

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea



ZTF-FCT
Zientzia eta Teknologia Fakultatea
Facultad de Ciencia y Tecnología



Trabajo Fin de Grado
Grado en Biología

Inserciones *Alu* con alto poder discriminante en el norte de Navarra

Autora:
Lara Santiago Sánchez
Director:
Jose Angel Peña García

ÍNDICE

| | Página |
|---------------------------------------|--------|
| Resumen | 2 |
| Abstract..... | 2 |
| Introducción..... | 3 |
| Objetivos..... | 5 |
| Material y métodos | 5 |
| Poblaciones de estudio..... | 5 |
| Marcadores utilizados | 6 |
| Identificación de los genotipos | 6 |
| Tratamientos estadísticos..... | 7 |
| Resultados..... | 10 |
| Discusión | 18 |
| Conclusiones..... | 19 |
| Bibliografía | 19 |

RESUMEN

En este trabajo se han analizado 4 inserciones *Alu* (*Ya5DP3*, *Ya5NBC37*, *YD3JX437* e *Ya5a2DPI*), todas pertenecientes al cromosoma X, en una muestra de 30 individuos originarios del norte de Navarra. La finalidad de dicho análisis ha sido valorar el grado de heterogeneidad genética que presenta la población de Navarra frente a poblaciones de diferentes continentes y comprobar el posible origen paleonorteafricano del patrimonio genético vasco.

Las inserciones *Alu* polimórficas permiten conocer el estado ancestral del marcador (caracterizado por la ausencia de inserción) y poseen una tasa mutacional prácticamente nula; además, todas las copias de una inserción son idénticas por descendencia. Todo ello las convierte en unos marcadores genéticos muy útiles en estudios de evolución humana. Así pues, son interesantes para analizar la heterogeneidad genética entre poblaciones con diferentes orígenes.

En general, el Análisis Molecular de la Varianza muestra unos valores similares a lo esperado para el conjunto del genoma humano, reflejando una heterogeneidad significativa entre continentes para 3 de las 4 inserciones *Alu* estudiadas. Por otra parte, los resultados de los análisis *Neighbour-Joining* y de escalamiento multidimensional revelan unas relaciones entre las poblaciones que encajan con la teoría del *Out of Africa* y muestran una clara distinción entre las poblaciones bereberes y vascas. Esta heterogeneidad genética no se debe a la selección, sino a alteraciones en el equilibrio entre el flujo génico y la deriva que acontecieron durante el proceso de colonización de los diferentes continentes.

ABSTRACT

In this paper, 4 *Alu* insertions (*Ya5DP3*, *Ya5NBC37*, *YD3JX437* and *Ya5a2DPI*), all belonging to the X chromosome, has been analyzed in a sample of 30 people from the north of Navarre. The aim of this analysis has been to value the degree of genetic heterogeneity that the population of Navarre presents with populations from different continents and to verify the possible paleo-North african origin of the Basque genetic pool.

The polymorphic *Alu* insertions allow to know the ancestral state of the marker (characterized by the absence of insertion) and possess a mutational rate practically null; in addition, all copies of an insertion are identical by descent. All this convert them in robust genetic markers for human evolutionary studies. Therefore, they are particularly interesting for analyze the genetic diversity of populations originated in different continents.

In general, the Molecular Analysis of Variance shows similar values to the expected for the whole human genome, reflecting a significant heterogeneity between continents for 3 of the 4 *Alu* insertions studied. On the other hand, the results of the *Neighbour-Joining* and multidimensional scaling analyzes reveal relationships between populations that fit the *Out of Africa* theory and show a clear distinction between Berber and Basque populations. This genetic heterogeneity is not due to selection, but to alterations in the balance between gene flow and drift that occurred during the colonization process of the different continents.

INTRODUCCIÓN

El uso de la reacción en cadena de la polimerasa (PCR) ha llevado, en los últimos años, a un aumento en las posibilidades de identificación de marcadores genéticos, lo que a su vez ha facilitado el estudio de los procesos de microdiferenciación en las poblaciones humanas a lo largo de su historia evolutiva. Entre estos marcadores están los SINE (de *Short INterspaced Elements*), los cuales engloban a una serie de secuencias menores de 500 pb que se encuentran distribuidas de forma dispersa por el genoma. Los SINEs más abundantes son las inserciones *Alu*, las cuales componen la mayor familia de elementos móviles presentes en el genoma humano (Batzer y Deininger, 2002).

Los elementos *Alu* representan alrededor del 10% del genoma humano, con aproximadamente 1.400.000 copias distribuidas por todo el genoma. Estructuralmente, se trata de dímeros de unos 300 pb constituidos por dos unidades similares (pero no idénticas) terminadas ambas en una secuencia de adeninas (dA) (Weiner *et al.*, 1986) (Figura 1). Se cree que las secuencias *Alu* se originaron por una retrotransposición del gen ARN 7SL (Batzer y Deininger, 2002), el cual forma parte del complejo ribosómico (Ullu y Tschudi, 1984).

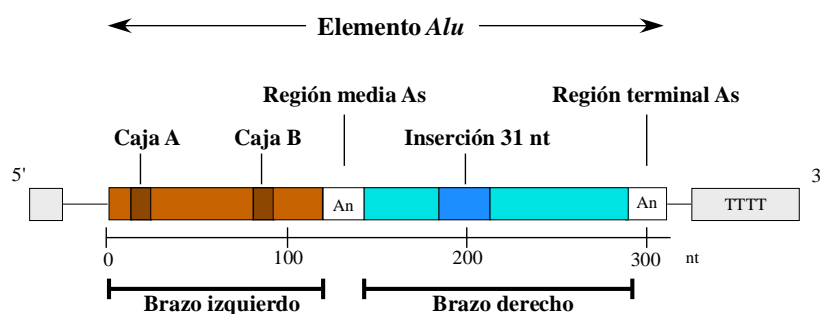


Figura 1: Esquema de una inserción *Alu*.

En la evolución de los elementos *Alu* se distinguen dos etapas bien diferenciadas: la monomérica, que es la más antigua y que concluye con el origen de las secuencias progenitoras de la familia *Alu*, y otra más reciente donde se expanden las secuencias diméricas de *Alu*. La etapa más antigua comienza con la aparición del elemento *Alu* ancestral, conocido como monómero fósil *Alu* o FAM (del inglés *Fossil Alu Monomer*), probablemente durante la radiación de los mamíferos (Labuda y Zietkiewicz, 1994; Roy-Engel *et al.*, 2008). La diversificación y consecuente evolución de la familia FAM originó las familias de monómeros FLAM (del inglés *Free Left Alu Monomer*) y FRAM (de *Free Right Alu Monomer*). La hipótesis más manejada al respecto de la evolución de las inserciones *Alu* sostiene que el primer progenitor de la familia de inserciones *Alu* diméricas fue consecuencia de la fusión de los monómeros FLAM y FRAM hace unos 65 millones de años (Quentin, 1992; Mighell *et al.*, 1997; Batzer y Deininger, 2002; Roy-Engel *et al.*, 2008), fecha que coincide con la radiación de los primates, de manera que estas inserciones sólo están presentes en el Orden Primates.

