Salzano FM, Elion J, Krishnamoorthy R, Schneider MP, De Guerra DC, Layrisse Z, Castellano HM, Weimer TD, Zago MA. 1999. Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. Am J Phys Anthropol 109, 425-437.
- De Fanti S, Barbieri C, Sarno S, Sevini F, Vianello D, Tamm E, et al. (2015) *Fine Dissection of Human Mitochondrial DNA Haplogroup HV Lineages Reveals Paleolithic Signatures from European Glacial Refugia*. PLoS ONE 10 (12)
- Demeter F, et al. (2012) Anatomically modern human in Southeast Asia (Laos) by 46 ka. Proc Natl Acad Sci USA 109(36):14375-14380.
- Dillehay TD (1997). Monte Verde: A Late Pleistocene Settlement in Chile: Volume 2: The Archaeological Context and Interpretation (Smithsonian Institution Press, Washington, DC).
- Dipierri JE, Alfaro E, Martínez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO. 1998. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. Hum Biol 70, 1001-1010.
- Dixon EJ. 2001. Human colonization of the Americas: Timing, technology and process. Quaternary Science Reviews 20, 277-299.
- Donnelly MP, Paschou P, Grigorenko E et al 2010 *The distribution and most recent common ancestor of the 17q21 inversion in humans*. The American Journal of Human Genetics 86: 161 - 171.
- Dobyns HF. 1993. Disease transfer at contact. Annu Rev Anthropol 22, 273-291.
- Drake, N. A. et al. *Palaeohydrology of the Fazzan Basin and surrounding regions: The last 7 million years*. Paleogeogr. Paleoclimatol. Paleoecol. 263, 131-145, doi: 10.1016/j.palaeo.2008.02.005 (2008).
- Dubourg C, et al. Clinical and molecular characterization of 17q21.31 microdeletion syndrome in 14 French patients with mental retardation. Eur J Med Genet. 2011; 54:144-51.
- Duda P & Zrzavý J. 2016. Human population history revealed by a supertree approach. Scientific Reports | 6:29890
- Duggen S. et al., 2003, Deep roots of the Messinian salinity crisis, Nature, 422, 602-606
- Dupanloup I, Bertorelle G, Chikhi L & Barbujani G (2004) *Estimating the impact of prehistoric admixture on the genome of Europeans*. Mol Biol Evol, 21:1361-1372

- Dyke AS. 2004. An outline of North American deglaciation with emphasis on central and northern Canada. En: Ehlers J, Gibbard PL (Eds). Quaternary Glaciations—Extent and Chronology, Part II: North America. Amsterdam, Netherlands: Elsevier.
- Eusko Jaurlaritza. V. Inkesta soziolinguistikoa (2011, 2013). Eusko Jaurlaritza, Hezkuntza, Hizkuntza politika eta kultura saila.
- [http://www.euskara.euskadi.eus/contenidos/informacion/argitalpenak/eu\\_6092/adjuntos/V.%20Inkesta.pdf](http://www.euskara.euskadi.eus/contenidos/informacion/argitalpenak/eu_6092/adjuntos/V.%20Inkesta.pdf)
- Escobar, CM (2005). Colombia Tierra de Esperanza. Ed. Aula XXI.
- Evans W, Fung HC, Steele J et al 2004 *The tau H2 haplotype is almost exclusively Caucasian in origin*. Neuroscience Letters 369: 183-185
- Excoffier L, Laval G, Shneider S. 2005. *Arlequin ver 3.0: An integrated software package for population genetics data analysis*. Evolutionary Bioinformatics Online 1, 47-50.
- Excoffier L & Lischer HE (2010) *Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows*. Molecular Ecology Resources, 10:564-567
- Excoffier, L. *Human demographic history: refining the recent African origin model*. (2002) Curr. Opin. Genet. Dev. 12, 675-682.
- Fagan BM. 2004. People of the earth. 11<sup>o</sup> Ed. New York, USA: Prentice-Hall
- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA Jr, Zago MA, Ribeiro-dosSantos AK, Santos SE, Petzl-Erler ML, Bonatto SL. 2008b. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82, 583592.
- Felsenstein, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5, 164-166.
- Fernández-Gómez AA, Velasco-Ortíz A. 2011. *Los primeros humanos en América*. Estrat Critic 5, 379-387.
- Fernández-Peris J, Barciela V, Blasco R, Cuartero F & Sañudo P (2008) El Paleolítico Medio en el territorio valenciano y la variabilidad tecno-económica de la Cova del Bolomor. Treballs
- d'Arqueologia, 14:141-169 Finlayson C., Giles Pacheco F., Rodríguez-Vidal J., et al. 2006. *Late survival of Neanderthals at the southernmost extreme of Europe*. Nature. 443, 850-853.
- Fledel-Alon A, et al. (2011) *Variation in human recombination rates and its genetic determinants*. PLoS One.
- Foissac A, Crouau-Roy B, Fauré S, Thomsen M, Cambon-Thomsen A. 1997. *Microsatellites in the HLA region: an overview*. Tissue Antigens 49, 197-214.
- Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59, 935-945.
- Fraser A. 1992. *The Gypsies*. Oxford: Blackwell Publishers.
- Fraser, A. (1998) Historia do Povo Cigano. Tradução de Costa T. Editorial Teorema, LDA
- Fullola i Pericot JM y Nadal Lorenzo J (2005). *Introducción a la prehistoria: La evolución de la cultura humana*. UOC (UNIVERSITAT OBERTA DE CATALUNYA)

- Gambetti, P., Velasco, M. E., Dahl, D., Bignami, A., Roessmann, U. & Sindley, S. P. (1980) in *Aging of the Brain and Dementia*, eds. Amaducci, L., Davison, A. N. & Antuono, P. (Raven, New York), pp. 55-63.
- Gamboa Martínez JC, Gómez Baos AD, Gómez Fuentes V, et. al. (2005) *Tras el Rastro de Melquíades. Memoria y Resistencia de los Rom de Colombia*. PRORROM. Bogotá, D.C.
- Gamble C, Davies W, Pettitt P, Richards M (2004). *Climate change and evolving human diversity in Europe during the last glacial*. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* 359: 243-254.
- Garagnani P, Laayouni H, Gonzalez-Neira A, Sikora M, Luiselli D, Bertranpetit J, Calafell F. 2009. Isolated populations as treasure troves in genetic epidemiology: the case of the Basques. *Eur J Hum Genet.* 17:1490-1494.
- Garcia O, Fregel R, Larruga JM, Alvarez V, Yurrebaso I, Cabrera VM, Gonzalez AM. 2011. Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge. *Heredity* 106:37-45.
- García-Castellanos, D. Estrada, F., Jiménez-Munt, I.; Gorini, C.; Fernández, M.; Vergés, J.; De Vidente, R. Catastrophic flood of the Mediterranean after the Messinian Crisis. (2009) *Nature* 462, 778-781
- García-Obregón S (2014). Diversidad genética de la península ibérica basada en marcadores con diferente grado de polimorfismo. Universidad del País Vasco UPV/EHU, Leioa.
- García-Obregón S, Alfonso-Sánchez MA, Gómez-Pérez L, Pérez-Miranda AM, Arroyo D, de Pancorbo MM, Peña JA. 2011. *Microsatellites and Alu elements from the human MHC in Valencia (Spain): analysis of Genetic relationships and linkage disequilibrium*. *Int J ImmunoGenet* 38, 483-491.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, Gómez-Pérez L, de Pancorbo MM, Peña JA (2010). *Genetic variability in autochthonous Basques from Guipuzcoa: a view from MHC microsatellites*. *Int J Immunogenet.* 37(4):279-87.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, de Pancorbo MM, Peña JA. 2007. *Polymorphic Alu insertions and the genetic structure of Iberian Basques*. *J Hum Genet* 52:317-327.
- Gimbutas M, *The Prehistory of Eastern Europe. Part I: Mesolithic, Neolithic and Copper Age Cultures in Russia and the Baltic Area*. 1956. Cambridge (Mass): Peabody Museum.
- Goebel T, Waters MR, O'Rourke DH. 2008. *The late Pleistocene dispersal of modern humans in the Americas*. *Science* 319, 1497-1502.
- Goedert M, Wischik CM, Crowther RA, Walker JE, Klug A. *Cloning and sequencing of the cDNA encoding a core protein of the paired helical filament of Alzheimer disease: identification as the microtubule-associated protein tau*. *Proc Natl Acad Sci USA* 1988; 85: 4051-55
- Goedert et al, 2005
- Goldstein, D.B., Ruiz Linares, A., Cavalli-Sforza, L.L. and Feldman, M.W. Genetic absolute dating based on microsatellites and the origin of modern humans (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 6723-6727
- Gómez Fuentes V, Gamboa Martínez JC y Paternina Espinosa HA (2000). *Los Rom de Colombia: Itinerario de un Pueblo Invisible*. Suport Mutu. PRORROM. Bogotá, D.C.
- Gómez-Pérez L, 2015. Procesos microevolutivos en poblaciones de Sudamérica. Tesis doctoral. Universidad del País Vasco / Euskal Herriko Unibertsitatea.

- Gonzalez AM, Garcia O, Larruga JM, Cabrera VM. 2006. The mitochondrial lineage U8a reveals a Paleolithic settlement in the Basque country. *BMC Genomics* 7:124.
- González-Fortes G., Jones E. R., Lightfoot E., Bonsall C., Lazar C., Grandal-d'Anglade A., ... & Boroneanț A. (2017). Paleogenomic evidence for multi-generational mixing between Neolithic farmers and Mesolithic hunter-gatherers in the lower Danube basin. *Current Biology*, 27(12), 1801-1810.
- Gorrochategui J, Lakarra JA (2013) Why Basque language cannot be, unfortunately, an Indo-European language. *J Indo-Eur Stud* 41(1-2):203-237.
- Gray RD, Atkinson QD (2003) Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426(6965):435-439.
- Green LD, Derr JN, Knight A. 2000. mtDNA affinities of the peoples of North-Central Mexico. *Am J Hum Genet* 66, 989-998
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Hsi-Yang Fritz M, Nancy F. Hansen, Eric Y. Durand, Anna-Sapfo Malaspinas, Jeffrey D. Jensen, Tomas Marques-Bonet,, Can Alkan, Kay Prüfer, Matthias Meyer, Hernán A. Burbano, Jeffrey M. Good,, Rigo Schultz, Ayinuer Aximu-Petri, Anne Butthof, Barbara Höber, Barbara Höffner, Madlen Siegemund, Antje Weihmann, Chad Nusbaum, Eric S. Lander, Carsten Russ, Nathaniel Novod, Jason Affourtit, Michael Egholm, Christine Verna, Pavao Rudan, O Dejana Brajkovic, Željko Kucan, Ivan Gušić, Vladimir B. Doronichev, Liubov V. Golovanova, Carles Lalueza-Fox, Marco de la Rasilla, Javier Fortea, Antonio Rosas, Ralf W. Schmitz, Philip L. F. Johnson, Evan E. Eichler, Daniel Falush, Ewan Birney, James C. Mullikin, Montgomery Slatkin, Rasmus Nielsen, Janet Kelso, Michael Lachmann, David Reich, Svante Pääbo (2010) A Draft Sequence of the Neandertal Genome *SCIENCE* VOL 328, 710.
- Gresham, D., Morar, B., Underhill, P. A., Passarino, G., Lin, A. A., Wise, C., Angelicheva, D., Calafell, F., Oefner, P. J., Shen, P., Tournay, I., de Pablo, R., Kucinskas, V., Perez-Lezaun, A., Marushiakova, E., Popov, V. & Kalaydjieva, L. (2001) Origins and divergence of the Roma (gypsies). *Am J Hum Genet* 69, 1314- 1331.
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48, 361-372.
- Gusmão A, Gusmão L, Gomes V, Alves C, Calafell F, Amorim A, Prata MJ. (2008) A perspective on the history of the Iberian gypsies provided by phylogeographic analysis of Y-chromosome lineages. *Ann Hum Genet.* 72(Pt 2):215-27.
- Gusmão A, Valente C, Gomes V, Alves C, Amorim A, Prata MJ, Gusmão L. (2010) A genetic historical sketch of European Gypsies: The perspective from autosomal markers. *Am J Phys Anthropol.* 141(4):507-14.
- Haak, W., Lazaridis, I., Patterson, N., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522(7555), 207-211.
- Haber M, Platt DE, Badro DA, et al. (13 co-authors). 2010. Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon. *Eur J Hum Genet.* 19:334-340.
- Haber M, Mezzavilla M, Xue Y and Tyler-Smith C (2016) Ancient DNA and the rewriting of human history: be sparing with Occam's razor. *Genome Biology* 17:1
- Hammer Ø, Harper DAT & Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4(1):9
- Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C., & Wall, J. D. (2011). Genetic evidence for archaic admixture in Africa. *Proceedings of the National Academy of Sciences*, 108(37): 15123-15128.



