

Las poblaciones han sido ordenadas según el valor de la media para el segmento de 360pb de menor a mayor, ya que la media de diferencias por parejas es también un índice de diversidad. La jerarquía de las poblaciones es prácticamente idéntica a la observada con el índice de diversidad nucleotídica (Tabla R-43), sin embargo, a diferencia de éste, los valores disminuyen con la longitud del fragmento analizado. Por otra parte, al igual que para otros índices, el orden de las poblaciones en cuanto al grado de diversidad no se ve alterado para el fragmento de 302pb, sin embargo si presenta cambios para el segmento de 153pb, que involucran a las poblaciones GALU, SAR y HAD. Estos cambios son equivalentes a los observados para el índice de diversidad nucleotídica, por lo que vale la pena tenerlos en cuenta para la interpretación de los resultados obtenidos en la reconstrucción filogenética, como ya se ha mencionado con anterioridad.

La comparación de la curva de la distribución de las diferencias por parejas en distintas poblaciones aporta información acerca de la fecha relativa de sus expansiones (asumiendo que todas presentan la misma tasa de mutación), ya que de acuerdo con Rogers y Harpending (1992), la curva de la distribución se desplaza con el tiempo hacia la derecha en función de la tasa de mutación (1 unidad del eje horizontal cada $1/2u$ generaciones, siendo u la tasa de mutación). Así, la comparación de la distribución que presentan las poblaciones BRI y CAT para el fragmento de 391pb (Fig. R-16) indica que la expansión de la población Catalana ocurrió con anterioridad a la de la población BRI.

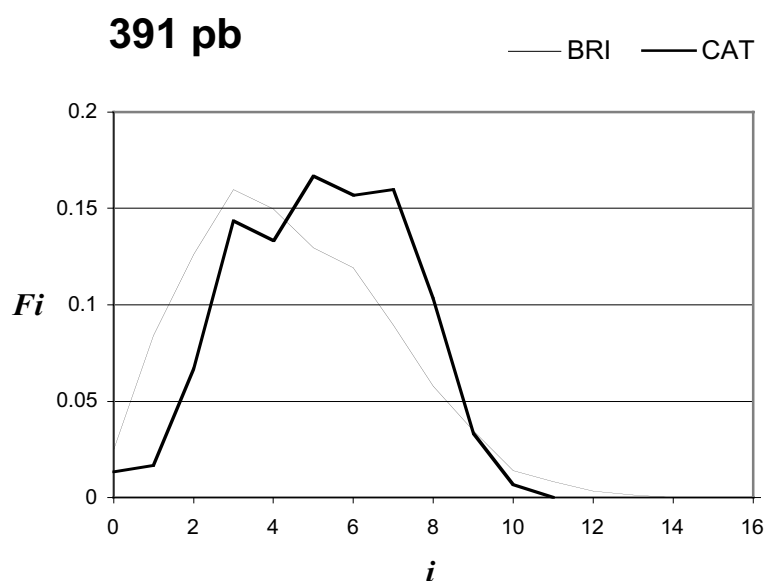


Figura R-16 Distribución de las diferencias por parejas que presentan las poblaciones CAT y BRI, para el fragmento de 391pb; i , número de diferencias; F_i , frecuencia relativa de i .

Un panorama más completo se puede obtener analizando el fragmento de 360pb. Como se puede observar en la Figura R-17, la curva de la distribución de la población CAT se sitúa en una posición muy diferente a las distribuciones de las poblaciones ibéricas aquí analizadas. En efecto, las poblaciones del País Vasco (VAS) y Galicia (GALI) presentan distribuciones que se sitúan en el extremo izquierdo del eje horizontal, indicando que el inicio de su expansión ocurrió en tiempos más recientes que para el resto de poblaciones (Bertranpetit et al. 1995; Salas et al. 1998).

En cambio, la curva de la población CAT se superpone en cierta medida con las curvas de las poblaciones TOS y TRK, indicando una expansión relativamente antigua. La irregularidad de las distribuciones que presentan las poblaciones de Galura y Cerdeña, hace difícil su comparación respecto a la población actual Catalana; en tanto que las posiciones de las distribuciones de BRI y FIN, indican que se expandieron después que la población CAT.

El análisis de la distribución que presentan las poblaciones para el fragmento de 302pb (Fig. R-18) resulta interesante porque permite incluir a la población Beréber de Argelia (ARG). La distribución de esta población indica una expansión más antigua incluso que la de la población Turca, resultado que está en acuerdo con la idea de que los beréberes representan una población que se estableció en tiempos pre-neolíticos, quizá hace 50,000 años (Macaulay et al. 1999a), aunque el extenso análisis de Rando y col. (1998) demuestra que la composición mitocondrial de las poblaciones beréberes contiene influencias paleolíticas, neolíticas y posiblemente otras más recientes tanto del norte de Europa como de fenicios, romanos, árabes y también de la península ibérica. El resto de poblaciones sigue el mismo patrón observado para el segmento de 360pb.

La distribución de las diferencias por parejas de la población de la Plaça Vella, se ha comparado con la distribución que presentan el resto de poblaciones para el segmento de 153pb (Fig. R-19). Un aspecto notable, es que la curva de la Plaça Vella se sitúa más a la izquierda respecto a la población actual de Cataluña, indicando un tiempo de expansión más reciente. Incluso, de acuerdo con esta comparación, la población de la Plaça Vella se habría expandido después que la población Vasca. La curva de la población Catalana actual, se sitúa claramente más a la izquierda que la curva de la población TOS y de la población TRK, contrariamente a lo observado en los segmentos más grandes.