Las inserciones *Alu* son idóneas para estudios genético-evolutivos en poblaciones humanas. Ello se debe a tres características: a) la posibilidad de conocer el estado ancestral del marcador, caracterizado por la ausencia de la inserción, b) la probabilidad de que se produzcan dos inserciones de un elemento *Alu*

en el mismo lugar cromosómico es extremadamente baja o virtualmente inexistente, de tal modo que estas inserciones constituyen fenómenos prácticamente irrepetibles en la historia evolutiva humana y c) como consecuencia, todas las inserciones encontradas en el mismo lugar del genoma son idénticas por naturaleza y por descendencia, lo cual implica la existencia de antepasados comunes (Batzer y Deininger, 1991; Batzer *et al.*, 1994; Lander *et al.*, 2001).

La población analizada en este estudio ha sido la población autóctona del norte de Navarra. La Comunidad Foral de Navarra, cuya capital es Pamplona, está localizada al norte de la Península Ibérica. Ocupa un área de 10.421 km² y sus fronteras administrativas limitan con el departamento francés de Pirineos Atlánticos al norte, las provincias aragonesas de Huesca y Zaragoza al este, La Rioja y nuevamente Zaragoza al sur y las provincias de Álava y Gipuzkoa al oeste (Figura 2). Navarra posee una población de aproximadamente 621.976 habitantes (INE, 2011), un tercio de los cuales vive en la capital, mientras que en torno al 10% lo hace en la franja septentrional de la provincia.

Desde el punto de vista lingüístico, en Navarra se distinguen tres zonas: la zona vascofona, situada en la frontera con la provincia de Gipuzkoa, en el extremo norte de la provincia; una zona mixta donde conviven el euskera y el castellano, predominando este último, y una zona no vascofona, que ocupa algo más de la mitad sur de la provincia (Gobierno Vasco, 2008). Para poder comprender correctamente la influencia del euskera, hay que tener en cuenta que algo más de la mitad de los habitantes de Navarra viven en la zona mixta, algo más de un tercio en la zona no vascofona y solamente una décima parte en la zona vascofona.

Como consecuencia del grado variable de persistencia del euskera en la región, puede considerarse que ha existido una barrera lingüística a la panmixia en el norte de Navarra, (García-Obregón *et al.*, 2007). Por otra parte, mientras que en el resto del País Vasco la Revolución Industrial comenzó a finales del siglo XIX, en Navarra no ocurrió hasta la década de 1960. La zona norte de Navarra es muy montañosa, lo que unido a la barrera lingüística y la tardía industrialización, pudo dificultar el flujo génico desde otras regiones próximas y la consiguiente entrada de diversidad genética foránea hasta finales del siglo XX (Alfonso-Sánchez *et al.*, 2005).

Por otro lado, es interesante mencionar que algunos autores han sugerido un origen paleonorteafricano del patrimonio genético vasco, de modo que se ha llegado a postular un origen pre-neolítico común con bereberes, marroquíes, argelinos y egipcios (Arnaiz-Villena *et al.*, 1995; 1997; 2000) basado en las afinidades lingüísticas y en los resultados derivados de estudios filogenéticos de genes HLA de las clases I y II. En dichos estudios se propone que los ancestros de los vascos estarían entre las poblaciones que habitaron hace 10.000 años el Sahara fértil, que posteriormente emigraron gradualmente hacia el norte conforme avanzaba la desertificación de la zona (Arnaiz-Villena *et al.*, 1997).



Figura 2: Mapa de Navarra y provincias próximas.

OBJETIVO

La finalidad de este trabajo ha consistido en valorar el grado de heterogeneidad genética que presenta la población de Navarra frente a otras poblaciones representativas de diferentes continentes, analizando un grupo de inserciones *Alu* situadas en el cromosoma X. Con ello se propone, por un lado, profundizar en el conocimiento de los procesos de microdiferenciación experimentados por la población del norte de Navarra, y, por el otro, comprobar el posible origen paleonorteafricano del patrimonio genético vasco.

MATERIALES Y MÉTODOS

Población de estudio

En el presente trabajo se han analizado 30 muestras de DNA de individuos originarios del norte de la provincia de Navarra, concretamente de los municipios de Elizondo, Lekaroz, Bera de Bidasoa y Leitza (Figura 3). Todos los individuos, sus padres y abuelos eran originarios del norte de Navarra; eran voluntarios informados, sanos y no emparentados y dieron su consentimiento informado. El DNA se encontraba extraído previamente a este trabajo. Las muestras de DNA se encontraban identificadas mediante un código, de modo que no se pudo acceder a ningún dato personal, excepto el sexo, clave para el análisis de inserciones *Alu* del cromosoma X.

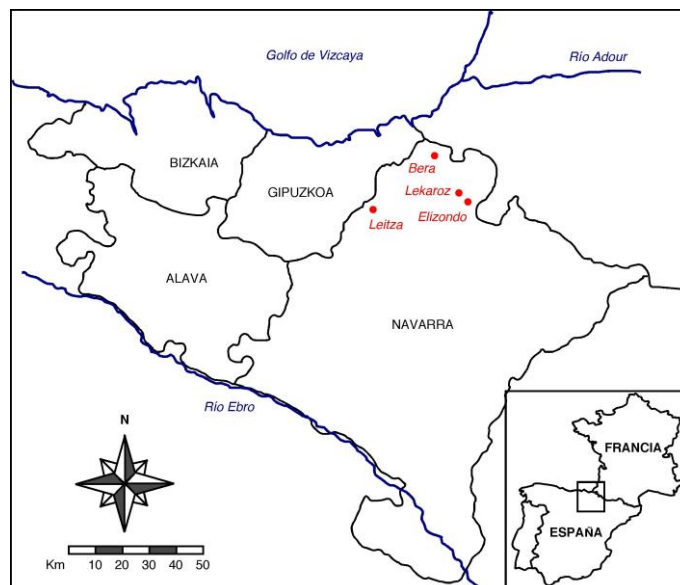


Figura 3: Mapa de Navarra con la ubicación de los lugares de muestreo.

Marcadores utilizados

Las secuencias *Alu* analizadas en este trabajo han sido *Ya5DP3*, *Ya5NBC37*, *YD3JX437* e *Ya5A2DP1*, todas ellas ubicadas en el cromosoma X y pertenecientes a las subfamilias Ya5, YD3 e Ya5a2, que se encuentran dentro de la familia conocida como *Young Alu* (Tabla 1).