- Hancock I (1993). The emergency of a union dialect of North American Vlax Romani, and its implications for an international standard. *International Journal of Sociology of Language* 91-104.
- Harcourt AH (2016) Human phylogeography and diversity. *PNAS* vol. 113 no. 29, 8072-8078
- Hardy J, Pittman A, Myers A et al 2005 *Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens*. *Biochemical Society Transactions* 33:582-585
- Harpending, H., Jenkins, T., 1973, Genetic distance among southern African populations. En *Methods and Theories of Anthropological Genetics*, editado por M. Crawford y P. Workman (Albuquerque: University of México Press) pp. 177-99.
- Harpending HC, Ward RH. (1982). Chemical systematics and human populations. In: Nitecki M, editors. *Biochemical aspects of evolutionary biology*. Chicago: University of Chicago. p 213-256.
- Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116). Sunderland: Sinauer associates.
- Harpending HC, Jenkins T. 1973. Genetic distance among southern African populations. En: Crawford M, Workman P (Eds). *Method and Theory in Anthropological Genetics*. Albuquerque, USA: University of New Mexico Press. pp 177-199. Hedrick PW. (2004) *Genetics of Populations* (3rd Edition). Jones & Bartlett Learning
- Hearne CM, Ghosh S, Todd JA. 1992. Microsatellites for linkage analysis of Genetic traits. *Trends Genet* 8, 288-294. Review.
- Hellenthal, G. et al. *A Genetic Atlas of Human Admixture History*. *Science* 343, 747-751, doi: 10.1126/science.1243518 (2014).
- Hernández, CL, Soares, P, Dugoujon JM, Novelletto A, Rodríguez JN, Rito T, Oliveira M, Melhaoui M, Baali A, Pereira L, Calderón R. (2015) *Early Holocene and Historic mt DNA African Signatures in the Iberian Peninsula: The Andalusian Region as a Paradigm*. *PLoS ONE* 10 (10)
- Hervella M, (2010) Variación temporal del ADNmt en poblaciones de la cornisa cantábrica. Contribución del AND antiguo. Tesis Doctoral. Universidad del País Vasco - Euskal Herriko Unibertsitatea
- Hewitt G (1996). *Some genetic consequences of ice ages, and their role in divergence and speciation*. *Biological Journal of the Linnean Society*, Volume 58, Issue 3, Pages 247-276
- Hewitt G (2000). *The genetic legacy of the Quaternary ice ages*. *Nature*. 405(6789):907-13.
- Higham T, Douka K, Wood R, Ramsey CB, Brock F, Basell L, Camps M, Arrizabalaga A, Baena J, Barroso-Ruiz C, Bergman C, Boitard C, Boscato P, Caparrós M, Conard NJ, Draily C, Froment A, Galván B, Gambassini P, Garcia-Moreno A, Grimaldi S, Haesaerts P, Holt B, Iriarte-Chiapusso MJ, Jelinek A, Jordá Pardo JF, Maíllo-Fernández JM, Marom A, Maroto J, Menéndez M, Metz L, Morin E, Moroni A, Negrino F, Panagopoulou E, Peresani M, Pirson S, de la Rasilla M, Riel-Salvatore J, Ronchitelli A, Santamaria D, Semal P, Slimak L, Soler J, Soler N, Villaluenga A, Pinhasi R, Jacobi R. (2014). *The timing and spatiotemporal patterning of Neanderthal disappearance*. *Nature*. 512, 306-309.
- Higham T., Douka K., Wood R., et al. (2014) *The timing and spatiotemporal patterning of Neanderthal disappearance*. *Nature*. 512(7514):306-9.
- Hiscock P (2013) The human colonization of Australia. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 55-60.

- Hoffmann DL, Rogerson M, Spötl C, Luetscher M, Vance D, Osborne AH, Fello NM & Moseley GE (2016). *Timing and causes of North African wet phases during the last glacial period and implications for modern human migration* Scientific Reports | 6:36367
- Hofmanová Z., Kreutzer S., Hellenthal G., Sell C., Diekmann Y., Díez-del-Molino D., & Kirsanow K. (2016). *Early farmers from across Europe directly descended from Neolithic Aegeans*. Proceedings of the National Academy of Sciences, 113(25), 6886-6891.
- Hoglinger GU, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat Genet. 2011; 43:699-705.
- Holzer, M., Craxton, M., Jakes, R., Arendt, T. and Goedert, M., 2004. *Tau gene (MAPT) sequence variation among primates*. Gene 341, 313-322.
- Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, Bergmann I, Le Cabec A, Benazzi S, Harvati K, Gunz P (2018) *New fossils from Jebel Irhoud, Morocco, and the pan-African origin of Homo sapiens*. Nature 546, 289-292.
- Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, Chakraverty S, Isaacs A, Grover A, Hackett J, Adamson J, Lincoln S, Dickson D, Davies P, Petersen RC, Stevens M, de Graaff E, Wauters E, van Baren J, Hillebrand M, Joosse M, Kwon JM, Nowotny P, Che LK, Norton J, Morris JC, Reed LA, Trojanowski J, Basun H, Lannfelt L, Neystat M, Fahn S, Dark F, Tannenberg T, Dodd PR, Hayward N, Kwok JB, Schofield PR, Andreadis A, Snowden J, Craufurd D, Neary D, Owen F, Oostra BA, Hardy J, Goate A, van Swieten J, Mann D, Lynch T, Heutink P (1998). *Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17*. Nature. 393(6686):702-5.
- Hsu K, 1983, The mediterranean was a desert, Princeton University Press
- INDEC. Censo General, 2001. Instituto Nacional De Estadística y Censos. <http://www.indec.gov.ar/>
- INEC. Censo General. 2014. Instituto Nacional de Estadística y Censos. <http://www.ecuadorencifras.gob.ec/>
- Instituto Nacional de Estadística y Censos de Argentina, INDEC, 2001. Disponible en <http://www.indec.mecon.ar/>
- Instituto Nacional de Estadística y Censos de Ecuador, INEC, 2014. Información Censal Cantonal. Disponible en <http://www.inec.gob.ec/>
- Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia, IDEAM, xxxx. Disponible en <http://institucional.ideam.gov.co/jsp/index.jsf>.
- Instituto Nacional de Estadística de España, INE, 2012. Disponible en <http://www.ine.es/>
- Fledel-Alon, A. et al. (2011). Variation in human recombination rates and its genetic determinants. PLoS ONE 6 (6).
- Iriondo M, Barbero MC, Manzano C. 2003. DNA polymorphisms detect ancient barriers to gene flow in Basques. Am J Phys Anthropol. 122:73-84.
- Ittner LM, Ke YD, Delerue F, et al. *Dendritic function of tau mediates amyloid-beta toxicity in Alzheimer's disease mouse models*. Cell 2010; 142: 387-97.
- Jobling MA, Hurles M, Tyler-Smith C (2004). Human evolutionary genetics : origins, peoples & disease. Garland Science, New York.
- Jochim M (1987). *Late Pleistocene Refugia in Europe*. In: The Pleistocene Old World. Soffer O, editor. Springer: 317 - 331.



- Jochim M (2011). *The Upper Paleolithic European prehistory*. In: Milisauskas S. *European Prehistory: A Survey*. Springer: 67-124.
- Kalaydjieva, L., Calafell, F., Jobling, M. A., Angelicheva, D., de Knijff, P., Rosser, Z. H., Hurles, M. E., Underhill, P., Tournev, I., Marushiakova, E. & Popov V (2001a) Patterns of inter- and intragroup genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur J Hum Genet* **9**, 97-104.
- Kalaydjieva, L., Gresham, D. & Calafell, F. (2001b) Genetic studies of the Roma (Gypsies): a review. *BMC Medical Genetics* **2**, 5.
- Kalaydjieva L, Morar B, Chaix R, Tang H. (2005) A newly discovered founder population: the Roma/Gypsies. *Bioessays*. (10):1084-94.
- Kang, J., Lemaire, H. G., Unterbeck, A., Salbaum, J. M., Masters, C. L., Grzeschik, K. H., Multhaup, G., Beyreuther, K. & Muller-Hill, B. (1987) *Nature* (London) 325, 733-736.
- Kirrish F. "Definition of ethnic boundaries - the case of the Druze of the Golan heights". *Cambridge Anthropology***11**:70-8. (1986)
- Klein RG (2009) *The Human Career: Human Biological and Cultural Origins* (University of Chicago, Chicago), 3rd Ed.
- Kolipakam V, Jordan FM, Dunn M, Greenhill SJ, Bouckaert R, Gray RD, Verkerk A. 2018 A Bayesian phylogenetic study of the Dravidian language family. *R. Soc. open sci.*5: 171504.
- Kong A, et al. (2004) *Recombination rate and reproductive success in humans*. *Nat Genet.* 36:1203
- Koolen, D.A., Kramer, J. M., Neveling, K., Nillesen, W. M., Moore-Barton, H. L., Elmslie, F. V., Toutain, A., Amiel, J., Malan, V., Tsai, A. C.-H., Cheung, S. W., Gilissen, C., and 13 others. (2012) Mutations in the chromatin modifier gene *KANSL1* cause the 17q21.31 microdeletion syndrome. *Nature Genet.* 44: 639-641.
- Koolen DA, Pfundt R, Linda K, Beunders G, Veenstra-Knol HE, Conta JH, Fortuna AM, Gillessen-Kaesbach G, Dugan S, Halbach S, Abdul-Rahman OA, Winesett HM, Chung WK, Dalton M, Dimova PS, Mattina T, Prescott K, Zhang HZ, Saal HM, Hehir-Kwa JY, Willemsen MH, Ockeloen CW, Jongmans MC, Van der Aa N, Failla P, Barone C, Avola E, Brooks AS, Kant SG, Gerkes EH, Firth HV, Öunap K, Bird LM, Masser-Frye D, Friedman JR, Sokunbi MA, Dixit A, Splitt M; DDD Study, Kukulich MK, McGaughran J, Coe BP, Flórez J, Nadif Kasri N, Brunner HG, Thompson EM, Gecz J, Romano C, Eichler EE, de Vries BB. (2016) *The Koolen-de Vries syndrome: a phenotypic comparison of patients with a 17q21.31 microdeletion versus a KANSL1 sequence variant* *European Journal of Human Genetics* 24, 652-659
- Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, et al. (2006) *A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism*. *Nat Genet* 38: 999-1001.
- Koolen DA, Sharp AJ, Hurst JA et al (2008) *Clinical and molecular delineation of the 17q21.31 microdeletion syndrome*. *J Med Genet.* 45: 710-720
- Kosik, K. S., Joachim, C. L. & Selkoe, D. J. (1986) *Microtubule-associated protein tau (tau) is a major antigenic component of paired helical filaments in Alzheimer disease*. *Proc. Natl. Acad. Sci. USA* 83, 4044-4048.
- Krauß R., Marinova E., De Brue H., & Weninger B. (2018). The rapid spread of early farming from the Aegean into the Balkans via the Sub-Mediterranean-Aegean Vegetation Zone. *Quaternary International*, 496, 24-41. (Añadir, de la memoria Ainhoa)