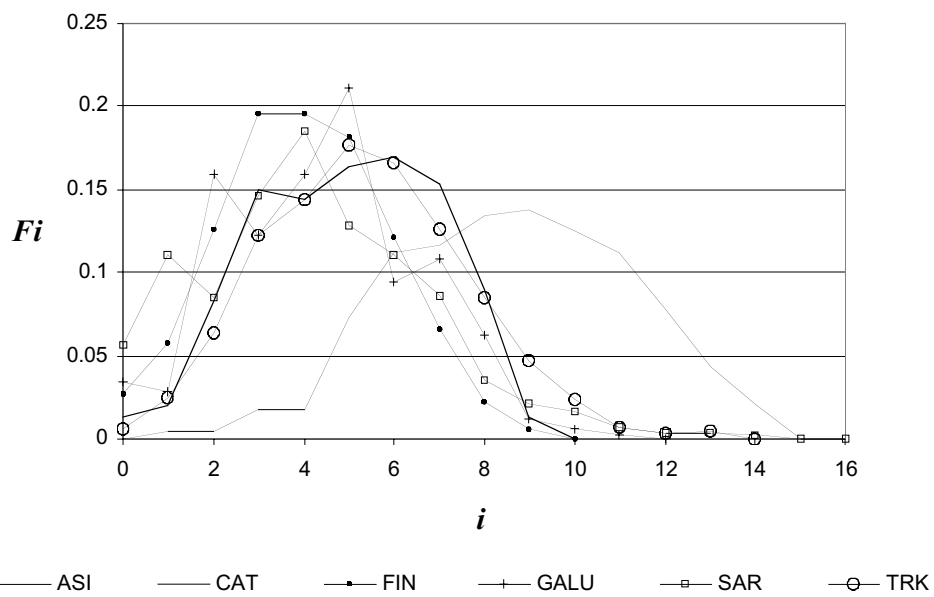
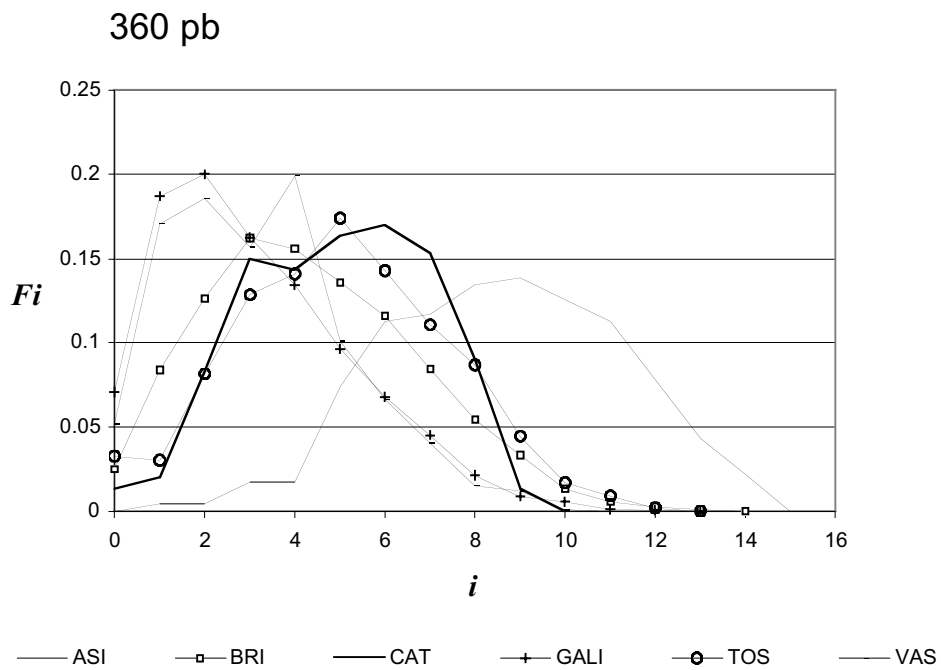


Figura R-17 Distribución de las diferencias por parejas de las poblaciones analizadas para el fragmento de 360pb; *i*, número de diferencias; *Fi*, frecuencia relativa de *i*.

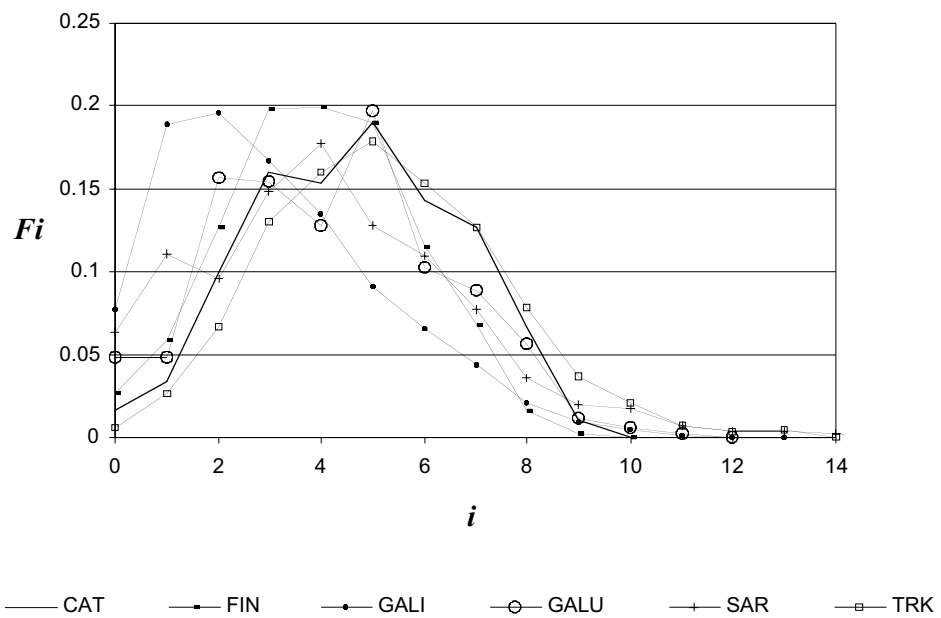
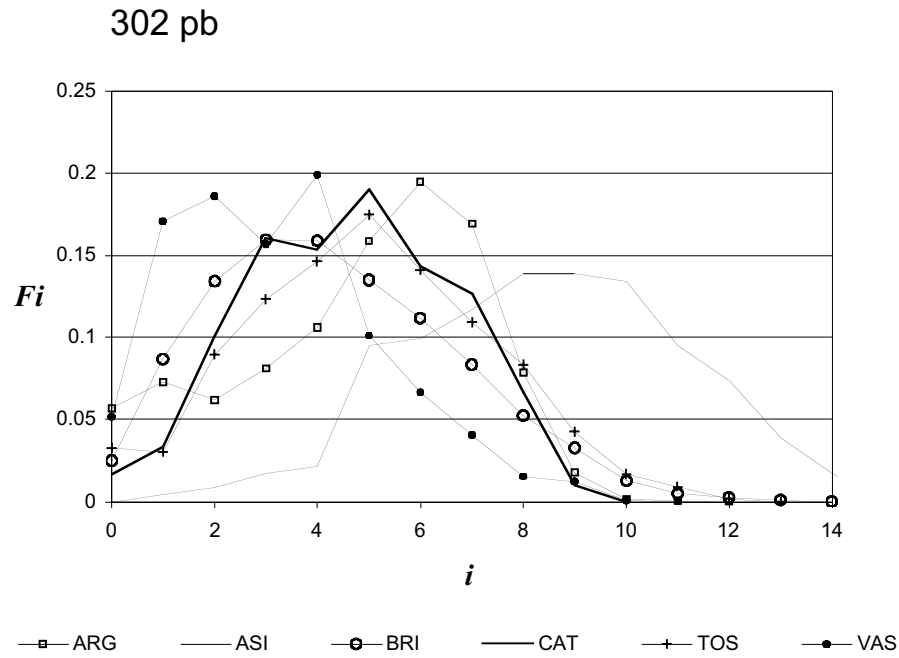


Figura R-18 Distribución de las diferencias por parejas para el fragmento de 302pb; *i*, número de diferencias; *Fi*, frecuencia relativa de *i*.