Tabla 1: Principales características de las inserciones *Alu* estudiadas (Callinan, 2005)

| Nombre | Ubicación | 5' Primer | 3' Primer | Tamaño del amplificado | |
|-----------------|--------------------------|-------------------------------------|-----------------------------------|------------------------|--------------|
| | | | | Inserción | No inserción |
| <i>Ya5DP3</i> | chrX:4095243-4260035 | ATCTTGAGAA TCTCTACCAC | TCCTCTGGAT TTCAGGGTTG | 487 | 162 |
| <i>Ya5NBC37</i> | chrX:120184952-120332053 | GCTTGAGGTTTTTCAT ACTACTCTTATCTTT | ACTGTATAAGCATT TCCTCTTTATCTTTC | 497 | 184 |
| <i>Yd3JX437</i> | chrX:46070143-46177191 | TGGTGTACCTTA GTCCAAAGACC | TTTGCATCTCA GAACTTTTTCCT | 547 | 235 |
| <i>Ya5a2DP1</i> | ChrX:46070143-46177191 | CACAACAAAGTA CTGCAAAGAGT | CTTTGTTTTCTG ATTTTGAAGG | 939 | 615 |

Identificación de los genotipos

Con el objeto de amplificar las regiones del genoma donde se encuentran las inserciones *Alu* analizadas en este trabajo, se utilizó la técnica de la reacción en cadena de la polimerasa, también conocida como PCR (del inglés: *Polymerase Chain Reaction*). Para todas las muestras se utilizó un volumen final de 10µL, con 50ng de DNA molde, 0,2µM de cada primer, 200µM de dNTPs, 1,5 mM de MgCl₂, 10 mM de Buffer y 1U de Taq DNA polimerasa.

Las reacciones PCR se llevaron a cabo en un termociclador *Gene Amp PCR System 9700*, de Applied Biosystems. Los ciclos de temperaturas y tiempos se muestran en la Tabla 2.

Tabla 2: Ciclos PCR realizados para la amplificación de 4 inserciones Alu del cromosoma X.

| | Alu Ya5DP3 | Alu Ya5NBC37 | Alu Yd3JX437 | Alu Ya5a2DP1 |
|---------------------------|---------------|-----------------|-----------------|-----------------|
| Desnaturalización Inicial | 94°C | 94°C | 94°C | 94°C |
| | 150 seg | 150 seg | 150 seg | 150 seg |
| Desnaturalización | 94°C | 94°C | 94°C | 94°C |
| | 1 min | 1 min | 1 min | 1.30 min |
| Anillamiento | 54°C | 59°C | 56°C | 57°C |
| | 1 min | 1 min | 1 min | 1.30 min |
| Extensión | 72°C | 72°C | 72°C | 72°C |
| | 1 min | 1 min | 1 min | 1.30 min |
| Nº ciclos | 32 | 32 | 32 | 32 |
| Extensión Final | 72°C | 72°C | 72°C | 72°C |
| | 10 min | 10 min | 10 min | 10 min |

Para visualizar los resultados, se realizaron electroforesis con los amplificadores obtenidos en la PCR, en geles de agarosa al 1,5%, a 115V durante 20 minutos. Para que las muestras fuesen visibles a la luz ultravioleta (*GelVue transiluminador UV - Syngene GVM20*) se utilizó el colorante Real Safe, según las indicaciones del fabricante. Además, para poder interpretar los resultados de manera correcta, se añadieron controles positivos (heterocigotos conocidos para la inserción), negativos (los mismos reactivos, sin DNA) y un marcador de tamaño del DNA, *DNA Molecular Weight Marker V* de Roche, el cual permitió estimar con fiabilidad el tamaño de cada banda.

Tratamientos estadísticos

Cálculo de las frecuencias alélicas

Las frecuencias alélicas para cada uno de los marcadores se calcularon por el método del conteo directo (Nei, 1987), según el cociente:

$$\text{Frecuencia alélica} = n_a / (2N_m + N_v)$$

donde N_m es el tamaño de la muestra en mujeres, N_v , el tamaño de la muestra en varones y n_a es la frecuencia observada del alelo.

El error estándar (SE) de las frecuencias alélicas se ha obtenido a partir de la fórmula (Li, 1968):

$$\text{Error estándar} = \pm \sqrt{p_i (1 - p_i) / 2N}$$

donde p_i es el valor de la frecuencia del i -ésimo alelo y N es el número de cromosomas analizados.

Probabilidad de coincidencia (PM):

La probabilidad de coincidencia (PM, del inglés *Matching Probability*) es la probabilidad de que dos individuos tomados al azar tengan el mismo genotipo. Para su cálculo se utilizó el programa ForensicMarkers 1.0 (Peña *et al*, 2016).

Contenido de información polimórfica (PIC) y diversidad genética (GD):

La diversidad genética (GD, del inglés *Gene Diversity*) o heterocigosidad esperada (*h*), representa la probabilidad de que dos alelos o haplotipos tomados al azar en una muestra sean diferentes. Al ser una de las medidas de variabilidad genética más comúnmente empleadas, es muy útil en análisis comparativos entre muestras poblacionales de diferentes zonas geográficas, grupos étnicos, troncos lingüísticos..., etc (Nei, 1987).

El grado de polimorfismo de un determinado marcador genético se puede cuantificar por diversos parámetros, dentro de los cuales destaca el contenido de información polimórfica, PIC (del inglés *Polymorphism Information Content*). Su valor depende esencialmente del número de alelos del locus examinado y de las frecuencias registradas en la población de estudio, como puede apreciarse en las fórmulas:

$$h = 1 - \left(\sum_{i=1}^n p_i^2 \right)$$

$$PIC = 1 - \left(\sum_{i=1}^n p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2$$

donde *h* es la heterocigosidad esperada por locus (cuya expresión matemática se corresponde con la diversidad genética de Nei (1987), *n* es el número de alelos para cada marcador genético y *p_i*, *p_j* la frecuencia de los alelos *i-ésimo* y *j-ésimo* en la población estudiada.

Se considera que valores superiores a 0,5 proporcionan mucha información; si están comprendidos entre 0,25 y 0,5 son medianamente informativos, y si se trata de valores inferiores a 0,25, apenas proporcionan información.

Para su cálculo se utilizó el programa ForensicMarkers 1.0 (Peña *et al*, 2016).

Índice de paternidad (Pi):

El índice de paternidad (Pi, del inglés *Paternity Index*) es un parámetro que se calcula para un locus como la proporción de verosimilitud X/Y, siendo X la probabilidad de que el padre alegado haya transmitido el alelo paterno, e Y, la probabilidad de que lo haya transmitido otro hombre cualquiera que se encuentre dentro del mismo grupo continental y resida en el mismo país.

Para su cálculo se utilizó el programa ForensicMarkers 1.0 (Peña *et al*, 2016).

Matriz de distancias R

La matriz de distancias R, o coeficiente R de Harpending y Jenkins (1973), se ha utilizado para establecer las distancias genéticas entre poblaciones. Se basa en la aplicación del concepto de parentesco entre poblaciones, que se puede definir como la probabilidad de que un alelo tomado en un individuo al azar en la población *i* sea idéntico por descendencia a otro alelo del mismo locus, tomado también en un individuo al azar de la población *j*.

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

donde p_{ik} es la frecuencia del *k*-ésimo alelo en la *i*-ésima población, y \bar{p}_k , la frecuencia alélica media del *k*-ésimo alelo en la población.

La matriz R puede transformarse en una distancia euclídea mediante la expresión:

$$d_y^2 = r_{ii} + r_{jj} - 2r_{ij}$$

Para calcular este coeficiente de distancia se utilizó el programa GeDis 2.0 (Peña *et al*, 2009).

Test de neutralidad de Ewens-Watterson:

Es un método utilizado para averiguar si existe una presión selectiva en un polimorfismo y una población determinadas. Se comparan las frecuencias observadas de homocigotos con las esperadas bajo la teoría neutral si el locus se encuentra en equilibrio mutación-deriva. De esta manera, las frecuencias alélicas pueden ser demasiado uniformes (selección estabilizadora) o demasiado desiguales (barrido selectivo) en comparación con la expectativa neutral.