- Krijgsman W. et al., (1999) Chronology, causes and progression of the Messinian salinity crisis, *Nature*, 400, 652-655
- Kurlansky M (1999) *The Basque History of the World* (Jonathan Cape, London).
- Kwok JB, Teber ET, Loy C, Hallupp M, Nicholson G, Mellick GD, Buchanan DD, Silburn PA, Schofield PR (2004). *Tau haplotypes regulate transcription and are associated with Parkinson's disease*. *Ann Neurol*. 55(3):329-34.
- Laayouni H, Calafell F, Bertranpetit J. 2010. *A genome-wide survey does not show the genetic distinctiveness of Basques*. *Hum Genet*. 127:455-458.
- Lazaridis, I., Patterson, N., Mittnik, A., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 513(7518), 409-413.
- Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, Katarzyna Hutnik, Ellen C. Royrvik, Barry Cunliffe, Wellcome Trust Case Control Consortium, International Multiple Sclerosis Genetics Consortium, Lawson DJ, Falush D, Freeman C, Pirinen M, Myers S, Robinson M, Donnelly P & Bodmer W (2015) The fine-scale genetic structure of the British population *Nature* **volume519**, 309-314
- LEY FORAL 18/86, de 15 de diciembre de 1986, del Vascuence. Regulación de su uso normal y oficial.
- Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100-1104.
- Lipson, M., Szécsényi-Nagy, A., Mallick, S., et al (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature*, 551(7680), 368.
- Liu, K. & Muse, S. (2005) PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics*, 21, 2128.
- Liu W, et al. (2015) The earliest unequivocally modern humans in southern China. *Nature* 526(7575):696-699.
- Livi-Bacci. 2006. The Depopulation of Hispanic America after the Conquest. *Popul dev rev* 32, 199-232.
- Lowery et al, (2013) *Neanderthal and Denisova genetic affinities with contemporary humans: Introgression versus common ancestral polymorphisms* *Gene*. 530(1):83-94
- Lupski JR: *Genomic rearrangements and sporadic disease*. *Nat Genet* 39: S43-S47 (2007).
- Lucotte G, Hazout S. 1996. Y-chromosome DNA haplotypes in Basques. *J Mol Evol*. 42:472-475.
- Malayarchuk, B. A., Grzybowski, T., Deremko, M. V., Czarny, J. & Miscicka-Sliwka, D. (2006) Mitochondrial DNA diversity in the Polish Roma. *Ann Hum Genet* **70**, 195-206.
- Mallory JP (2013) Twenty-first century clouds over Indo-European homelands. *J LangRelatsh* 9:145-154.
- Mandryk CAS. 2004. Invented traditions and the ultimate origin myth: in the beginning... there was an ice freecorridor. En: Barton CM, Clark GA, Yesner DR, Pearson GA (eds). *The settlement of the American continents*. Tucson, USA: University of Arizona Press. pp 113-122.
- Mann CC. 2006. 1491. Una nueva historia de las Américas antes de Colón. Madrid, España: Taurus.

- Manzano C, de la Rúa C, Iriondo M, Mazon LI, Vicario A, Aguirre A. 2002. Structuring the genetic heterogeneity of the Basque population: a view from classical polymorphisms. *Hum Biol.* 74:51-74.
- Manzano, C., A.I. Aguirre, M. Iriondo, M. Martin, L. Osaba, and C. de la Rúa. 1996a. Genetic polymorphisms of the Basques from Gipuzkoa: Genetic heterogeneity of the Basque population. *Ann. Hum. Biol.* 23: 285-296.
- Manzano, C., J.M. Orue, and C. de la Rúa. 1996B. The “basqueness” of Alava: A reappraisal from a multidisciplinary perspective. *Am. J. Phys. Anthropol.* 99: 249-258.
- Marcheco-Teruel B, Parra EJ, Fuentes-Smith E, Salas A, Buttenschøn HN, Demontis D, Torres-Español M, MarínPadrón LC, Gómez-Cabezas EJ, Alvarez-Iglesias V, Mosquera-Miguel A, Martínez-Fuentes A, Carracedo A, Børglum AD, Mors O. 2014. Cuba: exploring the history of admixture and the genetic basis of pigmentation using autosomal and uniparental markers. *PLoS Genet* 10, e1004488.
- Mardini AC, Rodenbusch R, Schumacher S, Chula FG, Michelon CT, Gastaldo AZ, Maciel LP, de Matos Almeida SE, da Silva CM. 2013. Mutation rate estimates for 13 STR loci in a large population from Rio Grande do Sul, Southern Brazil. *Int J Legal Med* 127, 45-47.
- Mariezkurrena Gastearena, K (1990) Caza y Domesticación durante el Neolítico y Edad de los Metales en el País Vasco *Munibe Antropologia - Arkeologia*, ISSN 1132-2217, Nº 42, págs. 241-252
- Marshall OJ. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* 2004 20(15):2471-2472
- Martínez-Cruz B, Harmant C, Platt DE, Haak W, Manry J, Ramos-Luis E, Soria-Hernanz DF, Bauduer F, Salaberria J, Oyharçabal B, Quintana-Murci L, Comas D; Genographic Consortium. Evidence of Pre-Roman Tribal Genetic Structure in Basques from Uniparentally Inherited Markers (2012) *Mol. Biol. Evol.* Research article 29(9):2211-2222.
- Martínez-Frías ML, Bermejo E (1992). *Prevalence of congenital anomaly syndromes in a Spanish gypsy population.* *J Med Genet.* 29(7):483-6.
- Martinez-Labarga C, Rickards O, Scacchi R, Corbo RM, Biondi G, Pena JA, De Vieira CV, Guevara AEB, Santurino MSM, De Stefano GF. 1999. Genetic population structure of two African- Ecuadorian communities of Esmeraldas. *Am J Phys Anthropol* 109, 159 -174.
- Martins, J. A., Silva, R. H., Freschi, A., Paneto, G. G., Oliveira, R. N. y Cicarelli, R. M. (2010b). Xchromosome genetic variation in Sao Paulo State (Brazil) population. *Ann Hum Biol* 37, 598- 603.
- Marushiakova E, Popov V (1997) *Gypsies (Roma) in Bulgaria.* Frankfurt am Main: Peter Lang.
- Matras Y (2006). Romani. In: Brown K (Ed.). *Encyclopedia of Language and Linguistics*, Oxford.
- McDougall I, Brown FH, Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433(7027):733-736.
- Mehan MR, Freimer NB, Ophoff RA (2004). *A genomewide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture.* *Hum Genomics* 1: 335-344.
- Mendizabal I, Comas D. La historia de los gitanos europeos (2013). *Investigación y Ciencia*, nº 446.
- Mendizabal I, Lao O, Marigorta UM, Wollstein A, Gusmão L, Ferak V, Ioana M, Jordanova A, Kaneva R, Kouvatzi 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. (2012)

- Reconstructing the population history of European Romani from genome-wide data. *Curr Biol.* 22(24):2342-9.
- Mendizabal 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. (2011) Reconstructing the Indian origin and dispersal of the European Roma: a maternal genetic perspective. *PLoS One.* 6(1)
- Merriwether DA, Huston S, Iyengar S, Hamman R, Norris JM, Shetterly SM, Kamboh MI, Ferrell RE. 1997. Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. *Am J Phys Anthropol* 102, 153-159.
- Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortíz-Barrientos D, García LF, Velez ID, Bravo ML, Múnera JG, Bedoya G, Bortolini MC, Ruiz-Linares A. 2000. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of Gene flow in South America. *Am J Hum Genet* 67, 1277-1286.
- Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF, Shunkov MV, Derevianko AP, Patterson N, Andrés AM, Eichler EE, Slatkin M, Reich D, Kelso J, Pääbo S. (2012) A High-Coverage Genome Sequence from an Archaic Denisovan Individual 12;338(6104):222-6.
- Michelena L, Pagola M, Agirrebaltzategi M, Intxausti J, Villasante L, et al. (1977) El Libro Blanco de La Lengua Vasca. Euskaltzaindia, Real Academia de la Lengua Vasca. Bilbao.
- Mirazón-Lahr, M (2016) The shaping of human diversity: filters, boundaries and transitions. *Phil. Trans. R. Soc. B* 371
- Mix AC, Bard E, Schneider R. 2001. Environmental processes of the ice age: land, oceans, glaciers (EPILOG). *Quaternary Science Reviews* 20, 627 - 657
- Morar, B., Gresham, D., Angelicheva, D., Tournev, I., Gooding, R., Guergueltcheva, V., Schmidt, C., Abicht, A., Lochmüller, H., Tordai, A., Kalmár, L., Nagy, M., Karcagi, V., Jeanpierre, M., Herczegfalvi, A., Beeson, D., Venkataraman, V., Carter, K. W., Reeve, J., de Pablo, R., Kučinskas, V. & Kalaydjieva, L. (2004) Mutation history of the Roma/Gypsies. *Am J Hum Genet* 75, 596-609.
- Mourant, A.E. 1947. The blood groups of the Basques. *Nature* 160: 505-506.
- MPH, Museo de la PreHistoria de Valencia.
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51, 263-273
- Myers AJ, et al. The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. *Neurobiol Dis.* 2007; 25:561-70.
- Nevo S. "Genetic blood markers in Arab Druze of Israel". *American Journal of Physical Anthropology* 77:183-190. (1988)
- O'Connell JF, Allen J (2004) Dating the colonization of Sahul (Pleistocene Australia- New Guinea): A review of recent research. *J Archaeol Sci* 31(6):835-853.
- Olalde I, et al. (2014) Derived immune and ancestral pigmentation alleles in a 7,000- year-old Mesolithic European. *Nature* 507(7491):225-228.
- Olalde I., Brace S., Allentoft M. E., Armit I., Kristiansen K., Booth T., ... Reich D. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*, 555(7695), 190-196.