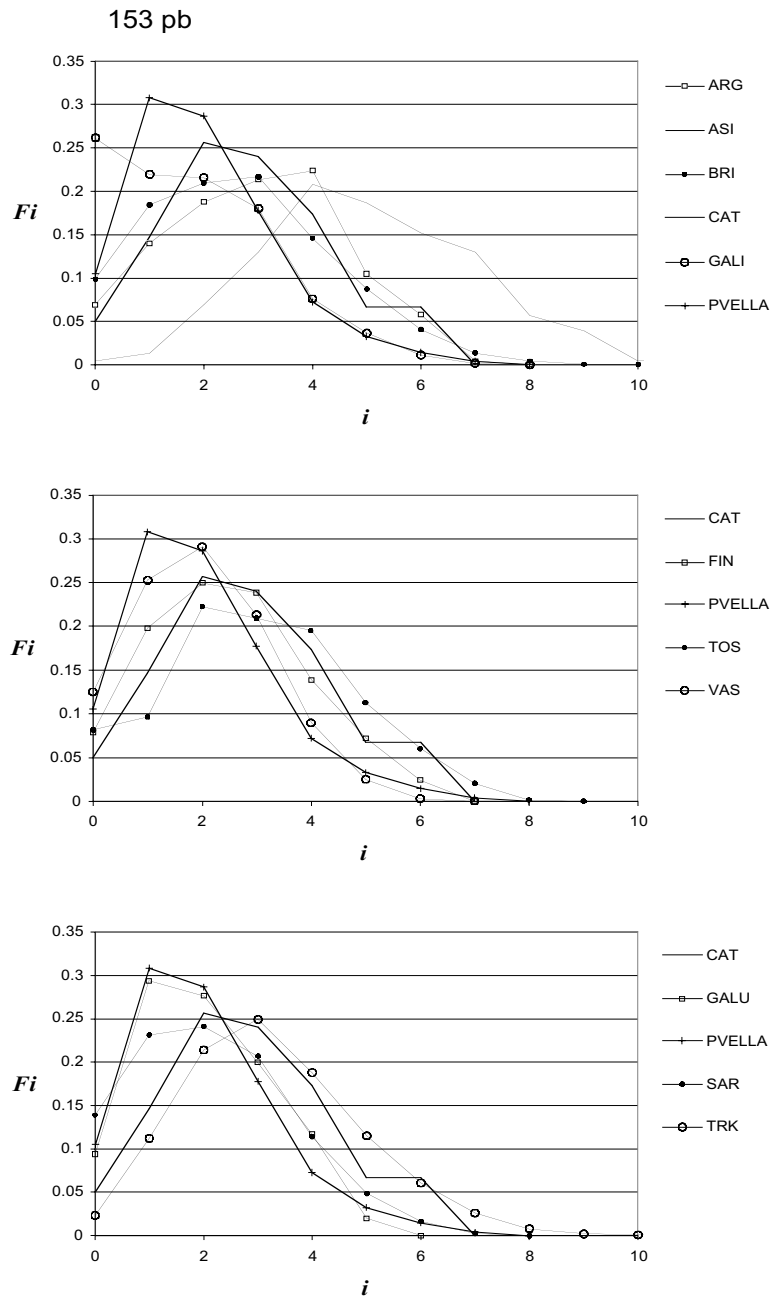


Figura R-19 Distribución de las diferencias por parejas en el fragmento de 153pb; i , número de diferencias; Fi , frecuencia relativa de i .

Para realizar una comparación más precisa se calcularon los tiempos de expansión de las poblaciones utilizando el parámetro τ (Tau) del modelo de expansión súbita y la tasa de mutación del segmento analizado (ver Material y Métodos). Este cálculo se realizó únicamente con fines comparativos y no para intentar determinar el tiempo real de la expansión, debido a los problemas, ya discutidos, que presenta la estimación de la tasa de mutación en el mtDNA.

Resulta interesante observar que mientras el valor de τ disminuye a medida que el fragmento analizado es más pequeño, no sucede lo mismo con los tiempos de expansión, que muestran la tendencia contraria. Los índices de diversidad nucleotídica, indican que los segmentos de 302 y 153pb presentan una tasa de evolución más rápida, y en esta comparación se ha utilizado la misma tasa para los cuatro segmentos, lo que en parte podría explicar la observación anteriormente descrita. Dicha tasa ha sido derivada por Horai y col. (1995) para un fragmento de 378pb y sería interesante intentar derivar tasas específicas para cada segmento aquí analizado y utilizarlas en la comparación. Aunque esta derivación puede hacerse de forma relativamente sencilla a partir de los datos de Horai y col., por el momento no realizaremos esta labor porque se sitúa fuera de los objetivos de este trabajo, sin embargo, dejamos este punto abierto para futuros trabajos.

Lo que ahora nos interesa es la observación de los diferentes tiempos de expansión de las poblaciones y determinar si el orden cronológico cambia respecto al tamaño del fragmento analizado. De acuerdo con esta comparación (Tabla R-48), los segmentos de 360 y 302pb indican que la población de Turquía (TRK) fue la primera población europea en expandirse, en tanto que la última fue la población de Galicia, seguida de la del País Vasco.

Sorprendentemente, la población actual de Cataluña presenta un tiempo de expansión muy antiguo, superada únicamente por la población de Turquía. Este resultado podría contradecir la hipótesis de una única expansión desde el sudeste de Europa por la zona del mediterráneo, defendida por algunos investigadores (Simoni et al. 2000a; ver Marco de Referencia). Sin embargo, el orden cronológico presentado por el resto de

poblaciones si concuerda con la hipótesis mencionada; por lo tanto, lo que estos resultados parecen indicar en realidad es que la población Catalana ha tenido una historia demográfica diferente. Esta historia, podría involucrar lo sucedido durante la última glaciación, ya que de acuerdo con varios autores, el Pirineo Catalán fue una de las zonas de refugio para los pobladores de la época (Torrioni et al. 1998b; Simoni et al. 2000a y las referencias citadas en ambos trabajos).

Población	N	Tamaño del segmento (pb)							
		391		360		302		153	
		<i>Tau</i>	<i>Tiempo</i>	<i>Tau</i>	<i>Tiempo</i>	<i>Tau</i>	<i>Tiempo</i>	<i>Tau</i>	<i>Tiempo</i>
GALI	92	-	-	1.913	25796	1.862	29930	1.065	33790
VAS	45	-	-	2.366	31904	2.366	38031	<u>1.978</u>	62758
P.VELLA	24	-	-	-	-	-	-	<u>1.978</u>	62758
SAR	69	-	-	2.696	36354	2.569	41294	1.925	61076
BRI	100	3.208	39828	3.227	43514	3.164	50858	2.136	67771
GALU	27	-	-	4.008	54045	3.357	53961	<u>2.011</u>	63805
FIN	47	-	-	<u>4.008</u>	54045	<u>3.966</u>	63750	<u>2.474</u>	78495
ARG	85	-	-	-	-	4.017	64570	<u>2.95</u>	93597
TOS	48	-	-	4.266	57524	4.206	67608	<u>3.031</u>	96167
CAT	25	<u>5.200</u>	64559	<u>4.960</u>	66882	<u>4.647</u>	74696	<u>2.807</u>	89060
TRK	45	-	-	<u>5.378</u>	72519	<u>5.246</u>	84325	<u>3.257</u>	103338
ASI	22	-	-	<u>8.610</u>	116100	<u>8.429</u>	135488	<u>5.061</u>	160575

Tabla R-48 Tiempo estimado de la expansión según el modelo de Rogers y Harpending (1992), utilizando una tasa evolutiva de 2.575×10^{-6} (Horai et al. 1995) y un tiempo generacional de 25 años. Los valores de Tau subrayados son iguales a la media de la diferencia por parejas (ver Tabla R-47).