Para obtener este índice se utilizó el programa Arlequin 3.5 (Excoffier y Lischer, 2010).

Análisis de escalamiento multidimensional

El MDS (del inglés: *Multidimensional Scaling*) proporciona una representación euclídea de una serie de poblaciones en un espacio multidimensional, generalmente limitado a 2 dimensiones (Kruskal, 1964). Las posiciones relativas de las poblaciones reflejan el grado de afinidad o proximidad genética entre ellas.

Para realizar el análisis MDS se utilizó el programa Past 3.0 (Hammer *et al.*, 2001).

Índice de fijación (F_{ST})

Es un estadístico acuñado por Wright (1943), que evalúa el nivel de diferenciación genética entre subpoblaciones, de acuerdo al modelo de población subdividida. Su valor depende del balance establecido entre el flujo genético y la deriva genética. El índice de fijación se emplea en estudios de genética de poblaciones, teniendo en cuenta que habrá una mayor diferenciación genética entre las subpoblaciones cuanto mayor sea el efecto de la deriva; esto se traduce a un mayor valor de F_{ST} . De esta manera, si el valor de F_{ST} es 0, significará que no existen diferencias genéticas entre las subpoblaciones, mientras que si toma un valor de 1, implicará un aislamiento genético máximo. A partir de este concepto, se ha desarrollado un coeficiente de distancia entre pares de poblaciones F_{ST} , con el que es posible obtener el grado de significación de las diferencias (Reynolds *et al*, 1983).

Para obtener este índice se utilizó el programa Arlequin 3.5 (Excoffier y Lischer, 2010).

Análisis de la varianza molecular (AMOVA)

El análisis molecular de la varianza o AMOVA (del inglés: *Analysis of Molecular Variance*) determina la proporción de varianza genética total que puede atribuirse a los diferentes niveles de organización poblacional establecidos. En general, un test AMOVA permite desglosar la diversidad genética en tres niveles jerárquicos fundamentales: dentro de la población (entre individuos), entre poblaciones dentro de grupos y entre grupos poblacionales. Aparte de criterios estrictamente geográficos, también pueden analizarse agrupamientos poblacionales basados en criterios lingüísticos, étnicos o de otra índole, siempre que constituyan agrupaciones jerarquizadas.

Para realizarlo se utilizó el programa Arlequin 3.5 (Excoffier y Lischer, 2010).

Neighbour Joining

Es un método que sirve para reconstruir fenogramas a partir de una matriz de distancias. Comienza por un árbol en forma de estrella (con todas las ramas iguales), en el que sucesivamente se desplazan los diferentes nodos que unen cada par de poblaciones y se modifican las longitudes de las correspondientes ramas, partiendo de la distancia más pequeña. Mediante sucesivas iteraciones se alcanzará un dendrograma consenso, en el que las distancias en la matriz y en el árbol se diferencian lo mínimo posible (Saitou y Nei, 1987).

Para realizar este análisis se usó el programa Past 3.0 (Hammer *et al.*, 2001), y para visualizarlo, TreeView 1.6.6 (Page, 1996).

Network (Redes de Haplotipos)

Son construcciones gráficas con un objetivo filogenético, que muestran las relaciones ancestro-descendiente entre un grupo de haplotipos. Se utilizan para construir las redes filogenéticas, inferir los tipos ancestrales, los tipos posibles, las ramificaciones evolutivas y eventualmente proporcionar estimaciones de la edad del ancestro común de los haplotipos seleccionados.

Las estimas de las frecuencias haplotípicas se realizaron por conteo directo, en el caso de los varones, y mediante Máxima Verosimilitud con el programa Arlequin versión 3.5 (Excoffier y Lischer, 2010), en el caso de las mujeres.

La filogenia de los haplotipos se ha obtenido mediante el programa Network v4.6 (Bandelt *et al* 1999).

RESULTADOS

Las frecuencias alélicas en la población de Navarra, junto con sus correspondientes errores estándar, se muestran en la Tabla 3. Como puede observarse, el rango de frecuencias es alto, abarcando tanto valores cercanos a 0 como próximos a 1. Las inserciones *Alu Ya5DP3*, *Ya5NBC37* e *Yd3JX437* están muy lejos de la fijación, puesto que sus valores son muy bajos, especialmente el de *Ya5DP3*; en

cambio, *Ya5a2DP1* posee el valor más cercano a la unidad y por lo tanto es la inserción que más posibilidades tendría de fijarse en la población.

Tabla 3: Frecuencia alélica y error estándar para las cuatro inserciones Alu en la población del norte de Navarra

| | Frecuencia | Error estándar |
|-----------------|------------|----------------|
| <i>Ya5DP3</i> | 0,0870 | 0,0282 |
| <i>Ya5NBC37</i> | 0,1800 | 0,0388 |
| <i>Yd3JX437</i> | 0,1633 | 0,0385 |
| <i>Ya5a2DP1</i> | 0,8696 | 0,0351 |

En la Tabla 4 se muestran los parámetros forenses para los 4 marcadores genéticos. Los valores de la heterocigosidad y del PIC son muy bajos, mientras que la probabilidad de coincidencia (PM) y el índice de paternidad (Pi) son altos. No puede decirse que sean valores sorprendentes, puesto que son marcadores bialélicos y, por tanto, poseen un escaso grado de polimorfismo, si se comparan con los microsatélites, que son los marcadores habituales en los análisis forenses. Así, el valor promedio de PM para los microsatélites del sistema CODIS en europeos es de 0,084, de Pi es 2,505, de PIC es 0,750 y de heterocigosidad 0,779.

Tabla 4: Parámetros forenses de las cuatro inserciones Alu analizadas. PM: Probabilidad de coincidencia. Pi: Índice de paternidad. PIC: Contenido de información polimórfica. Het: Heterocigosidad.

| | PM | Pi | PIC | Het |
|-----------------|-------|-------|-------|-------|
| <i>Ya5DP3</i> | 0,720 | 0,594 | 0,146 | 0,159 |
| <i>Ya5NBC37</i> | 0,540 | 0,709 | 0,252 | 0,295 |
| <i>Yd3JX437</i> | 0,565 | 0,688 | 0,236 | 0,273 |
| <i>Ya5a2DP1</i> | 0,624 | 0,647 | 0,201 | 0,227 |

Con objeto de poder comparar la población de Navarra con otras poblaciones de varios continentes, se han tomado de la bibliografía todas las frecuencias alélicas publicadas hasta la fecha, las cuales se muestran en la Tabla 5 (Gayá-Vidal *et al.*, 2010; Athanasiadis *et al.*, 2007).