- Oppenheimer S (2006) *The Origins of the British* (Constable & Robinson Ltd., London).
- Pakendorf, B. 2014. Coevolution of languages and genes. *Current Opinion in Genetics & Development* 29, 39–44.
- Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, et al. *Mitochondrial DNA signals of late glacial recolonization of Europe from near Eastern refugia*. *Am.J. Hum.Genet.*2012; 90:915–924.
- Palacios Preciado, J (1989) Nueva Historia de Colombia Vol. 1. La esclavitud y la sociedad esclavista. Jorge Palacios Preciado. Ed: PLANETA COLOMBIANA EDITORIAL S.A. Pag. 168-170.
- Pascale E, Di Battista ME, Rubino A, Purcaro C, Valente M, Fattapposta F, Ferraguti G and Meco G (2016) *Genetic Architecture of MAPT Gene Region in Parkinson Disease Subtypes*. *Front. Cell. Neurosci.* 10:96.
- Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, Kjeldsen KK, Mendoza ML, Beaudoin AB, Zutter C, Larsen NK, Potter BA, Nielsen R, Rainville RA, Orlando L, Meltzer DJ, Kjær KH, Willerslev E (2016) Postglacial viability and colonization in North America's ice-free corridor. *Nature*. 537(7618):45-49.
- Pennisi E 2008 *Genetics. 17q21.31: not your average genomic address*. *Science* 322, 842–845
- Peña JA, Alfonso-Sánchez MA, Pérez-Miranda AM, García-Obregón S, Gómez-Pérez L. 2009. GeDis: *Un programa para análisis de datos en Antropogenética*. *Antropo* 20, 49-56. [www.didac.ehu.es/antropo](http://www.didac.ehu.es/antropo).
- Peña JA, García-Obregón S, Pérez-Miranda AM, De Pancorbo MM & Alfonso-Sánchez MA (2006) *Gene flow in the Iberian Peninsula determined from Y-chromosome STR loci*. *Am J Hum Biol*, 18:532-539
- Peña JA, Alfonso-Sánchez MA, Pérez-Miranda AM, García-Obregón S (2004) Gene flow and selective pressures in Europe. An approach from genic and nongenetic DNA clines. *Antropo*, 8:177-122
- Peña JA, Calderón R, Pérez-Miranda AM, Vidales C, Dugoujon JM, Carrion M, Crouau-Roy B (2002). *Microsatellite DNA markers (D6S105, D6S265 and TFINa) from HLA region in autochthonous Basques from Northern Navarre (Spain)*. *Ann Hum Biol* 29, 176-191.
- Pereira L, Richards M, Goios A, Alonso A, Albarrán C, Garcia O, Behar DM, Gölge M, Hatina J, Al-Gazali L, Bradley DG, Macaulay V & Amorim A (2005) *High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium*. *Genome Res*, 15:19-24
- Pérez-Miranda AM, Alfonso-Sánchez MA, Kalantar A, García-Obregón S, de Pancorbo MM, Peña JA, Herrera RJ. 2005. Microsatellite data support subpopulation structuring among Basques. *J Hum Genet* 50:403–414.
- Perez-Miranda, A. M., Alfonso-Sanchez, M. A., Kalantar, A., García-Obregón, S., de Pancorbo, M. M., Peña, J. A., & Herrera, R. J. (2005). Microsatellite data support subpopulation structuring among Basques. *Journal of human genetics*, 50(8), 403.
- Pérez-Miranda AM, Alfonso-Sánchez MA, Vidales MC, Calderón R, Peña JA. 2004. Genetic polymorphism and linkage disequilibrium of the HLA-DP region in Basques from Navarre (Spain). *Tissue Antigens* 64:264–275.
- Pickrell JK, Patterson N, Loh PR, Lipson M, Berger B, Stoneking M, Pakendorf B, Reich D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. *Proc Natl Acad Sci U S A*. 111(7):2632-7

- Pinhasi R, Thomas MG, Horeiter M, Currat M, Burger J (2012). The genetics history of Europeans. *Trends in Genetics*, vol.28, No, 10.
- Pittman AM, et al. *The structure of the tau haplotype in controls and in progressive supranuclear palsy*. *Hum Mol Genet*. 2004; 13:1267-74.
- Pla-Ballester E (1985) Estado actual del conocimiento de la Prehistoria en el País Valenciano.
- Zerphyrus: *Revista de Prehistoria y Arqueología*, 37-38:7-16 Poorkaj P, Bird TD, Wijsman E, Nemens E, Garruto RM, Anderson L, Andreadis A, Wiederholt WC, Raskind M, Schellenberg GD (1998). *Tau is a candidate gene for chromosome 17 frontotemporal dementia*. *Ann Neurol*. 43(6):815-25.
- Pope KO, Terrell JE (2008) *Environmental setting of human migrations in the circum-Pacific region*. *J Biogeogr* 35(1):1-21.
- Prieto, L., Zimmermann, B., Goios, A., Rodríguez-Monge, A., Paneto, G.G., Alves, C., Alonso, A., Fridman, C., Cardoso, S., Lima, G., et al. (2011). The GHEP-EMPOP collaboration on mtDNA population data—A new resource for forensic casework. *Forensic Sci. Int. Genet.* 5, 146-151.
- Prüfer K, et al. 2014. The complete genome sequence of a Neandertal from the Altai Mountains. *Nature*. 505:43-49.
- Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspinas AS, Eriksson A, Moltke I, Metspalu M, Homburger JR, Wall J, Cornejo OE, Moreno-Mayar JV, Korneliussen TS, Pierre T, Rasmussen M, Campos PF, de Barros Damgaard P, Allentoft ME, Lindo J, Metspalu E, Rodríguez-Varela R, Mansilla J, Henriksen C, Seguin-Orlando A, Malmström H, Stafford T Jr, Shringarpure SS, Moreno-Estrada A, Karmin M, Tambets K, Bergström A, Xue Y, Warmuth V, Friend AD, Singarayer J, Valdes P, Balloux F, LeBoreiro I, Vera JL, Rangel-Villalobos H, Pettener D, Luiselli D, Davis LG, Heyer E, Zollikofer CPE, Ponce de León MS, Smith CI, Grimes V, Pike KA, Deal M, Fuller BT, Arriaza B, Standen V, Luz MF, Ricaut F, Guidon N, Osipova L, Voevodina M, Posukh OL, Balanovsky O, Lavryashina M, Bogunov Y, Khusnutdinova E, Gubina M, Balanovska E, Fedorova S, Litvinov S, Malyarchuk B, Derenko M, Moshier MJ, Archer D, Cybulski J, Petzelt B, Mitchell J, Worl R, Norman PJ, Parham P, Kemp BM, Kivisild T, Tyler-Smith C, Sandhu MS, Crawford M, Vilems R, Smith DG, Waters MR, Goebel T, Johnson JR, Malhi RS, Jakobsson M, Meltzer DJ, Manica A, Durbin R, Bustamante CD, Song YS, Nielsen R, Willerslev E (2015) *POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans*. *Science*.; 349 (6250)
- Ramal LM, de Pablo R, Guadix MJ, Sánchez J, Garrido A, Garrido F, Jiménez-Alonso J, López-Nevo MA. 2000. HLA class II allele distribution in the Gypsy community of Andalusia, southern Spain *Tissue Antigens* 57: 138-143
- Rao PN, Li W, Vissers LE, Veltman JA, Ophoff RA. (2010) *Recurrent Inversion Events at 17q21.31 Microdeletion Locus Are Linked to the MAPT H2 Haplotype* *Cytogenet Genome Res* 129:275-279
- Reed, D.L., Smith, V.S., Hammond, S.L., Rogers, A.R. & Clayton, D.H. *Genetic analysis of lice supports direct contact between modern and archaic humans*. *PLoS*
- Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. et al. (2001). *Linkage disequilibrium in the human genome*. *Nature*, 411, 199-204.
- Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AM, Ko YC, Jinam TA, Phipps ME, et al. 2011. *Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania*. *Am J Hum Genet*. 89:516-528.



- Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, et al. 2010. *Genetic history of an archaic hominin group from Denisova Cave in Siberia*. *Nature* 468:1053–1060.
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, García LF, Triana O, Blair S, Maestre A, Dib JC, Bravi CM, Bailliet G, Corach D, Hünemeier T, Bortolini MC, Salzano FM, Petzl-Erler ML, Acuña-Alonzo V, Aguilar-Salinas C, Canizales-Quinteros S, Tusié-Luna T, Riba L, Rodríguez-Cruz M, Lopez-Alarcón M, Coral-Vazquez R, Canto-Cetina T, Silva-Zolezzi I, Fernandez-Lopez JC, Contreras AV, Jimenez-Sanchez G, Gómez-Vázquez MJ, Molina J, Carracedo A, Salas A, Gallo C, Poletti G, Witonsky DB, Alkorta-Aranburu G, Sukernik RI, Osipova L, Fedorova SA, Vasquez R, Villena M, Moreau C, Barrantes R, Pauls D, Excoffier L, Bedoya G, Rothhammer F, Dugoujon JM, Larrouy G, Klitz W, Labuda D, Kidd J, Kidd K, Di Rienzo A, Freimer NB, Price AL, Ruiz-Linares A. (2012) *Reconstructing Native American population history*. *Nature*. 488(7411):370-4.
- Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term Genetic distance. *Genetics* 105, 767-779.
- Richards, M., Córte-Real, H., Forster, P., Macaulay, V., Wilkinson- Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.J., and Sykes, B. (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* 59, 185–203.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A & Bandelt HJ (2000) *Tracing European founder lineages in the Near Eastern mtDNA pool*. *Am J Hum Genet*, 67:1251-1276
- Richter D, Grün R, Joannes-Boyau R, Steele TE, Amani F, Rué M, Fernandes P, Raynal JP, Geraads D, Ben-Ncer A, Hublin JJ, McPherron SP (2017) *The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age*. *Nature* 546, 293–296.
- Rios-Garaizar J, Garate D, Bourrillon R, Gómez-Olivencia A and Karampaglidis T (2015) *The Venuses Block from Arlanpe Cave (Northern Iberian Peninsula): Implications for the Origins and Dispersion of Gönnersdorf-Lalinde Style Depictions Throughout the European Magdalenian*. *Oxford Journal of Archaeology* Volume 34, Issue 4, pages 321–341.
- Rodriguez-Ezpeleta N, Alvarez-Busto J, Imaz L, Regueiro M, Azcarate MN, Bilbao R, Iriondo M, Gil A, Estonba A, Aransay AM. 2010. High-density SNP genotyping detects homogeneity of Spanish and French Basques and confirms their genomic distinctiveness from other European populations. *Hum Genet.* 128:113–117.
- Roewer L, Nothnagel M, Gusmão L, Gomes V, González M, Corach D, Sala A, Alechine E, Palha T, Santos N, Ribeiro-Dos-Santos A, Geppert M, Willuweit S, Nagy M, Zweynert S, Baeta M, Núñez C, Martínez-Jarreta B, González-Andrade F, Fagundes de Carvalho E, da Silva DA, Builes JJ, Turbón D, Lopez Parra AM, Arroyo-Pardo E, Toscanini U, Borjas L, Barletta C, Ewart E, Santos S, Krawczak M. 2013. Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. *PLoS Genet* 9.
- Sajantila, A. and S. Pääbo. 1995. Language replacement in Scandinavia. *Nat. Genet.* 11: 359–360.
- Sajantila, A., P. Lahermo, T. Anttinen, M. Lukka, P. Sistonen, M. Savontaus, P. Aula, L. Beckman, L. Tranebjaerg, T. Gedde-Dahl et al. 1995. Genes and languages in Europe: An analysis of mitochondrial lineages. *Genome Res.* 5: 42–52.