Así, la zona pudo conglomerar un número de individuos de distintas procedencias, alterando la dinámica del crecimiento y dando origen a una estructura poblacional diferente. De acuerdo con Torrioni y col. (1998b), después de la última glaciación, las poblaciones se expandieron hacia el norte desde los refugios glaciales, pero Simoni y col. (2000a) no encuentran evidencia de esta expansión. Nuestros datos dan cuenta de una historia demográfica de la población Catalana diferente a la que se deduce de la estructura genética mitocondrial del resto de poblaciones analizadas, pero no es posible inferir a partir de ellos si existió la expansión hacia el norte de Europa. También es importante aclarar que el valor de τ para la población Catalana ha sido obtenido directamente de la media de las diferencias por parejas, dado que ésta ha resultado ser mayor que su varianza (ver Material y Métodos). No sabemos hasta que punto el haber obtenido este parámetro de esta forma haya podido sesgar el resultado.

El orden cronológico observado mediante el análisis del fragmento de 153pb, es en general similar al obtenido con los fragmentos grandes (Tabla R-48), sin embargo, puede observarse que las poblaciones de Cerdeña y Galura no quedan situadas en la misma posición. Estos cambios son el resultado lógico de los cambios observados en el índice de diversidad nucleotídica y de la diferencia media por parejas, y corroboran que una parte importante de la estructura genética mitocondrial en estas dos poblaciones está definida en la zona que queda situada fuera del fragmento de 153pb.

La población de la Plaça Vella muestra un tiempo de expansión más antiguo que el de las poblaciones de Galicia, País Vasco y Galura, por lo que el resultado de la comparación gráfica de las distribuciones de este segmento, que mostraba una expansión más reciente de la población de la Plaça Vella que de las poblaciones de Galicia y País Vasco queda refutado con la comparación hecha a través del cálculo del tiempo de expansión utilizando una tasa de mutación.

Tamaño inicial de la población

El modelo de expansión súbita permite estimar también el tamaño que tenía la población al inicio de la población. Dado que θ (Theta, número esperado de diferencias por parejas), es proporcional al tamaño de la población, a partir de θ_0 es posible calcular el tamaño de la población inicial (Rogers y Harpending, 1992). Desafortunadamente, el método de los momentos de Rogers (1995) no permite estimar este parámetro si el promedio de las diferencias por parejas es mayor que su varianza (ver Material y Métodos) y en esta situación se encuentran diversas poblaciones aquí analizadas, entre ellas la población actual de Cataluña y la de la Plaça Vella (ver Tabla R-47). Dado que las diferencias que presenta la población CAT en su distribución de diferencias por parejas podrían estar relacionadas con este parámetro (especialmente la pendiente tan pronunciada de la cara derecha de la curva), se ha buscado otro método para estimarlo. Los resultados se presentan a continuación.

Historia demográfica: inferencia de máxima verosimilitud

Para los datos de la población Catalana actual, se realizó un análisis exhaustivo mediante cuatro paneles correspondientes a valores de ρ de 1,000, 10,000, 100,000 y 1,000,000 con $B = 25,000$ (Fig. R-20). Cada panel fue compuesto con 10 valores de θ (de 0.5 a 5) y 10 valores de τ (de 1 a 10), analizando así 100 combinaciones de parámetros distintas. El valor máximo verosímil (L_A^{pop}), que define la historia poblacional más probable de la población Catalana actual, fue obtenido con los parámetros $\theta = 2.5$, $\tau = 1$ y $\rho = 10,000$. En

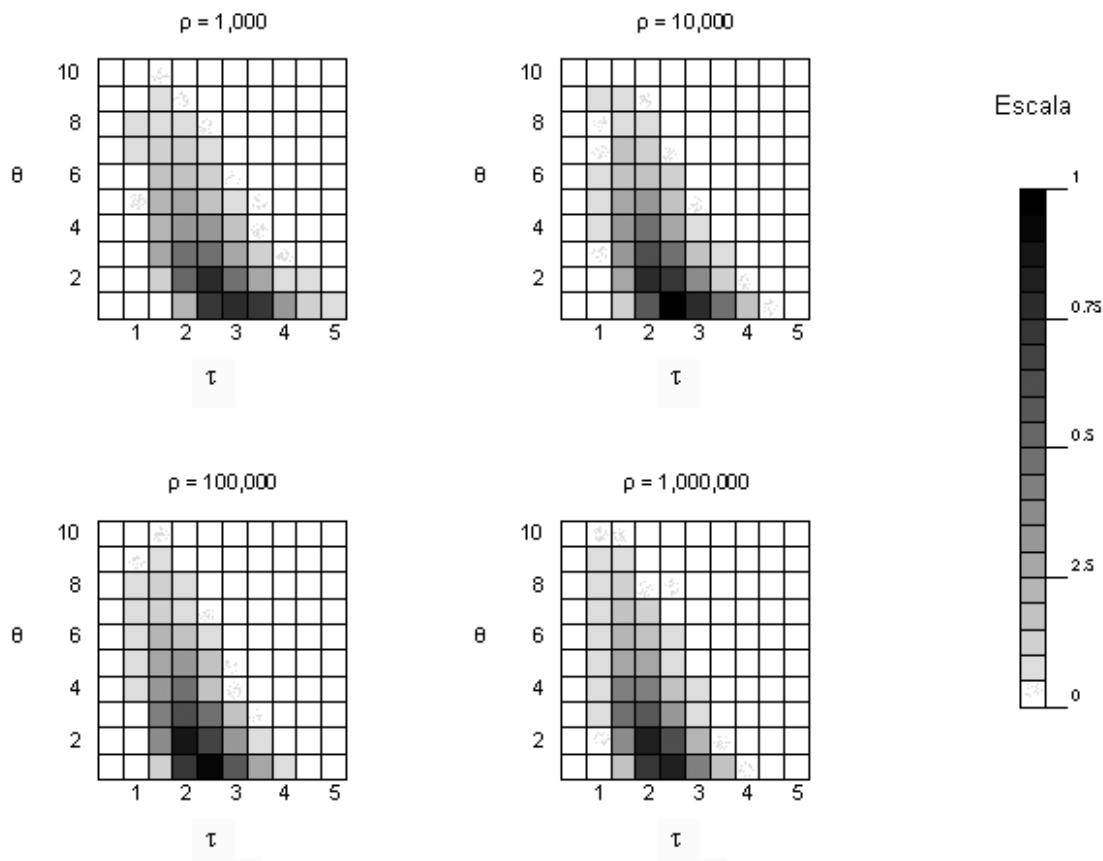


Figura R-20 Resultado de la inferencia de la historia poblacional máximo verosímil de los datos del fragmento de 360pb de la población Catalana actual (CAT). Los colores oscuros corresponden a valores altos de la razón de verosimilitud (ver escala). La combinación de parámetros $\theta = 2.5$, $\tau = 1$ y $\rho = 10,000$ dio el valor más alto. Si se aplica la teoría χ^2 , las celdas coloreadas pertenecen a un conjunto de confianza del 95% (Weiss y von Haeseler, 1998).

la Figura R-20 se muestra una representación gráfica de los valores de la razón de verosimilitud $lik(\underline{t}_0, \underline{t}_0, \underline{t}_0 | k, s) / L_A^{pop}$ obtenidos con cada combinación de parámetros.