Tabla 5: Frecuencias alélicas para las cuatro inserciones Alu en diferentes poblaciones

| | Europa | | | Costa de Marfil | África | | | América | | Asia |
|-----------------|---------|-------|------------|-----------------|-------------|------------|-------|---------|---------|---------|
| | Navarra | Creta | País Vasco | | Monte Atlas | Oasis Siwa | Túnez | Aimaras | Quechua | Yakutia |
| <i>Ya5DP3</i> | 0,087 | 0,067 | 0,134 | 0,230 | 0,199 | 0,239 | 0,232 | 0,000 | 0,007 | 0,185 |
| <i>Ya5NBC37</i> | 0,180 | 0,339 | 0,324 | 0,313 | 0,432 | 0,275 | 0,412 | 0,088 | 0,029 | 0,326 |
| <i>Yd3JX437</i> | 0,163 | 0,143 | 0,104 | 0,552 | 0,213 | 0,246 | 0,033 | 0,520 | 0,554 | 0,538 |
| <i>Ya5a2DP1</i> | 0,870 | 0,874 | 0,869 | 0,294 | 0,669 | 0,629 | 0,830 | 0,863 | 0,894 | 0,682 |

Con esta base de datos es posible valorar si existe una heterogeneidad genética significativa entre continentes para estos marcadores. Con esta finalidad, se ha realizado un análisis de la varianza molecular (AMOVA), cuyos resultados pueden observarse en la Tabla 6.

Tabla 6: Análisis molecular de la varianza. Los valores se muestran en porcentajes. * $p < 0,05$; ** $p < 0,01$

| | Entre grupos | Entre poblaciones dentro de grupos | Entre individuos dentro de poblaciones |
|-----------------|--------------|---------------------------------------|---|
| <i>Ya5DP3</i> | 9,58** | 0,00 (-0,49) | 90,91** |
| <i>Ya5NBC37</i> | 10,69* | 2,01* | 87,30** |
| <i>Yd3JX437</i> | 14,38 | 10,50** | 75,12** |
| <i>Ya5a2DP1</i> | 29,02* | 10,24** | 60,74** |
| Media | 15,92 | 5,69 | 78,52 |

Atendiendo a la variación entre continentes, se observa que el valor medio del porcentaje de varianza entre grupos es del 15,92%, relativamente alto, obteniéndose valores dispares entre las cuatro inserciones *Alu*, particularmente *Ya5a2DP1*. En todo caso, 3 de las 4 inserciones muestran significación en la heterogeneidad entre continentes. En cuanto a las diferencias entre poblaciones dentro de grupos, la media (5,69%), entra dentro del rango establecido para el conjunto del genoma, aunque los resultados entre marcadores también varían. Finalmente, respecto a la variación entre individuos dentro de poblaciones, la media (78,52%), es relativamente baja, siendo de nuevo el valor dispar el de *Ya5a2DP1*, cuyo porcentaje de variación es considerablemente más bajo que el del resto.

Es preciso mencionar que, debido al diseño del método, pueden encontrarse valores negativos, lo que en la práctica equivale a 0, como es el caso de la varianza entre poblaciones dentro de grupos para *Ya5DP3*. No obstante, esto puede suponer que la suma de los 3 componentes sea superior a 100.

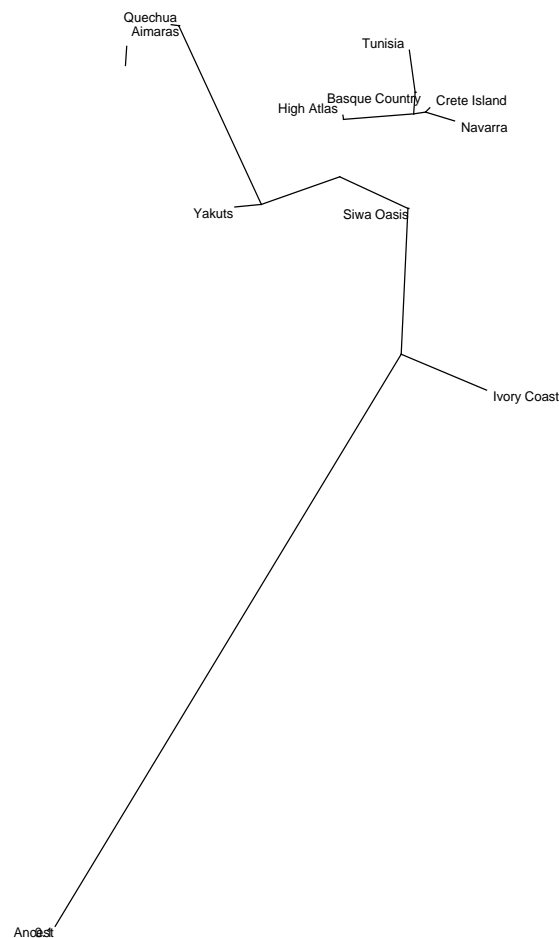


Figura 4: Análisis NJ realizado sobre una matriz de distancias R.

Puesto que el AMOVA ha mostrado, en general, una heterogeneidad significativa entre continentes, se ha procedido a analizar las relaciones entre las diferentes poblaciones consideradas, con el fin de caracterizar sus relaciones de parentesco.

La Figura 4 muestra un análisis Neighbour-Joining, realizado a partir de una matriz de distancias R, pudiendo apreciarse la posición de las poblaciones humanas en función de su proximidad microevolutiva respecto de la población ancestral. Se observa una distribución que se aproxima al modelo *Out of Africa* de expansión de nuestra especie por los diferentes continentes. Así, atendiendo a su posición en el árbol filogenético, la población más cercana a la ancestral es Costa de Marfil, la única población del África subsahariana; aparecen relativamente próximas a ella las poblaciones bereberes de África del norte (oasis de Siwa, en Egipto, y Alto Atlas de Marruecos) y más alejadas, el resto de poblaciones. Navarra aparece cercana a Creta y el País Vasco, formando un grupo de poblaciones europeas, y próxima a ellas se encuentra la población de origen árabe de Túnez. Finalmente, en una nueva rama se encuentra la población asiática de Yakutia, puente de unión con las poblaciones amerindias, Quechuas y Aimaras.

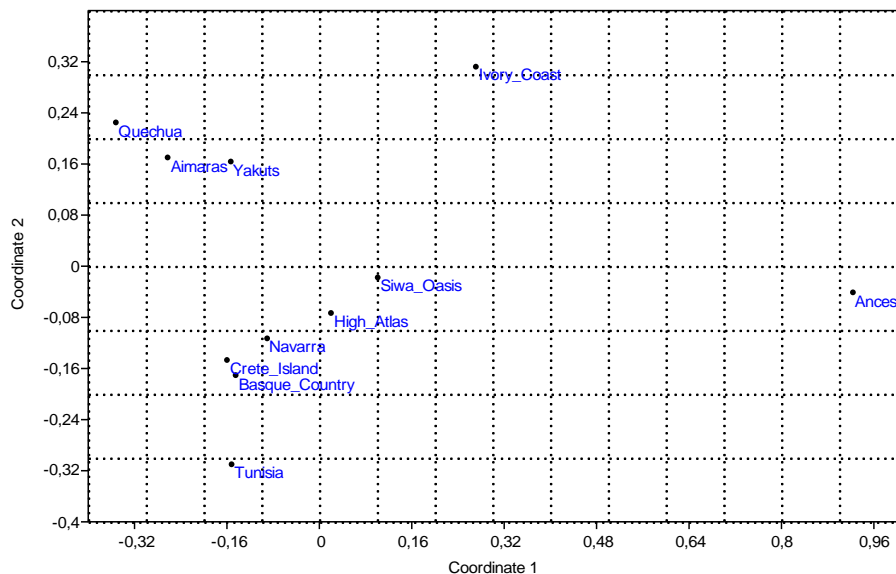


Figura 5: Análisis de escalamiento multidimensional (MDS) obtenido a partir de la matriz R de Harpending y Jenkins para poblaciones de África, Europa, América y Asia. Se ha tenido en cuenta la población ancestral (Root). Coeficiente de estrés: 0,018.