- Salas, A., Comas, D., Lareu, M.V., Bertranpetit, J., and Carracedo, A. (1998). mtDNA analysis of the Galician population: A genetic edge of European variation. *Eur. J. Hum. Genet.* 6, 365–375.
- Salas A, Richards M, Lareu MV, Sobrino B, Silva S, Matamoros M, Macaulay V, Carracedo A. 2005. Shipwrecks and founder effects: divergent demographic histories reflected in Caribbean mtDNA. *Am J Phys Anthropol.* 128(4):855-60.
- Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet.* 74(3):454-65.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Sánchez-Quinto F, et al. (2012) Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr Biol* 22(16):1494–1499.
- Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N and Reich D (2014) The landscape of Neandertal ancestry in present-day humans. *Nature.* 507(7492): 354–357.
- Sans M, Salzano FM, Chakraborty R. 1997. Historical Genetics in Uruguay: estimates of biological origins and their problems. *Hum Biol* 69, 161-170.
- Sans M, Weimer TA, Franco MHL, Salzano FM, Bentancor N, Alvarez I, Bianchi NO, Chakraborty R. 2002. Unequal contributions of male and female Gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am J Phys Anthropol* 118, 33–44.
- Sans M. 2000. Admixture studies in Latin America: from the 20th to the 21st century. *Hum Biol.* 72(1):155-77.
- Santos Alonso and John A.L. Armour. 1998. MS205 Minisatellite Diversity in Basques: Evidence for a Pre-Neolithic Component.
- Skipper L, et al. Linkage disequilibrium and association of MAPT H1 in Parkinson disease. *Am J Hum Genet.* 2004; 75:669–77.
- Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, Petzl-Erler ML, Salzano FM, Patterson N, Reich D (2015). *Genetic evidence for two founding populations of the Americas.* *Nature.* ;525(7567):104-8.
- Seto-Salvia N, et al. *Dementia risk in Parkinson disease: disentangling the role of MAPT haplotypes.* *Arch Neurol.* 2011; 68:359–64.
- Setó-Salvia N, Sánchez-Quinto F, Carbonell E, Lorenzo C, Comas D, Clarimón J. (2012) *Using the neanderthal and denisova genetic data to understand the common MAPT 17q21 inversion in modern humans.* *Hum Biol.* 2012 Dec;84(6):633-40.
- Seto-Salvia N, et al. (2011) *Dementia risk in Parkinson disease: disentangling the role of MAPT haplotypes.* *Arch Neurol.* 68:359–64.
- Shahani N, Brandt R (2002). *Functions and malfunctions of the tau proteins.* *Cell Mol Life Sci.* (10):1668-80.
- Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, et al. (2006) *Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome.* *Nat Genet* 38: 1038–1042.



- Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, et al: Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* 38: 1032–1037 (2006).
- Simon-Sanchez J, et al. (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet.* 41:1308–12.
- Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, Crosslin DR, Hebring SJ, Jarvik GP, Kullo IJ, Li R, Pathak J, Ritchie MD, Roden DM, Verma SS, Tromp G, Prato JD, Bush WS, Akey JM, Denny JC, Capra JA. (2016). *The phenotypic legacy of admixture between modern humans and Neandertals.* *Science.* 351(6274):737-41.
- Singhal DP, 1982. Gypsies: Indians in exile. Berkeley CA: Archana Publications.
- Skipper, L. et al. Linkage disequilibrium and association of MAPT H1 in Parkinson disease. *Am. J. Hum. Genet.* 75, 669–677 (2004).
- Smith, E. R., Cayrou, C., Huang, R., Lane, W. S., Cote, J., Lucchesi, J. C. (20065) *A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16.* *Molec. Cell. Biol.* 25: 9175-9188 Note: Erratum: *Molec. Cell. Biol.* 26: 387
- Sokal RR, Harding RM, Oden NL (1989). *Spatial patterns of human gene frequencies in Europe.* *Am J Phys Anthropol.* 80(3):267-94.
- Spillantini MG and Goedert M, 2013. Tau pathology and neurodegeneration. *Lancet Neurology* 12, 609 – 22.
- Spillantini MG, Murrell JR, Goedert M, Farlow MR, Klug A, Ghetti B (1998). *Mutation in the tau gene in familial multiple system tauopathy with presenile dementia.* *Proc Natl Acad Sci U S A;* 95:7737–7741.
- Steele, J. & Kandler, A. 2010. Language trees not equal gene trees. *Theory in Biosciences* 129, 223–233.
- Stringer, C. and Davies, W. (2001) Those elusive Neanderthals *Nature* (London) 413, 791–792
- Stefansson H, Helgason A, Thorleifsson G et al 2005 *A common inversion under selection in Europeans.* *Nature Genetics* 37:129-137
- Steinberg KM, Antonacci F et al 2012. *Structural Diversity and African origin of the 17q21.31 inversion polymorphism.* *Nature Genetics* 44, 872 – 880.
- Stewart JR and Stringer CB (2012). *Human evolution Out of Africa: the role of Refugia and Climate Change.* *Science* 335, 1317
- Stewart JR. (2009). *The evolutionary consequence of the individualistic response to climate change.* *J Evol Biol.* 22(12):2363-75
- Sun H, Liu S, Zhang Y, Whittle MR. 2014. Comparison of southern Chinese Han and Brazilian Caucasian mutation rates at autosomal short tandem repeat loci used in human forensic genetics. *Int J Legal Med* 128, 1-9.
- Sündermann F, Fernandez MP, Morgan RO (2016). *An evolutionary roadmap to the microtubule-associated protein MAP Tau.* *BMC Genomics* 17:264
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) *Comparative phylogeography and postglacial colonization routes in Europe.* *Mol Ecol.* (4):453-64.
- Tautz D, Schlötterer C. 1994. Simple sequences. *Curr Opin Genet Dev* 4, 832-837

- Thomas H (1998) The slave trade—the history of the Atlantic slave trade: 1440–1840. Macmillan, London.
- Tobin JE, et al. Haplotypes and gene expression implicate the MAPT region for Parkinson disease: the GenePD Study. *Neurology*. 2008; 71:28–34.
- Torrioni A, Achilli A, Macaulay V, Richards M & Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet*, 22:339–345
- Torrioni A, Bandelt H, Dúrbano L, Lahermo P, Moral P, Sellito D, Rengo C, Forster P, Savontaus M, Bonn -Tamir B & Scozzari R (1998) mtDNA analysis reveals a major late Paleolithic population expansion from Southwestern to Northeastern Europe. *Am J Hum Genet*, 62:1137–1152
- Torrioni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk HV, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus ML, Huoponen K, Laitinen V, Koivum ki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O & Scozzari R (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet*, 69:844–852
- Torsten G nthera, Cristina Valdiosera, Helena Malmstr ma, Irene Ure a, Ricardo Rodr guez-Varela,  ddny Osk Sverrisd ttira, Evangelia A. Daskalakis, Pontus Skoglund, Thijessen Naidoo, Emma M. Svensson, Jos  Mar a Berm dez de Castro, Eudald Carbonell, Michael Dunn, Jan Stor , Eneko Iriarte, Juan Luis Arsuaga, Jos -Miguel Carretero, Anders G therstr me, and Mattias Jakobsson (2015) Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *PNAS* | September 22, 2015 | vol. 112 | no. 38 | 11917–11922
- Tuzun E, et al. Fine-scale structural variation of the human genome. *Nat Genet*. 2005; 37:727–32.
- United Nations, Department of Economic and Social Affairs, Population Division (2017). *World Population Prospects: The 2017 Revision*, custom data acquired via website.
- Valdiosera C., G nther T., Vera-Rodr guez J. C., Ure a I., Iriarte E., Rodr guez-Varela R., ... & Jakobsson M. (2018). Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proceedings of the National Academy of Sciences of the United States of America*, 115(13), 3428–3433.
- Vander Linden, M. (2007). What linked the Bell Beakers in third millennium BC Europe? *Antiquity*, 81(312), 343–352.
- Vennemann T, 2003. Europa Vasconica-Europa Semitica. *Lingua*, 116
- Vernot B, Akey JM. (2014) *Resurrecting Surviving Neandertal Lineages from Modern Human Genomes*. *Science* Vol. 343, Issue 6174, pp. 1017–1021
- Verpoorte A. *Limiting factors on early modern human dispersals: the human biogeography of late Pleniglacial Europe*. *Quatern. Int.* 2009; **201**: 77–85
- Vulto-van Silfhout AT, Hehir-Kwa JY, van Bon BW et al: Clinical significance of de novo and inherited copy-number variation. *Hum Mutat* 2013; 34: 1679–1687.
- Weber, J.L. and Wong, C. (1993) Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2, 1123–1128
- Wegmann S, Medalsy ID, Mandelkow E, M ller DJ. The fuzzy coat of pathological human tau fibrils is a two-layered polyelectrolyte brush. *Proc Natl Acad Sci USA* 2013; 110: 313–21.