Los valores máximo verosímiles que definen la historia de la población Vasca, estimados por Weiss y von Haeseler (1998), son $\underline{t}_0 = 2.25$, $\underline{t}_0 = 1$ y $\underline{t}_0 = 100$. Estos resultados nos indican que el tiempo en el que se expandieron ambas poblaciones (determinados por el valor de \underline{t}_0) fue similar, siendo quizá un poco más antiguo para la población Catalana. Asimismo, este resultado sugiere que el tamaño inicial de las dos poblaciones (determinado por el valor de \underline{t}_0) era prácticamente idéntico. Por lo tanto, la principal diferencia entre la historia demográfica de las dos poblaciones radica en sus respectivos valores de \underline{t}_0 ; siendo cien veces mayor para la población de Cataluña. Cabe destacar, que el segundo valor de verosimilitud más alto en la población Catalana se obtuvo en el panel de $\underline{t}_0 = 100,000$ y el tercer valor más alto en el de $\underline{t}_0 = 1,000,000$; por lo que la probabilidad de que exista un valor verosímil mayor en un conjunto de parámetros con \underline{t}_0 entre 10,000 y 100,000, es mayor que la probabilidad de que exista en un conjunto con \underline{t}_0 entre 1,000 y 10,000. Como este parámetro determina la magnitud de la expansión, debemos interpretar que la expansión de la población Catalana fue por lo menos 100 veces mayor que la de la población Vasca.

En principio, las expansiones posteriores no pueden modificar las características producidas por una primera expansión (Rogers, 1995), por lo que el relativo elevado valor de \underline{t}_0 debería, en teoría, ser el producto de las características de una expansión primigenia. No obstante, este importante incremento de la población pudo estar condicionado por la dinámica poblacional durante el último pleniglacial, cuando Cataluña constituyó una zona de refugio, ya que ni el modelo de expansión súbita de Rogers y Harpending (1992) ni los modelos analizados mediante el método de Weiss y von Haeseler (1998) contemplan los efectos de la migración. Por este motivo, no es posible descartar que la historia de la población Catalana inferida con estos métodos esté sesgada por inmigraciones tanto antiguas como relativamente recientes. Por otra parte, el modelo de cuello de botella tampoco puede ser evaluado de manera exhaustiva con estos métodos.

De acuerdo con Weiss y von Haeseler (1998) la extensión de su método para evaluar diferentes escenarios de crecimiento poblacional, e incluso para evaluar modelos simples con migración, es relativamente sencilla, por lo que no descartamos que en un futuro estas herramientas estén disponibles para profundizar en la inferencia de la historia demográfica de las poblaciones.

El sondeo que realizamos con los datos de la población Toscana nos indica que los parámetros máximo verosímiles que definen su historia se encuentran alrededor de $\mu = 3.5$, $\theta = 1$ y $\rho = 1,000$. Aunque este resultado requiere de una verificación exhaustiva, concuerda con los resultados que apoyan la hipótesis de la ola migratoria este – oeste, ya que indican un tiempo de expansión más antiguo para la población Toscana que el determinado para las poblaciones VAS y CAT. Por otra parte, el valor de μ de esta población está más cercano al de la población Vasca, lo que resalta aún más la peculiaridad de la población Catalana. El valor de ρ es el mismo para las tres poblaciones, indicando que la base poblacional que existía antes de la expansión este – oeste era más o menos homogénea, en cuanto al tamaño, a lo largo del gradiente.

Análisis filogenético intrapoblacional

En la Figura R-21 se muestran los árboles filogenéticos construidos con el método de Neighbor-Joining (N-J) a partir de la matriz de distancias de Tamura-Nei para los fragmentos de 391 y 153pb de la población Catalana actual. Se utilizaron únicamente los haplotipos diferentes, por lo que el árbol de 391pb contiene 22 ramas y el de 153 sólo 17. En la Tabla R-49 se presentan los haplotipos utilizados para cada segmento y población analizada. Los árboles de 360 y 302pb no son presentados ya que el árbol obtenido con el fragmento de 360pb muestra exactamente la misma topología y longitud de las ramas que el de 391pb, en tanto que el obtenido con el fragmento de 302pb es casi idéntico con la diferencia de que algunas ramas resultaron más cortas y de que contiene una rama menos debido a la pérdida de diferenciación de un haplotipo.

Fragmento	Sec. en el gráfico	Secuencias idénticas no incluidas
CAT 391	CAT17-K	CAT18-K, CAT36-K.
CAT 153	CAT24-T2	CAT66-T2.
	CAT1-H	CAT15-H, CAT58-H, CAT60-H, CAT63-H.
	CAT17-K	CAT18-K, CAT36-K.
	CAT24-T2	CAT66-T2.
	CAT19-X	CAT26-Otro.
TOS 360	TOS10-H	(TOS20, TOS31, TOS41, TOS47, TOS3, TOS36, TOS39, TOS49)-H.
	TOS23-J	TOS46-J.
TOS 153	TOS10-H	(TOS20, TOS31, TOS41, TOS47, TOS3, TOS36, TOS39, TOS18, TOS49)-H,
	TOS22-T	(TOS1, TOS23, TOS46)-J, TOS14-Otro. TOS48-T.
GALU 360	G21-J2a	G9-J2a.
	G4-H	G10-H, G14-H, G29-H, G30-H.
	G12-H	G26-H.
GALU 153	G4-H	(G7, G10, G14, G29, G30)-H, G5-J, G28-U.
	G21-J2a	G9-J2a, G2-J2a.
	G6-U5	G23-U5.
	G12-H	G26-H.
PVella 153	PVC48-H	PVE80-H, PVE85-H, (PVC1, PVC16, PVC45)-J, PVC68-Otro.
	PVCE3-K	PVC13-K, PVE30-K.
	PVE35-T	PVC72-H.
	PVE47-H	PVE98-H.
	PVC29-V	PVE40-V, PVE41-V.

Tabla R-49 Grupos de secuencias idénticas encontradas en los distintos fragmentos de las distintas poblaciones. Para la reconstrucción filogenética, las secuencias idénticas fueron excluidas, no obstante, el método de las redes medias permite incluir la frecuencia de los distintos haplotipos. Algunas secuencias del haplogrupo J, pierden su diferenciación en los fragmentos de 153pb y quedan representadas por secuencias del haplogrupo H.