La Figura 5 muestra un análisis de escalamiento multidimensional obtenido a partir de una matriz de distancias R, en el que, de forma similar a la Figura 4, se pueden observar varios grupos claramente diferenciados: las poblaciones pertenecientes a América del Sur (Aimaras y Quechua) y la población asiática de Yakutia en el cuadrante superior izquierdo; hacia el centro y en la zona superior, Costa de Marfil; en el cuadrante inferior izquierdo y alrededor del centroide aparecen agrupadas las poblaciones europeas (Navarra, Creta y País Vasco) y las del norte de África (Túnez, Monte Atlas y oasis de Siwa), aunque Túnez, de origen árabe, se separa claramente de las poblaciones de origen bereber; finalmente, en el extremo derecho y notablemente apartada del resto, está la población ancestral.

Debido a la polarización que introduce la población ancestral, se ha realizado un segundo MDS en el que no se incluye, con el fin de aumentar la resolución del análisis (Figura 6).

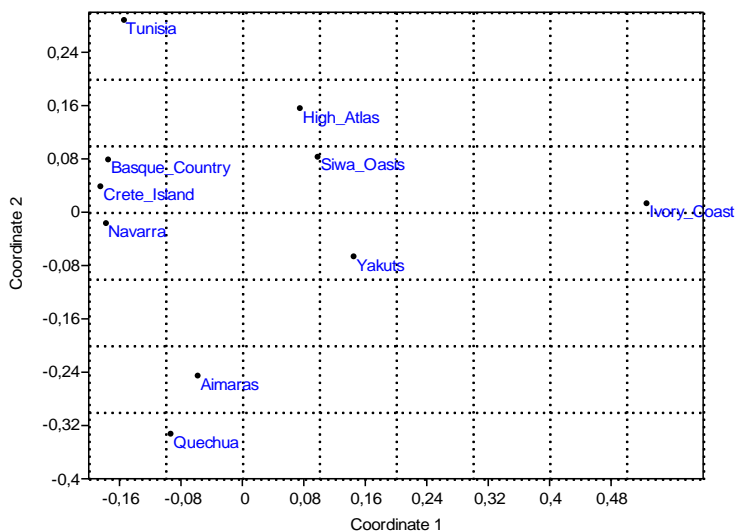


Figura 6: Análisis de escalamiento multidimensional (MDS) obtenido a partir de la matriz R de Harpending y Jenkins para poblaciones de África, Europa, América y Asia. No se ha tenido en cuenta la población ancestral.
Coeficiente de estrés: 0,048.

La Figura 6 muestra que, en realidad, las poblaciones de Europa y del norte de África son genéticamente heterogéneas a pesar de su proximidad geográfica y del contacto que ha habido entre ellas a lo largo de la historia, apareciendo claramente discriminadas las poblaciones europeas respecto de la población árabe de Túnez y de las norteafricanas de origen bereber. Quechua y Aïmaras se mantienen alejadas del resto y genéticamente próximas entre sí, en tanto que Costa de Marfil aparece en el extremo derecho, muy separada del resto de poblaciones. En cuanto a los Yakutos, aparecen entre las poblaciones africanas y las americanas, como cabría esperar de acuerdo a la teoría del *Out of Africa*, puesto que los antecesores de los actuales amerindios cruzaron el estrecho de Bering desde el centro norte de Asia, en territorios próximos a la actual Yakutia, y, en origen, provenían de África.

Tabla 7: Resultados del índice F_{ST} para la población del norte de Navarra frente al resto de poblaciones para las cuatro inserciones Alu. Significación: * $p < 0,05$; ** $p < 0,01$

| | Ancestral | Aïmaras | Quechua | Creta | País Vasco | Costa Marfil | Monte Atlas | Oasis de Siwa | Túnez | Yakutos |
|----------|-----------|---------|---------|--------|------------|--------------|-------------|---------------|---------|---------|
| Ya5DP3 | 0,081** | 0,081** | 0,056** | 0 | 0 | 0,061** | 0,038* | 0,069** | 0,061** | 0,025 |
| Ya5NBC37 | 0,172** | 0,024 | 0,104** | 0,055* | 0,041* | 0,035 | 0,128** | 0,018 | 0,111* | 0,048** |
| Yd3JX437 | 0,151** | 0,245** | 0,278** | 0 | 0,006 | 0,278** | 0 | 0,015 | 0,084** | 0,267** |
| Ya5a2DP1 | 0,869** | 0 | 0 | 0 | 0 | 0,508** | 0,098** | 0,134** | 0 | 0,089** |

Con el objeto de valorar las distancias observadas en los análisis Neighbour-Joining y MDS, se han calculado los valores del índice F_{ST} y su correspondiente significación. En la Tabla 7 se muestran los

valores de la distancia entre el norte de Navarra y el resto de poblaciones. En general se observan diferencias significativas en 2 o 3 inserciones con todas las poblaciones excepto Creta y el País Vasco, que sólo muestran 1 valor significativo. Si se aplica la corrección de Bonferroni, en general se mantiene la significación de las diferencias, excepto en las poblaciones europeas, que no muestran ninguna diferencia significativa. Los valores más altos se observan en las comparaciones con la población ancestral, ya que es la población de origen y no ha sufrido cambios microevolutivos. En conjunto, la existencia de un gran número de diferencias significativas confirma la verosimilitud de las agrupaciones poblacionales obtenidas en las figuras 4 a 6.

Tabla 8: **Resultados para el test de neutralidad de Ewens-Watterson; * $p < 0,05$**

| | Europa | | | África | | | Túnez | América | | Asia |
|----------|---------|--------|------------|-----------------|-------------|---------------|-------|---------|---------|-------|
| | Navarra | Creta | País Vasco | Costa de Marfil | Monte Atlas | Oasis de Siwa | | Aimaras | Quechua | |
| Ya5DP3 | 0,482* | 0,518* | 0,395 | 0,249 | 0,298 | 0,245 | 0,255 | N.A | 1 | 0,295 |
| Ya5NBC37 | 0,278 | 0,141 | 0,156 | 0,139 | 0,058 | 0,162 | 0,072 | 0,451* | 0,726 | 0,127 |
| Yd3JX437 | 0,327 | 0,392 | 0,424* | 0,04 | 0,272 | 0,226 | 0,703 | 0,02 | 0,055 | 0,031 |
| Ya5A2DP1 | 0,339 | 0,364 | 0,368 | 0,183 | 0,145 | 0,107 | 0,301 | 0,376 | 0,435* | 0,151 |

¿Podrían las distancias entre poblaciones ser, al menos en parte, debidas a procesos selectivos? En la Tabla 8 se puede observar que, en general, existe neutralidad en la mayor parte de las poblaciones para cada inserción *Alu*, esto es, no está actuando la selección natural. Los escasos valores significativos desaparecen cuando se aplica la corrección de Bonferroni para múltiples comparaciones. Por tanto, las diferencias observadas entre la población del norte de Navarra con el resto de poblaciones serán debidas a procesos que afectan al equilibrio entre flujo génico y deriva y no a procesos selectivos.