- Weingarten MD, Lockwood AH, Hwo SY and Kirschner MW. 1975. A protein factor essential for microtubule assembly. *Proc. Nat. Acad. Sci. USA*. Vol 72, n5, 1858 – 1862.
- Whiteheart SW, Schraw T, Matveeva EA. 2001 N-ethylmaleimide sensitive factor (NSF) structure and function. *Int Rev Cytol*. 207:71-112.
- Wiik (2008) Where did European men come from? *J Genetic Genealogy*, 4:35-85
- Wischik CM, Novak M, Thogersen HC, et al. Isolation of a fragment of tau derived from the core of the paired helical filament of Alzheimer disease. *Proc Natl Acad Sci USA* 1988; 85: 4506–10.
- Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*. 329:75–8.
- Yoshida H, Ihara Y. Tau in paired helical filaments is functionally distinct from fetal tau: assembly incompetence of paired helical filament-tau. *J Neurochem* 1993; 61: 1183–86.
- Zhivotovsky LA, Bennett L, Bowcock AM, Feldman MW. 2000. Human population expansion and microsatellite variation. *Mol Biol Evol* 17, 757-767.
- Zhivotovsky LA, Rosenberg NA, Feldman MW. 2003. Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am J Hum Genet* 72, 1171-1186.
- Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 74, 50-61.
- Zilhao J (2001) Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proc Natl Acad Sci*, 98:14180-14185
- Zlojutro, M., Roy, R., Palikij, J., and Crawford, M.H. (2006). Autosomal STR variation in a Basque population: Vizcaya Province. *Hum. Biol.* 78, 599–618.
- Zody MC, Jiang Z, Fung HC et al 2008 *Evolutionary toggling of the MAPT 17q21.31 inversion region*. *Nature Genetics* 40, 1076-1083
- Zvelebil M & Zvelebil KV (1988) *Agricultural transition and Indo-European dispersals*. *Antiquity*, 62:574-583



# MATERIAL ANEXO

---



**Tabla A1.** Frecuencias de los haplotipos H1', H2' y H2D en diferentes poblaciones.

(Sub) Continente	País / Región	Población	N	2N	Frecuencias			Referencia
					H1'	H2'	H2D	
Africa	Cameroon	Bakola	19		0,947	0,053	0	Steinberg 2012
Africa	Kenya	BantuNE	11		1	0	0	Steinberg 2012
Africa	South Africa	BantuS	8		1	0	0	Steinberg 2012
Africa	Sudan	Beja	18		0,861	0,111	0,028	Steinberg 2012
Africa	Central African Republic	Biaka	23		0,935	0,044	0,022	Steinberg 2012
Africa	Central African Republic SW corner	Biaka2		138/130	0,985	0,015	0	Alfred*
Africa	Kenya	Boni	18		0,921	0	0,079	Steinberg 2012
Africa	Kenya	Borana	19		0,974	0	0,026	Steinberg 2012
Africa	Chad	Bulala	15		1	0	0	Steinberg 2012
Africa	Dar-es-Salaam, Tanzania	Chagga		90	0,989	0,001	0,01	Alfred*
Africa	Tanzania	Datog	18		0,917	0,056	0,028	Steinberg 2012
<b>Africa</b>	<b>Nigeria</b>	<b>Esan</b>	<b>198</b>		<b>1</b>	<b>0</b>	<b>0</b>	
Africa	Cameroon	Fulani	19		1	0	0	Steinberg 2012
<b>Africa</b>	<b>Gambia</b>	<b>Gambia</b>	<b>226</b>		<b>0,987</b>	<b>0,009</b>	<b>0,004</b>	
Africa	Tanzania	Hadza	36		0,972	0,028	0	Steinberg 2012
Africa	Zaria, North Central Nigeria	Hausa		78/70	0,986	0,014	0	Alfred*
Africa	Enugu, SE Nigeria	Ibo		96/94	0,979	0,021	0	Alfred*
Africa	Tanzania	Iraqw	18		0,972	0	0,028	Steinberg 2012
Africa	Cameroon	Lemande	19		1	0	0	Steinberg 2012
Africa	Kenya	Luhya1	101		1	0	0	Steinberg 2012
<b>Africa</b>	<b>Kenia</b>	<b>Luhya2</b>	<b>202</b>		<b>0,98</b>	<b>0,02</b>	<b>0</b>	
Africa	Kenya	Luo	19		1	0	0	Steinberg 2012
Africa	Senegal	Mandenka	22		0,977	0	0,023	Steinberg 2012
Africa	Cameroon	Mandera	18		1	0	0	Steinberg 2012
Africa	Kenya	Masai	143		0,927	0,025	0,049	Steinberg 2012
Africa	close to the Tanzanian-Kenyan border	Masai		40/38	1	0	0	Alfred*
Africa	Democratic Republic of Congo	Mbuti	13		0,923	0,077	0	Steinberg 2012
Africa	Democratic Republic of the Congo, NE Ituri Forest	Mbuti2		78	0,949	0,051	0	Alfred*
<b>Africa</b>	<b>Sierra Leona</b>	<b>Mende</b>	<b>170</b>		<b>0,994</b>	<b>0,006</b>	<b>0</b>	
Africa	Namibia	San	40		0,975	0,025	0	Steinberg 2012
Africa	Tanzania	Sandawe	46		0,913	0,054	0,033	Steinberg 2012
Africa	Kenya	Sengwer	18		0,917	0,083	0	Steinberg 2012

## Haplotipos MAPT en *Homo sapiens*

Africa	Nigeria	Yoruba1	157	1	0	0	Steinberg 2012
Africa	Benin City, Nigeria	Yoruba2	156/154	0,994	0,007	0	Alfred*
<b>Africa</b>	<b>Nigeria?</b>	<b>Yoruba3</b>	<b>218</b>	<b>0,995</b>	<b>0,005</b>	<b>0</b>	
<b>Africa</b>	<b>Barbados</b>	<b>Afroamer1</b>	<b>192</b>	<b>0,953</b>	<b>0,026</b>	<b>0,021</b>	
<b>Africa</b>	<b>USA</b>	<b>Afroamer2</b>	<b>132</b>	<b>0,879</b>	<b>0,038</b>	<b>0,083</b>	
Africa	USA	Afroamer3	180/182	0,929	0,031	0,04	Alfred*
<b>Africa</b>	<b>Colombia</b>	<b>Chocó</b>	<b>35</b>	<b>0,943</b>	<b>0,028</b>	<b>0,029</b>	<b>Present study</b>
Europe	Russia	Adygei	108	0,843	0,077	0,08	Alfred*
Europe	France	BasqueF	48/166	0,645	0,105	0,25	Alfred*
Europe	southern Ural Mountains region of Russia	Chuvash	84	0,893	0,007	0,1	Alfred*
Europe	Dinamarca	Danes	98/102	0,843	0,017	0,14	Alfred*
<b>Europe</b>		<b>España</b>	<b>212</b>	<b>0,736</b>	<b>0</b>	<b>0,264</b>	
<b>Europe</b>		<b>Finlandia</b>	<b>196</b>	<b>0,893</b>	<b>0</b>	<b>0,107</b>	
<b>Europe</b>	<b>Spain</b>	<b>GitanosPV</b>	<b>68</b>	<b>0,875</b>	<b>0</b>	<b>0,125</b>	<b>Present study</b>
<b>Europe</b>	<b>Spain</b>	<b>Guipuzcoa</b>	<b>91</b>	<b>0,5</b>	<b>0,006</b>	<b>0,494</b>	<b>Present study</b>
Europe	Ireland	Irish	228	0,825	0,005	0,17	Alfred*
<b>Europe</b>	<b>Spain</b>	<b>Navarra</b>	<b>92</b>	<b>0,69</b>	<b>0,027</b>	<b>0,283</b>	<b>Present study</b>
Europe	Vologda, Russia	Russia1	96	0,906	0,004	0,09	Alfred*
Europe	SW Arkhangel,N Russia	Russia2	66	0,924	0,026	0,05	Alfred*
Europe	Italy	Sardinian	56/70	0,643	0,127	0,23	Alfred*
<b>Europe</b>	<b>Italy</b>	<b>Toscana</b>	<b>214</b>	<b>0,636</b>	<b>0,037</b>	<b>0,327</b>	
<b>Europe</b>		<b>UK</b>	<b>184</b>	<b>0,755</b>	<b>0,022</b>	<b>0,223</b>	
<b>Europe</b>	<b>Spain</b>	<b>Valencia</b>	<b>77</b>	<b>0,74</b>	<b>0,026</b>	<b>0,234</b>	<b>Present study</b>
<b>Europe</b>	<b>USA</b>	<b>ResidentesUSA1</b>	<b>198</b>	<b>0,793</b>	<b>0,01</b>	<b>0,197</b>	
Europe	USA	ResidentesUSA2	182/170	0,747	0,023	0,23	Alfred*
<b>Europe</b>	<b>Spain</b>	<b>ResidentesPV</b>	<b>23</b>	<b>0,717</b>	<b>0,022</b>	<b>0,261</b>	<b>Present study</b>
WAsia	Israel	Ashkenazi	162	0,735	0,025	0,24	Alfred*
WAsia	Northern Israel	Druze	254/204	0,691	0,009	0,3	Alfred*
WAsia	Israel	Samaritans	78/80	0,638	0,012	0,35	Alfred*
<b>CAsia</b>	<b>Bengalí</b>	<b>Bangladesh</b>	<b>170</b>	<b>0,953</b>	<b>0,006</b>	<b>0,041</b>	
<b>CAsia</b>	<b>USA</b>	<b>Gujarati</b>	<b>210</b>	<b>0,91</b>	<b>0,01</b>	<b>0,081</b>	
CAsia	Pakistan	Hazara	48/204	0,931	0,069	0	Alfred*
CAsia	E Ural Montains, Russia	Khanty	100/96	0,969	0,001	0,03	Alfred*
CAsia	Kazak, near Ural M, Russia	Komi-Zyrian	94	0,968	0,022	0,01	Alfred*
<b>CAsia</b>		<b>Punjab</b>	<b>192</b>	<b>0,901</b>	<b>0,031</b>	<b>0,068</b>	



<b>CAsia</b>	<b>UK</b>	<b>Tamiles</b>	<b>206</b>	<b>0,971</b>	<b>0,005</b>	<b>0,024</b>	
<b>CAsia</b>	<b>UK</b>	<b>Telugu</b>	<b>206</b>	<b>0,951</b>	<b>0,01</b>	<b>0,039</b>	
CAsia	Rusia	Yakut	102/100	0,98	0	0,02	Alfred*
EastAsia	Taiwan	Ami	80/78	1	0	0	Alfred*
EastAsia	Taiwan	Atayal	84/82	1	0	0	Alfred*
<b>EastAsia</b>	<b>Chinos</b>	<b>Dai</b>	<b>198</b>	<b>1</b>	<b>0</b>	<b>0</b>	
EastAsia	Taiwan	Han	100/94	0,989	0,001	0,01	Alfred*
EastAsia	USA	Han	116/110	1	0	0	Alfred*
<b>EastAsia</b>	<b>Beijing</b>	<b>Han</b>	<b>206</b>	<b>1</b>	<b>0</b>	<b>0</b>	
<b>EastAsia</b>	<b>S. China</b>	<b>Han</b>	<b>216</b>	<b>1</b>	<b>0</b>	<b>0</b>	
EastAsia	USA	Japanese	100	0,99	0	0,01	Alfred*
<b>EastAsia</b>		<b>Japoneses</b>	<b>208</b>	<b>0,995</b>	<b>0</b>	<b>0,005</b>	
<b>EastAsia</b>	<b>Vietnam</b>	<b>Kinh</b>	<b>202</b>	<b>1</b>	<b>0</b>	<b>0</b>	
EastAsia	Korea	Koreans	108	1	0	0	Alfred*
Oceania	Micronesia	Micronesians	72/74	0,973	0,017	0,01	Alfred*
Oceania	Bougainville Island, Solomon Islands	Nasioi	44/46	1	0	0	Alfred*
Oceania	New Guinea	PapNewGuinean	34/40	1	0	0	Alfred*
NorthAmerica	Western Oklahoma	Cheyenne	112/110	0,982	0,018	0	Alfred*
NorthAmerica	Phoenix, Arizona, EEUU	Pima	102/100	1	0	0	Alfred*
NorthAmerica	Campeche, central Yucatan, Mexico	Maya	50/100	0,94	0,02	0,04	Alfred*
<b>NorthAmerica</b>	<b>California</b>	<b>Mejicanos</b>	<b>134</b>	<b>0,828</b>	<b>0,015</b>	<b>0,157</b>	
SouthAmerica	Rondonia Province, Brazil	Karitiana	110	1	0	0	Alfred*
SouthAmerica	Remote Andes, Peru	Quechua	46	0,935	0,025	0,04	Alfred*
SouthAmerica	W Amazonas, Brasil	Ticuna	130/122	1	0	0	Alfred*
<b>SouthAmerica</b>	<b>Ecuador</b>	<b>Waoranis</b>	<b>24</b>	<b>1</b>	<b>0</b>	<b>0</b>	Present study
<b>SouthAmerica</b>	<b>Argentina</b>	<b>Jujuy</b>	<b>68</b>	<b>0,993</b>	<b>0</b>	<b>0,007</b>	Present study
	<b>SouthAmerica</b>	<b>Peruanos</b>	<b>172</b>	<b>0,93</b>	<b>0</b>	<b>0,07</b>	
	<b>SouthAmerica</b>	<b>Portorriqueños</b>	<b>208</b>	<b>0,803</b>	<b>0,005</b>	<b>0,192</b>	
<b>SouthAmerica</b>	<b>Medellín</b>	<b>Colombianos</b>	<b>188</b>	<b>0,819</b>	<b>0,005</b>	<b>0,176</b>	