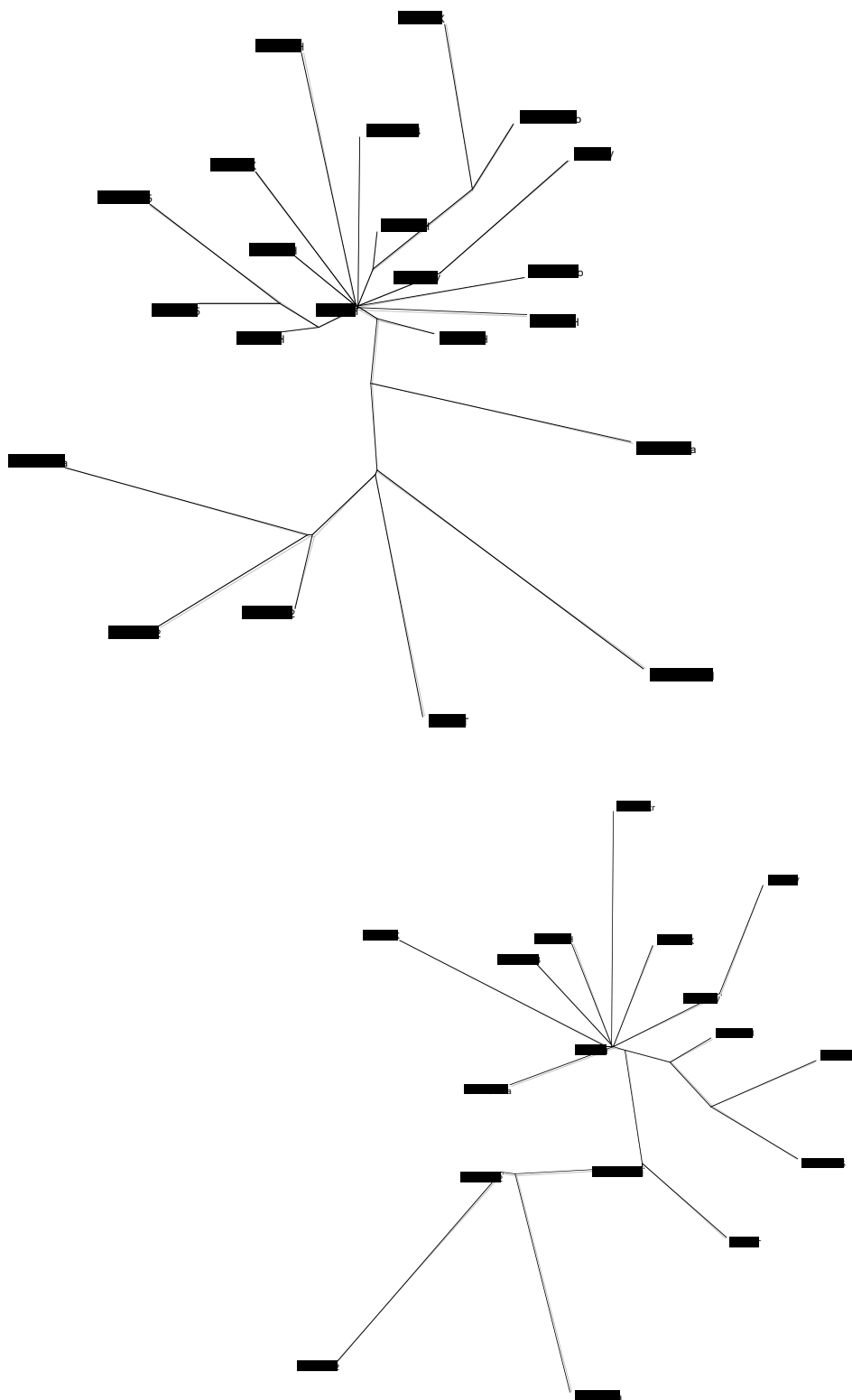


Figura R-21 Árboles filogenéticos construidos con el método de N-J a partir de la matriz de distancias de Tamura-Nei, para los fragmentos de 391pb (arriba) y de 153pb (abajo) de la población CAT. Las secuencias idénticas fueron agrupadas según la Tabla R-49. La secuencia CAT1-H es idéntica a la secuencia de referencia de Cambridge (CRS).

En los nombres de las secuencias analizadas se incluyó el haplogrupo (y el subgrupo cuando se conocía), para evaluar el funcionamiento de los métodos de reconstrucción. Si se asume que los haplogrupos son monofiléticos, la reconstrucción de la filogenia a partir de las secuencias debería diferenciar los distintos haplogrupos. Obviamente existen motivos por los cuales es posible que esto no sea siempre así, a pesar de que los haplogrupos sean efectivamente monofiléticos. Uno de ellos, quizá el más relevante, es la presencia de sitios con una alta tasa de mutación en la región de control. No obstante, los modelos de evolución más complejos, como el de Tamura-Nei con distribución gamma, toman en cuenta la presencia de estos sitios, así como la presencia de sitios invariables (Tamura y Nei, 1993). Por este motivo es razonable esperar que si el método utilizado funciona adecuadamente tendrá la capacidad de realizar una diferenciación de las secuencias en concordancia con el haplogrupo al que pertenecen.

En el árbol de 391pb se observa una clara diferenciación de las secuencias del haplogrupo T cuyo grupo muestra una estructura compatible, de manera general, con la nomenclatura cladística propuesta por algunos investigadores (Macaulay et al. 1999a; Francalacci et al. 1999). La secuencia del subgrupo J2a (CAT23), también resulta diferenciada a pesar de que es la única secuencia que pertenece a este haplogrupo. Asimismo, las secuencias del subgrupo U5 quedan separadas en una rama diferenciada. También se observa cierta clasificación de las secuencias del haplogrupo V, ya que las dos quedan incluidas en la misma rama. Por otra parte, la secuencia del subgrupo U4 no presenta una diferenciación clara, al igual que las secuencias del haplogrupo K (representadas por CAT17).

Las secuencias del haplogrupo H presentan una filogenia en forma de estrella, por lo que son en general las más conflictivas. Por ejemplo, la secuencia CAT20-H, queda incluida en la rama de las secuencias del subgrupo U5 probablemente debido a que comparte la sustitución 16304C con una de las secuencias de U5. El haplogrupo H es el de mayor frecuencia en las poblaciones europeas y de acuerdo con Richards y col. (1998) la estructura que presenta es compatible con una expansión poblacional relativamente reciente, por lo que es normal esperar un alto número de nuevas mutaciones. El problema

radica en que muchas de ellas han podido producirse en sitios con tasas de mutación elevada y esto a fin de cuentas se traduce en eventos de retromutación que pueden confundir los análisis filogenéticos.