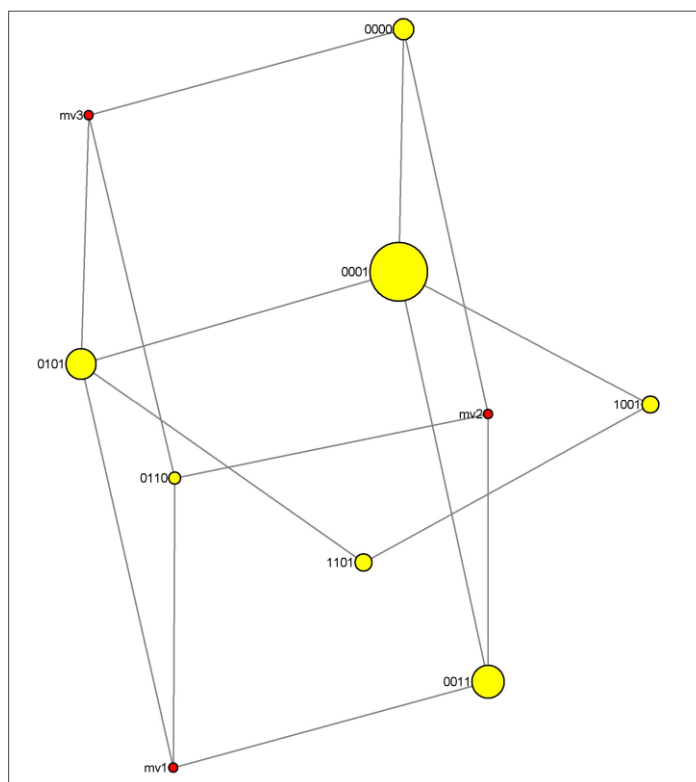


Figura 7: Network para los haplotipos encontrados en el norte de Navarra

Con objeto de profundizar en el conocimiento de las diferencias entre poblaciones para este grupo de marcadores, se calcularon las frecuencias haplotípicas en la población del norte de Navarra, con las que se realizó un Network (Figura 7). Lamentablemente, al no disponer de datos individuales, este análisis no se pudo realizar para el resto de poblaciones.

En el gráfico pueden observarse 7 haplotipos. El más abundante es 0001, con inserción en *Ya5a2DP1*, del que presumiblemente derivan tres haplotipos: 0011 (inserciones en *Yd3JX437* e *Ya5a2DP1*), 0101 (inserciones en *Ya5NBC37* e *Ya5a2DP1*) y 1001 (inserciones en *Ya5DP3* e *Ya5a2DP1*), este último menos frecuente. Por otra parte, 1101 (inserciones en *Ya5DP3*, *Ya5NBC37* e *Ya5a2DP1*) podría derivar de 0101 o de 1001. El haplotipo 0110, con inserciones en *Ya5NBC37* e *Yd3JX437*, es muy poco frecuente y no deriva directamente de los demás haplotipos presentes en el gráfico. Pudiera ocurrir que un aumento del tamaño muestral hiciera aparecer alguno de los haplotipos intermedios, aunque también es posible que proceda de otra población, por algún episodio de flujo génico en el pasado.

Finalmente, el haplotipo ancestral (0000, sin inserciones), es poco abundante; su baja frecuencia es fruto de la acumulación de las 4 inserciones, que se encuentran en todos los continentes y también presumiblemente de episodios de cuello de botella que parecen haber afectado al menos a *Ya5a2DP1*. En efecto, esta inserción, que conforma el haplotipo más frecuente en Navarra, sólo se encuentra con una

frecuencia por debajo de 0,6 en Costa de Marfil (0,294), por lo que su frecuencia podría haber aumentado por efecto de la deriva durante el comienzo de las migraciones iniciales fuera de África.

DISCUSIÓN

Las secuencias *Alu* conforman un valioso registro fósil molecular que puede permanecer prácticamente invariable de generación en generación en virtud de unas tasas de mutación extremadamente bajas ($2,3 \times 10^{-9}$), lo cual es una característica muy útil en estudios filogenéticos (Relethford, 2012). Así pues, las inserciones *Alu* necesariamente serán marcadores muy conservadores, revelando procesos con una notable antigüedad en nuestra especie, como el *Out of Africa* y la colonización de los diferentes continentes. Esta es la razón por la que, en promedio, nuestras inserciones ofrecen unos valores de varianza entre grupos, entre poblaciones dentro de grupos y entre individuos dentro de poblaciones similares al promedio del conjunto del genoma humano (15,91; 5,69 y 78,52% respectivamente en nuestras 4 inserciones, frente a 10,00; 5,50 y 84,50% en promedio para el conjunto del genoma) (Barbujani & Goldstein, 2004).

En relación a estos marcadores, se han observado unas frecuencias elevadas ($>0,6$) de *Ya5a2DPI* en todos los continentes excepto en el África subsahariana (0,294), por lo que cabe pensar en un efecto fundador al comienzo del proceso *Out of Africa* que provocase la aparición de mayores frecuencias. La inserción *Yd3JX437*, por el contrario, presenta valores elevados ($>0,5$) en el África subsahariana, Asia y América, en tanto que más bajos ($<0,25$) en Europa y norte de África, por lo que pudo haberse dado una pérdida por efecto de la deriva antes de la colonización de Próximo Oriente y Europa. Por último, *Ya5NBC37* presenta valores relativamente altos en todos los continentes excepto en América, por lo que el efecto fundador que supuso la colonización de este continente pudo haber afectado a la baja a sus frecuencias.

Sin embargo, por el propio carácter conservador de estos marcadores y por su escaso grado de polimorfismo, su capacidad de discriminación entre individuos y también entre poblaciones a escala local es escasa, como confirman los resultados obtenidos al calcular varios parámetros forenses (Tabla 4). Esta es la razón por la que Navarra y el País Vasco no muestran un mayor grado de parentesco que Navarra y Creta, quedando agrupadas como poblaciones europeas, sin una mayor resolución (figuras 4 a 6). En cualquier caso, estas 3 poblaciones aparecen claramente diferenciadas de las norteafricanas, separándose asimismo las de origen bereber (Alto Atlas y oasis de Siwa) de la de origen árabe (Túnez).

Como se ha mencionado previamente Arnaiz-Villena *et al.*, (1995; 1997; 2000), postulan un posible origen paleonorteafricano del patrimonio genético vasco, basándose en las afinidades lingüísticas y en los resultados derivados de estudios filogenéticos de genes HLA de las clases I y II. Sin embargo, como se observa en los análisis Neighbour-Joining y MDS, las poblaciones europeas aparecen claramente separadas de las norteafricanas, como confirma el índice F_{ST} , mostrando diferencias significativas para varias inserciones *Alu* con las poblaciones árabes y bereberes (Tabla 7). Además, diversos especialistas argumentan que, a pesar de que la Península Ibérica está separada de África por los 15 km que tiene el estrecho de Gibraltar, éste podría haber actuado como una fuerte barrera al flujo génico entre ambos continentes (Simoni *et al.*, 1999; Comas *et al.*, 2000a). La heterogeneidad genética entre las poblaciones

de ambos márgenes también podría haberse visto reforzada por la baja densidad de las poblaciones del norte de África a lo largo de su historia demográfica, lo cual habría promovido la microdiferenciación local por efecto de la deriva genética (Simoni *et al.*, 1999).