**Tabla A2a.** Frecuencias de los haplotipos extendidos (Indel, microsatélites y SNP rs199451) hallados en las ocho poblaciones analizadas.

Código	Nº cromosomas	Frecuencia	Haplotipo
1	29	0,0396	1 207 199 125 131 G
2	25	0,0342	2 209 199 109 131 A
3	16	0,0219	1 209 199 125 123 G
4	3	0,0041	1 205 199 125 135 G
5	23	0,0314	2 209 199 109 103 A
6	1	0,0014	1 211 199 125 139 A
7	17	0,0232	2 209 199 109 139 A
8	17	0,0232	1 207 199 125 135 G
9	15	0,0205	2 209 199 109 135 A
10	5	0,0068	2 209 199 109 119 A
11	1	0,0014	1 211 199 123 119 G
12	1	0,0014	1 209 199 123 139 G
13	1	0,0014	1 209 199 123 135 G
14	4	0,0055	1 211 199 125 139 G
15	9	0,0123	1 209 195 125 139 G
16	9	0,0123	1 209 195 125 123 G
17	6	0,0082	2 209 199 109 123 A
18	1	0,0014	1 205 195 125 131 G
19	6	0,0082	1 209 203 125 139 G
20	1	0,0014	2 209 199 109 143 G
21	1	0,0014	1 209 195 125 143 G
22	4	0,0055	2 209 199 125 103 A
23	1	0,0014	1 211 203 125 131 G
24	5	0,0068	2 209 203 109 131 A
25	35	0,0478	1 209 199 125 135 G
26	21	0,0287	1 209 203 125 131 G
27	14	0,0191	1 209 195 125 135 G
28	40	0,0546	1 207 195 125 135 G
29	2	0,0027	1 209 203 125 115 G
30	19	0,0260	1 207 195 125 119 G
31	1	0,0014	2 219 199 109 139 A
32	1	0,0014	1 207 199 125 115 G
33	7	0,0096	1 203 199 125 131 G

---

34	5	0,0068	2 209 203 109 103 A
35	23	0,0314	1 209 195 125 131 G
36	1	0,0014	2 213 199 109 131 A
37	17	0,0232	1 207 195 125 131 G
38	2	0,0027	2 205 199 109 135 A
39	6	0,0082	1 203 199 125 135 G
40	4	0,0055	1 203 195 125 131 G
41	15	0,0205	1 207 195 125 127 G
42	6	0,0082	1 209 199 125 139 G
43	3	0,0041	1 207 195 125 139 G
44	7	0,0096	1 205 199 125 131 G
45	12	0,0164	2 209 199 109 127 A
46	2	0,0027	1 205 199 125 127 G
47	49	0,0669	1 207 195 125 123 G
48	9	0,0123	1 211 199 125 131 G
49	1	0,0014	2 209 195 109 135 A
50	3	0,0041	1 211 195 125 135 G
51	1	0,0014	2 209 199 125 123 A
52	1	0,0014	1 209 203 125 123 G
53	19	0,0260	1 209 199 125 127 G
54	1	0,0014	1 209 203 109 123 G
55	11	0,0150	1 209 199 125 119 G
56	1	0,0014	2 209 199 109 99 A
57	6	0,0082	1 207 199 125 119 G
58	6	0,0082	1 209 195 125 119 G
59	12	0,0164	1 207 199 125 127 G
60	4	0,0055	2 209 203 109 123 A
61	2	0,0027	2 205 199 109 131 A
62	4	0,0055	1 209 195 125 127 G
63	1	0,0014	2 213 199 109 103 A
64	1	0,0014	2 211 199 109 103 A
65	9	0,0123	1 207 199 125 139 G
66	22	0,0301	1 209 199 125 131 G
67	1	0,0014	2 209 203 107 103 A
68	1	0,0014	2 209 199 109 139 G
69	1	0,0014	2 209 199 109 131 G
70	4	0,0055	1 209 203 125 119 G
71	3	0,0041	1 213 199 125 119 G
72	1	0,0014	1 207 199 127 123 G

---

## Haplotipos MAPT en *Homo sapiens*

---

73	1	0,0014	2 207 195 125 143 G
74	2	0,0027	2 209 195 125 135 G
75	1	0,0014	1 209 199 109 131 A
76	1	0,0014	1 213 199 125 147 G
77	5	0,0068	1 211 199 125 123 G
78	1	0,0014	2 207 195 109 131 A
79	1	0,0014	1 207 203 125 139 G
80	1	0,0014	2 207 199 125 131 G
81	3	0,0041	2 209 199 107 131 A
82	1	0,0014	2 207 199 107 139 A
83	1	0,0014	1 205 199 125 119 G
84	1	0,0014	1 207 195 125 127 A
85	1	0,0014	1 211 195 125 131 G
86	1	0,0014	2 215 195 125 119 G
87	2	0,0027	2 207 195 125 123 G
88	15	0,0205	1 207 199 125 123 G
89	1	0,0014	1 207 195 123 155 G
90	1	0,0014	1 207 195 125 151 G
91	1	0,0014	2 209 195 109 123 A
92	1	0,0014	1 215 199 125 119 G
93	4	0,0055	1 217 199 125 119 G
94	1	0,0014	1 207 203 125 119 G
95	1	0,0014	1 211 199 125 135 G
96	1	0,0014	1 203 203 125 139 G
97	1	0,0014	2 211 199 109 127 A
98	2	0,0027	1 203 199 125 123 G
99	1	0,0014	1 207 199 125 155 G
100	1	0,0014	1 201 199 125 127 G
101	1	0,0014	1 217 195 125 119 G
102	1	0,0014	1 213 207 125 119 G
103	1	0,0014	1 207 191 125 127 G
104	3	0,0041	1 207 195 125 115 G
105	1	0,0014	1 205 195 125 115 G
106	1	0,0014	1 207 191 125 131 G
107	1	0,0014	1 211 199 123 135 G
108	1	0,0014	1 211 195 123 123 G
109	1	0,0014	1 203 203 129 123 G
110	1	0,0014	1 205 191 125 131 G
111	1	0,0014	1 217 199 125 123 G

---

112	1	0,0014	1 217 195 125 127 G
113	1	0,0014	1 211 199 125 143 G
114	2	0,0027	1 213 199 125 143 G
115	1	0,0014	1 211 195 125 115 G
116	2	0,0027	2 209 199 109 143 A
117	1	0,0014	1 209 199 125 143 G
118	2	0,0027	1 213 195 125 131 G
119	1	0,0014	1 217 195 123 127 G
120	1	0,0014	1 205 195 125 127 G
121	1	0,0014	2 205 199 109 123 G
122	2	0,0027	1 213 199 125 135 G
123	1	0,0014	1 213 199 125 139 G
124	7	0,0096	1 213 199 125 131 G
125	7	0,0096	1 209 191 125 131 G
126	1	0,0014	2 209 191 125 131 G
127	1	0,0014	1 209 203 125 151 G
128	1	0,0014	1 209 195 125 151 G
129	1	0,0014	1 215 199 109 119 G
130	1	0,0014	2 209 195 125 131 G
131	1	0,0014	1 209 199 109 139 G
132	3	0,0041	1 213 199 125 127 G
133	1	0,0014	1 213 199 125 123 G
134	1	0,0014	2 205 191 109 103 A
135	1	0,0014	1 207 203 125 143 G
136	2	0,0027	1 207 195 125 143 G
137	1	0,0014	2 209 199 109 103 G
138	1	0,0014	1 211 195 127 123 G
139	1	0,0014	1 209 187 125 131 G

---

**Tabla A2b.** Frecuencias de los haplotipos extendidos (Indel, microsatélites y SNP rs199451) hallados en las ocho poblaciones analizadas desglosadas por población.

Código	Frecuencias								Haplotipo
	Gipuzkoa	Nafarroa	Valencia	Gitanos Vascos	Chocó	Waoranis	Jujuy	Residentes	
1	0,0811	0,0750	0	0,0441	0	0	0,0172	0	1 207 199 125 131 G
2	0,1081	0,0417	0,0313	0,0074	0	0	0	0,0250	2 209 199 109 131 A
3	0,0270	0,0333	0,0313	0	0	0,0238	0,0259	0,0500	1 209 199 125 123 G
4	0,0068	0,0083	0,0156	0	0	0	0	0	1 205 199 125 135 G
5	0,1014	0,0333	0	0,0074	0	0	0	0,0750	2 209 199 109 103 A
6	0,0068	0	0	0	0	0	0	0	1 211 199 125 139 A
7	0,0743	0,0167	0,0156	0,0147	0	0	0	0,0250	2 209 199 109 139 A
8	0,0135	0,0333	0	0,0294	0	0	0,0603	0	1 207 199 125 135 G
9	0,0338	0,0417	0,0156	0,0221	0,0152	0	0	0	2 209 199 109 135 A
10	0,0068	0,0167	0,0156	0,0074	0	0	0	0	2 209 199 109 119 A
11	0,0068	0	0	0	0	0	0	0	1 211 199 123 119 G
12	0,0068	0	0	0	0	0	0	0	1 209 199 123 139 G
13	0,0068	0	0	0	0	0	0	0	1 209 199 123 135 G
14	0,0068	0,0083	0	0	0,0152	0	0,0086	0	1 211 199 125 139 G
15	0,0203	0,0167	0	0	0	0	0,0345	0	1 209 195 125 139 G
16	0,0068	0,0083	0,0156	0	0,0152	0,0952	0,0086	0	1 209 195 125 123 G
17	0,0135	0,0167	0	0	0	0	0	0,0500	2 209 199 109 123 A
18	0,0068	0	0	0	0	0	0	0	1 205 195 125 131 G
19	0,0068	0,0167	0,0313	0	0	0	0	0,0250	1 209 203 125 139 G
20	0,0068	0	0	0	0	0	0	0	2 209 199 109 143 G
21	0,0068	0	0	0	0	0	0	0	1 209 195 125 143 G
22	0,0135	0	0	0,0147	0	0	0	0	2 209 199 125 103 A
23	0,0068	0	0	0	0	0	0	0	1 211 203 125 131 G
24	0,0135	0,0167	0,0156	0	0	0	0	0	2 209 203 109 131 A
25	0,0203	0,0667	0,0625	0,0368	0,0152	0,0476	0,0776	0,0750	1 209 199 125 135 G
26	0,0270	0,0750	0,0313	0,0147	0,0152	0,0476	0	0,0250	1 209 203 125 131 G
27	0,0270	0,0083	0	0,0368	0,0455	0	0,0086	0	1 209 195 125 135 G
28	0,0135	0,0583	0,0625	0,1324	0,0455	0,0476	0,0259	0,0250	1 207 195 125 135 G
29	0,0135	0	0	0	0	0	0	0	1 209 203 125 115 G
30	0,0068	0,0083	0,0313	0,0441	0,0606	0	0,0345	0,0250	1 207 195 125 119 G
31	0,0068	0	0	0	0	0	0	0	2 219 199 109 139 A
32	0,0068	0	0	0	0	0	0	0	1 207 199 125 115 G