La secuencia del haplogrupo X (CAT19) muestra cierta diferenciación y no es extraña la presencia en la misma rama de una de las secuencias que pertenecen al conjunto “otro”, pues comparte con la secuencia CAT19 las sustituciones 16145A y 16223T.

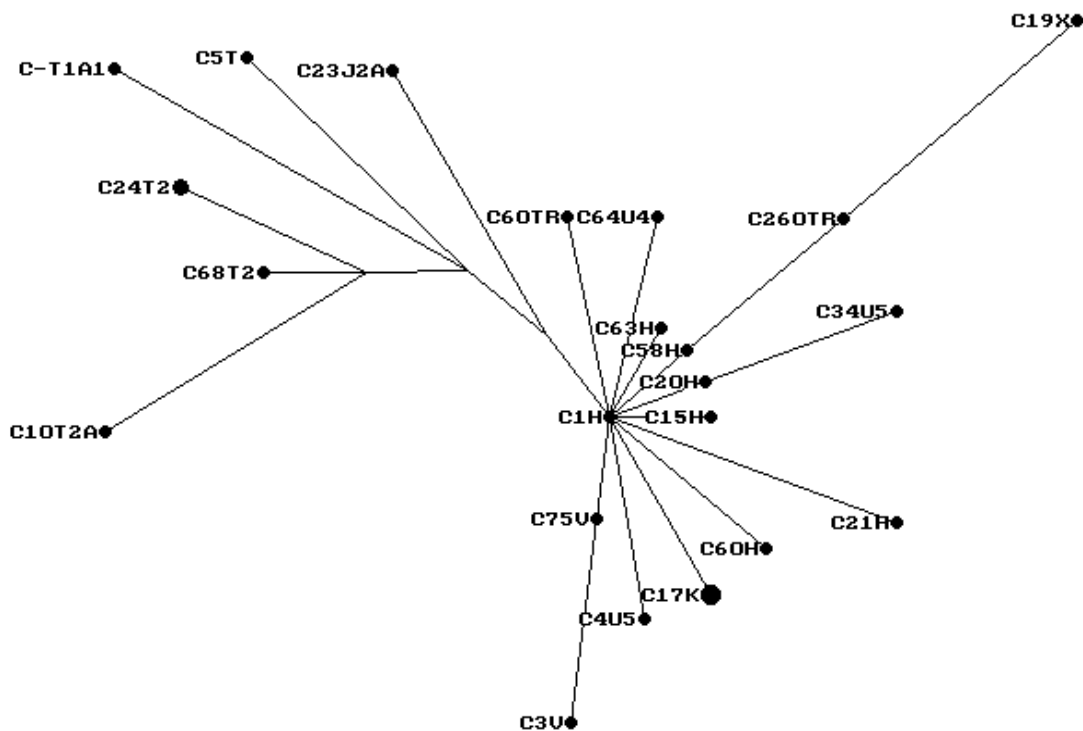
Si se analiza el árbol del fragmento de 153pb, se observa, al margen de la pérdida haplotípica (ver Tabla R-49), que las secuencias de los haplogrupos T y V y del subgrupo U5 no pierden su diferenciación, en tanto que las secuencias del haplogrupo X y del subgrupo J2a, quedan absorbidas dentro del conjunto central.

En la Figura R-22 se presentan las redes medias reducidas calculadas para los fragmentos de 391 y 153pb de la población CAT. Al igual que con los árboles de Neighbor-Joining, las redes obtenidas con los fragmentos de 360 y 302pb no presentan diferencias substanciales con la red del fragmento de 391pb, por lo que tampoco son presentadas. La estructura general de la red del fragmento de 391pb es compatible con la estructura del árbol filogenético, sobretodo en lo que respecta a la diferenciación de las secuencias de los haplogrupos T y V y a las secuencias del subgrupo J2a y del haplogrupo X. La secuencia del haplogrupo X muestra incluso asociaciones con las secuencias CAT26-Otro y CAT58-H equiparables con las observadas en el árbol filogenético. No obstante las secuencias del subgrupo U5 no son clasificadas de la misma forma, pues en la red, la secuencia CAT34-U5 se encuentra asociada con la secuencia CAT20-H, pero no con la secuencia CAT4-U5, en tanto que en el árbol filogenético estas tres secuencias están agrupadas en una misma rama.

La red del fragmento de 153pb muestra también la diferenciación y agrupamiento de las secuencias del haplogrupo T, así como la pérdida de diferenciación de las secuencias del subgrupo J2a y del haplogrupo X. Sin embargo, resulta notoria la

agrupación de las secuencias del subgrupo U5 en contraste con lo observado en la red del fragmento de 391pb pero en concordancia con los árboles filogenéticos de Neighbor-Joining.

Estos resultados destacan, por una parte, la utilidad de conocer el haplogrupo de las secuencias analizadas para no caer en errores en la interpretación de los resultados filogenéticos y por otra parte, muestran los haplogrupos que pueden resultar más útiles cuando se analiza el fragmento de 153pb. En este sentido, el haplogrupo J sería uno de los haplogrupos desfavorecidos al analizar el segmento comprendido entre los nucleótidos



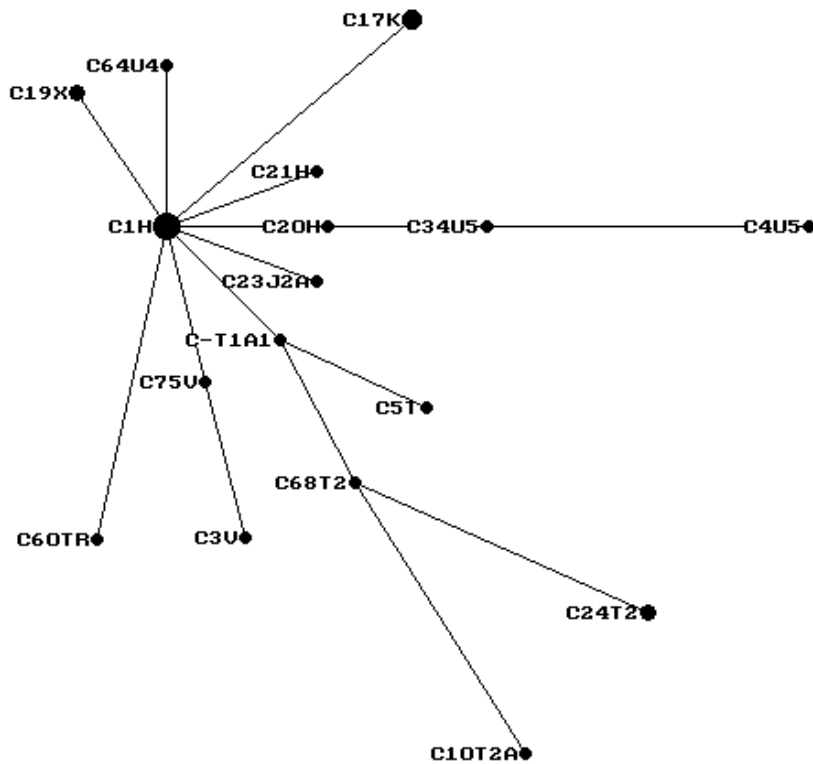


Figura R-22 Redes medias reducidas obtenidas para los fragmentos de 391 (arriba) y 153pb (abajo) de la población Catalana actual (CAT). El tamaño de los círculos es proporcional a la frecuencia del haplotipo (ver Tabla R-49). El programa NETWORK sólo admite nombres de 7 caracteres en mayúscula, por lo que los nombres de los haplotipos han sido adaptados a esta limitación, por ejemplo, substituyendo el prefijo CAT por la letra C. La secuencia C1H (CAT1-H) corresponde a la secuencia de referencia (CRS). La secuencia C-TIA1 corresponde a la secuencia CAT16-T1a1.