Respecto a los resultados del análisis Network (Figura 7), se ha observado que el haplotipo mayoritario en Navarra no es el ancestral, sino el que presenta inserción para *Ya5a2DPI*, probablemente debido a un efecto fundador sobre las frecuencias de esta inserción, como se ha mencionado anteriormente.

CONCLUSIONES

A la luz de lo analizado y discutido en el presente trabajo, se pueden extraer las siguientes conclusiones:

- Existen diferencias significativas entre Navarra y las poblaciones no europeas. Esta heterogeneidad genética se debe presumiblemente a alteraciones en el equilibrio entre el flujo génico y la deriva durante el proceso de colonización de los diferentes continentes y no a la acción de la selección. Es por ello que la distribución de la heterogeneidad refleja correctamente las huellas del proceso *Out of Africa*.
- Los vascos no poseen un origen común con los bereberes. El estudio de las cuatro inserciones *Alu* para las poblaciones europeas y norteafricanas revela una gran heterogeneidad genética entre Navarra y País Vasco frente a las poblaciones de Alto Atlas y oasis de Siwa, confirmando que el estrecho de Gibraltar ha supuesto una barrera al flujo génico entre África y Europa.

BIBLIOGRAFÍA

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

Arnaiz-Villena A, Benmamar D, Alvarez M, Diaz-Campos N, Varela P, Gomez-Casado E & Martinez-Laso J (1995) HLA allele and haplotype frequencies in Algerians. Relatedness to Spaniards and Basques. *Hum Immunol*, 43(4):259-268

Arnaiz-Villena A, Martínez-Laso J, Gomez-Casado E, Diaz-Campos N, Santos P, Martinho A & Breda-Coimbra H (1997) Relatedness among Basques, Portuguese, Spaniards, and Algerians studied by HLA allelic frequencies and haplotypes. *Immunogenetics*, 47:37-43

Arnaiz-Villena A & Lubell D (2000) Prehistoric Iberia: Genetics, Anthropology and Linguistics. *Current Anthropology*, 41:636-638

Athanasiadis G, Esteban E, Via M, Dugoujon JM, Moschonas N, Chaabani H & Moral P (2007) The X chromosome *Alu* insertions as a tool for human population genetics: data from European and African human groups. *European Journal of Human Genetics*. 15: 578–583

Bandelt HJ, Forster P, Röhl A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evo*, 16:37-48

Barbujani G & Goldstein DB (2004) Africans and Asians abroad: Genetic Diversity in Europe *Annu. Rev. Genomics Hum. Genet*, 5:119–150

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ & Deininger PL (1994) African origin of human specific polymorphic *Alu* insertions. *Proc Natl Acad Sci USA*, 91:12288-12292

Batzer MA & Deininger PL (1991) A human-specific subfamily of *Alu* sequences. *Genomics*, 9:481-487

Batzer MA & Deininger PL (2002) *Alu* repeats and human genomic diversity. *Nat Rev Genet*, 3:370-379

Callinan (2005) *Alu* retrotransposition-mediated genomic variation within the primate order. Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J & Sajantila A (2000a) *Alu* insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet*, 107:312-319

Excoffier L & Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*, 10:564-567
<http://cmpg.unibe.ch/software/arlequin35/>

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. *Journal of human genetics*, 52(4):317-327.

Gayá –Vidal M, Dugoujon JM, Esteban E, Athanasiadis G, Rodriguez A, Villena M, Vasquez R & Moral P (2010) Autosomal and X Chromosome *Alu* Insertions in Bolivian Aymaras and Quechuas: Two Languages and One Genetic Pool. *American journal of human biology*: 22:154–162

Gobierno Vasco (2008) IV Mapa sociolingüístico 2006. EDK Servicio Central de Publicaciones del Gobierno Vasco, Vitoria-Gasteiz, Spain www.euskadi.net

Hammer Ø, Harper DAT & Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4(1):9 http://palaeo-electronica.org/2001_1/past/issue1_01.htm

Harpending H & Jenkins T (1973) Genetic distance among southern African populations. In: *Methods and theories of anthropological genetics*. EDK University of New Mexico Press, Albuquerque, USA.

Instituto Nacional de Estadística. (2011). Madrid: INE <http://www.ine.es/>

Kruskal J (1964) Non metric multidimensional scaling: A numeric method. *Psychometrika*, 29:115-129

Labuda D & Zietkiewicz E (1994) Evolution of secondary structure in the family of 7SL-like RNAs. *J Mol Evol*, 39:506-518

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Showkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chisoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglu S, Birney E, Bork P, Brown DG, Burge CB,

Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409:860–921.

Li C (1968) Population genetics. EDK The University of Chicago Press, Chicago USA

Mighell AJ, Markham AF & Robinson PA (1997) *Alu* sequences. *FEBS lett*, 417:1-5

Nei M (1987) Molecular Evolutionary Genetics EDK Columbia University Press, New York USA

Page RDM (1996) TreeView: An application to display phylogenetic trees on personal computers. *CABIOS*, 12:357-358.

Peña JA, Alfonso-Sanchez MA, Pérez-Miranda AM, García-Obregón S & Gómez-Pérez L (2009) GeDis: un programa para análisis de datos en antropogenética. *Antropo*, 20:49-59

Peña JA, Alfonso-Sánchez MA (2016) ForensicMarkers: un programa para el cálculo de parámetros forenses. *Antropo*, 35: 79-82. www.didac.ehu.es/antropo

Quentin Y (1992) Origin of the *Alu* family: a family of *Alu*-like monomers gave birth to the left and the right arms of the *Alu* elements. *Nucleic Acids Res*, 20:3397-3401

Relethford JH (2012) Human population Genetics. Hoboken, New Jersey, USA: Wiley- Blackwell

Reynolds J & Weir BS & Cockerham CC (1983) Estimation of the coancestry coefficient: bases for a short-term genetic distance. *Genetics*, 105: 767-779.

Roy-Engel AM, Batzer, MA & Deininger, PL (2008) Evolution of Human Retrosequences: *Alu*. In: Encyclopedia of Life Sciences (ELS). EDK John Wiley & Sons, Ltd: Chichester, UK.

Saitou N & Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*: 4(4):406-25.

Simoni L, Gueresi P, Pettener D & Burbujani G (1999) Patterns of gene flow inferred from genetic distances in the Mediterranean region. *Hum Biol*, 71:399-415

Ullu E & Tschudi C (1984) *Alu* sequences are processed 7SL RNA genes. *Nature*, 312:171-172

Weiner A, Deininger P & Efstradiatis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu Rev Biochem*, 55:631-661

Wright S (1943) Isolation by distance. *Genetics*, 28(2):114.