33	0,0135	0,0083	0	0,0147	0,0152	0	0,0086	0	1 203 199 125 131 G
34	0,0135	0,0083	0	0,0074	0	0	0	0,0250	2 209 203 109 103 A
35	0,0135	0,0250	0,0156	0,0221	0,0606	0,0476	0,0690	0	1 209 195 125 131 G
36	0,0068	0	0	0	0	0	0	0	2 213 199 109 131 A
37	0,0135	0,0167	0,0313	0	0,0303	0	0,0431	0,1	1 207 195 125 131 G
38	0,0135	0	0	0	0	0	0	0	2 205 199 109 135 A
39	0,0135	0,0083	0	0,0147	0	0	0	0,0250	1 203 199 125 135 G
40	0,0068	0,0083	0	0	0,0303	0	0	0	1 203 195 125 131 G
41	0,0270	0,0083	0	0,0588	0	0	0,0172	0	1 207 195 125 127 G
42	0,0068	0,0167	0	0	0,0303	0,0238	0	0	1 209 199 125 139 G
43	0,0068	0	0,0156	0,0074	0	0	0	0	1 207 195 125 139 G
44	0,0068	0	0	0	0,0152	0	0,0345	0,0250	1 205 199 125 131 G
45	0,0203	0,0167	0,0313	0,0294	0	0	0	0,0250	2 209 199 109 127 A
46	0,0068	0	0,0156	0	0	0	0	0	1 205 199 125 127 G
47	0,0203	0,0250	0,0469	0,1250	0,0606	0,1905	0,0776	0,0500	1 207 195 125 123 G
48	0,0135	0,0167	0,0156	0,0221	0	0	0,0086	0	1 211 199 125 131 G
49	0,0068	0	0	0	0	0	0	0	2 209 195 109 135 A
50	0,0068	0	0	0	0,0303	0	0	0	1 211 195 125 135 G
51	0,0068	0	0	0	0	0	0	0	2 209 199 125 123 A
52	0,0068	0	0	0	0	0	0	0	1 209 203 125 123 G
53	0,0135	0,0167	0,0313	0,0588	0,0152	0	0,0259	0,0250	1 209 199 125 127 G
54	0,0068	0	0	0	0	0	0	0	1 209 203 109 123 G
55	0,0068	0,0333	0,0156	0,0074	0	0	0,0259	0,0250	1 209 199 125 119 G
56	0,0068	0	0	0	0	0	0	0	2 209 199 109 99 A
57	0,0068	0,0083	0,0156	0,0147	0	0	0,0086	0	1 207 199 125 119 G
58	0,0068	0,0083	0,0313	0	0,0152	0	0,0086	0	1 209 195 125 119 G
59	0,0068	0,0333	0,0156	0,0074	0,0152	0	0,0345	0	1 207 199 125 127 G
60	0,0068	0,0167	0,0156	0	0	0	0	0	2 209 203 109 123 A
61	0	0,0083	0,0156	0	0	0	0	0	2 205 199 109 131 A
62	0	0,0083	0	0,0074	0,0303	0	0	0	1 209 195 125 127 G
63	0	0,0083	0	0	0	0	0	0	2 213 199 109 103 A
64	0	0,0083	0	0	0	0	0	0	2 211 199 109 103 A
65	0	0,0250	0	0,0147	0	0	0,0172	0,0500	1 207 199 125 139 G
66	0	0,0250	0,0156	0,0074	0,0303	0,1190	0,0776	0,0250	1 209 199 125 131 G
67	0	0,0083	0	0	0	0	0	0	2 209 203 107 103 A
68	0	0,0083	0	0	0	0	0	0	2 209 199 109 139 G
69	0	0,0083	0	0	0	0	0	0	2 209 199 109 131 G
70	0	0,0083	0,0313	0,0074	0	0	0	0	1 209 203 125 119 G
71	0	0,0083	0	0	0,0152	0	0,0086	0	1 213 199 125 119 G

## Haplotipos MAPT en *Homo sapiens*

---

72	0	0	0,0156	0	0	0	0	0	1 207 199 127 123 G
73	0	0	0,0156	0	0	0	0	0	2 207 195 125 143 G
74	0	0	0,0156	0	0	0	0,0086	0	2 209 195 125 135 G
75	0	0	0,0156	0	0	0	0	0	1 209 199 109 131 A
76	0	0	0,0156	0	0	0	0	0	1 213 199 125 147 G
77	0	0	0,0313	0,0074	0,0152	0	0	0,0250	1 211 199 125 123 G
78	0	0	0,0156	0	0	0	0	0	2 207 195 109 131 A
79	0	0	0,0156	0	0	0	0	0	1 207 203 125 139 G
80	0	0	0,0156	0	0	0	0	0	2 207 199 125 131 G
81	0	0	0,0469	0	0	0	0	0	2 209 199 107 131 A
82	0	0	0,0156	0	0	0	0	0	2 207 199 107 139 A
83	0	0	0,0156	0	0	0	0	0	1 205 199 125 119 G
84	0	0	0,0156	0	0	0	0	0	1 207 195 125 127 A
85	0	0	0,0156	0	0	0	0	0	1 211 195 125 131 G
86	0	0	0	0,0074	0	0	0	0	2 215 195 125 119 G
87	0	0	0	0,0074	0,0152	0	0	0	2 207 195 125 123 G
88	0	0	0	0,0147	0	0,0476	0,0948	0	1 207 199 125 123 G
89	0	0	0	0,0074	0	0	0	0	1 207 195 123 155 G
90	0	0	0	0,0074	0	0	0	0	1 207 195 125 151 G
91	0	0	0	0,0074	0	0	0	0	2 209 195 109 123 A
92	0	0	0	0,0074	0	0	0	0	1 215 199 125 119 G
93	0	0	0	0,0294	0	0	0	0	1 217 199 125 119 G
94	0	0	0	0,0074	0	0	0	0	1 207 203 125 119 G
95	0	0	0	0,0074	0	0	0	0	1 211 199 125 135 G
96	0	0	0	0,0074	0	0	0	0	1 203 203 125 139 G
97	0	0	0	0,0074	0	0	0	0	2 211 199 109 127 A
98	0	0	0	0,0074	0,0152	0	0	0	1 203 199 125 123 G
99	0	0	0	0,0074	0	0	0	0	1 207 199 125 155 G
100	0	0	0	0,0074	0	0	0	0	1 201 199 125 127 G
101	0	0	0	0,0074	0	0	0	0	1 217 195 125 119 G
102	0	0	0	0,0074	0	0	0	0	1 213 207 125 119 G
103	0	0	0	0	0,0152	0	0	0	1 207 191 125 127 G
104	0	0	0	0	0,0303	0	0,0086	0	1 207 195 125 115 G
105	0	0	0	0	0,0152	0	0	0	1 205 195 125 115 G
106	0	0	0	0	0,0152	0	0	0	1 207 191 125 131 G
107	0	0	0	0	0,0152	0	0	0	1 211 199 123 135 G
108	0	0	0	0	0,0152	0	0	0	1 211 195 123 123 G
109	0	0	0	0	0,0152	0	0	0	1 203 203 129 123 G
110	0	0	0	0	0,0152	0	0	0	1 205 191 125 131 G



## Material anexo

111	0	0	0	0	0,0152	0	0	0	1 217 199 125 123 G
112	0	0	0	0	0,0152	0	0	0	1 217 195 125 127 G
113	0	0	0	0	0,0152	0	0	0	1 211 199 125 143 G
114	0	0	0	0	0,0152	0	0,0086	0	1 213 199 125 143 G
115	0	0	0	0	0,0152	0	0	0	1 211 195 125 115 G
116	0	0	0	0	0,0152	0	0	0,0250	2 209 199 109 143 A
117	0	0	0	0	0,0152	0	0	0	1 209 199 125 143 G
118	0	0	0	0	0,0152	0	0,0086	0	1 213 195 125 131 G
119	0	0	0	0	0,0152	0	0	0	1 217 195 123 127 G
120	0	0	0	0	0,0152	0	0	0	1 205 195 125 127 G
121	0	0	0	0	0,0152	0	0	0	2 205 199 109 123 G
122	0	0	0	0	0,0152	0	0,0086	0	1 213 199 125 135 G
123	0	0	0	0	0,0152	0	0	0	1 213 199 125 139 G
124	0	0	0	0	0	0,0714	0,0345	0	1 213 199 125 131 G
125	0	0	0	0	0	0,1667	0	0	1 209 191 125 131 G
126	0	0	0	0	0	0,0238	0	0	2 209 191 125 131 G
127	0	0	0	0	0	0,0238	0	0	1 209 203 125 151 G
128	0	0	0	0	0	0,0238	0	0	1 209 195 125 151 G
129	0	0	0	0	0	0	0,0086	0	1 215 199 109 119 G
130	0	0	0	0	0	0	0,0086	0	2 209 195 125 131 G
131	0	0	0	0	0	0	0,0086	0	1 209 199 109 139 G
132	0	0	0	0	0	0	0,0259	0	1 213 199 125 127 G
133	0	0	0	0	0	0	0,0086	0	1 213 199 125 123 G
134	0	0	0	0	0	0	0	0,0250	2 205 191 109 103 A
135	0	0	0	0	0	0	0	0,0250	1 207 203 125 143 G
136	0	0	0	0	0	0	0	0,0500	1 207 195 125 143 G
137	0	0	0	0	0	0	0	0,0250	2 209 199 109 103 G
138	0	0	0	0	0	0	0	0,0250	1 211 195 127 123 G
139	0	0	0	0	0	0	0	0,0250	1 209 187 125 131 G