16210 y 16362 de 153pb, ya que algunos de sus polimorfismos característicos se encuentran fuera de este fragmento. Esto ocasiona que la simple reducción del fragmento conlleve la pérdida de diferenciación de algunas secuencias del haplogrupo J, que resultan idénticas a la secuencia más común del haplogrupo H, la secuencia de referencia de Cambridge; en tanto que las secuencias restantes de este haplogrupo no conservan substituciones relevantes que permitan su diferenciación en el análisis filogenético ya sea mediante distancias o mediante redes medias reducidas. Por otra parte, los haplogrupos T, V y X y el subgrupo U5 se muestran más robustos para el análisis del fragmento de 153pb (16210-16362).

Para investigar más sobre el poder resolutivo de estos métodos en la diferenciación de las secuencias de los distintos haplogrupos, se analizaron las secuencias

de las poblaciones Toscana y Galuresa, de las que se cuenta también con la caracterización de los haplogrupos mediante enzimas de restricción.

En la Figura R-23 se presentan los árboles de Neighbor-Joining (N-J) de los fragmentos de 360 y 153pb de la población Toscana (TOS). En el árbol de 360pb se observa la diferenciación de las secuencias de los haplogrupos T, J y X al igual que en el árbol de la población CAT. Además, la secuencia del haplogrupo W se encuentra relacionada con las secuencias del haplogrupo X. Asimismo, en este árbol se observa también la diferenciación de las secuencias del haplogrupo K. En esta población no existen secuencias del subgrupo U5, por lo que no es posible evaluar su diferenciación. Las secuencias de otros subgrupos de U no presentan una especial diferenciación y se encuentran esparcidas en distintas ramas. Las secuencias idénticas a la secuencia de referencia de Cambridge, representadas aquí por la secuencia TOS10-H se encuentran en el centro de la filogenia y nuevamente se observa la inestabilidad de las secuencias del haplogrupo H. Por ejemplo, la secuencia TOS17-H se sitúa en el grupo de las secuencias X y W, debido probablemente a que contiene la sustitución 16278C. Si no contáramos con la caracterización del haplogrupo, esta muestra sería considerada erróneamente como perteneciente al clúster de las secuencias X y W.

El árbol del fragmento de 153pb continúa presentando la diferenciación de las secuencias del haplogrupo T y del K. No obstante, las secuencias del haplogrupo J pierden en general su diferenciación, salvo las del subgrupo J1a caracterizadas por la sustitución 16261T. Un aspecto interesante, es que las secuencias de los haplogrupos I, X y W se diferencian en un grupo con tres ramas principales, cada una conducente a las secuencias de

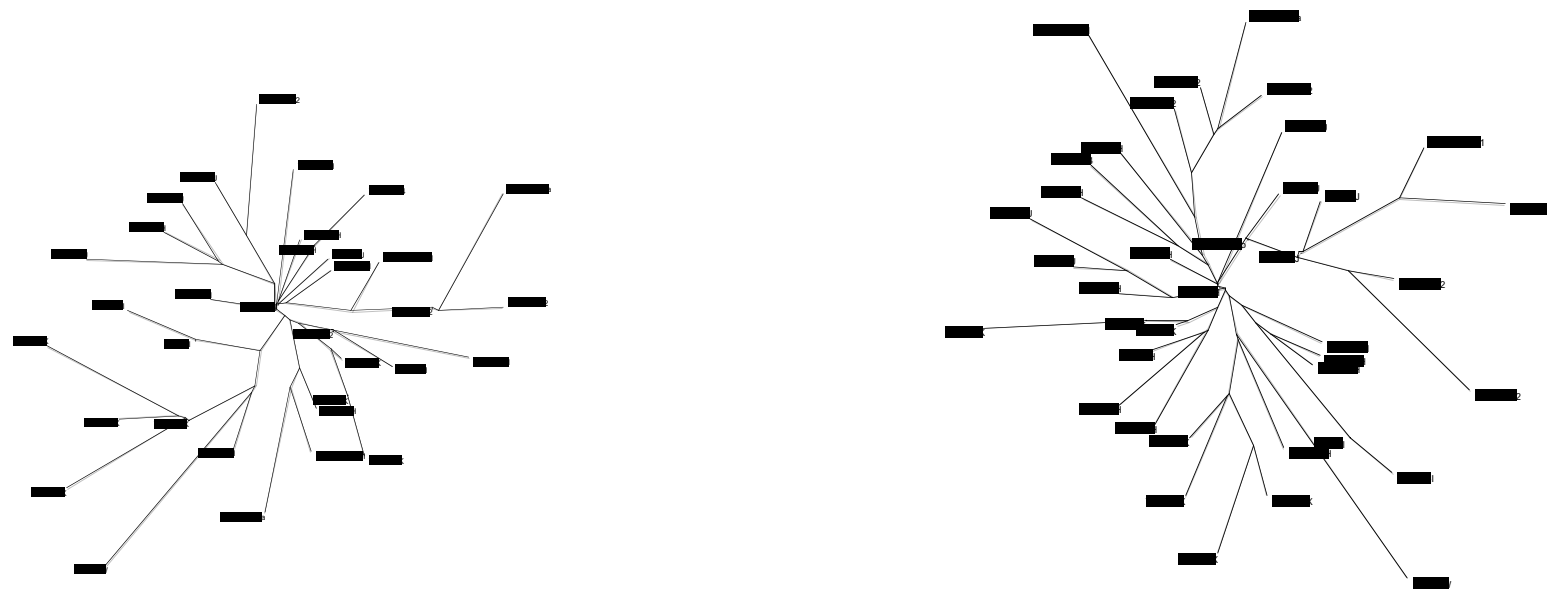


Figura R-23 Árboles filogenéticos construidos con el método de N-J a partir de la matriz de distancias de Tamura-Nei, para los fragmentos de 360 (izquierda) y 153pb (derecha) de la población TOS. Las secuencias idénticas fueron agrupadas según la Tabla R-49. La secuencia TOS10-H corresponde a la secuencia de referencia (CRS).

cada haplogrupo. Esta característica podría indicar que para la definición de las relaciones filogenéticas dentro de algunos haplogrupos resulta útil la eliminación de determinados fragmentos que pueden contener sustituciones que ocasionan un “ruido de fondo”.

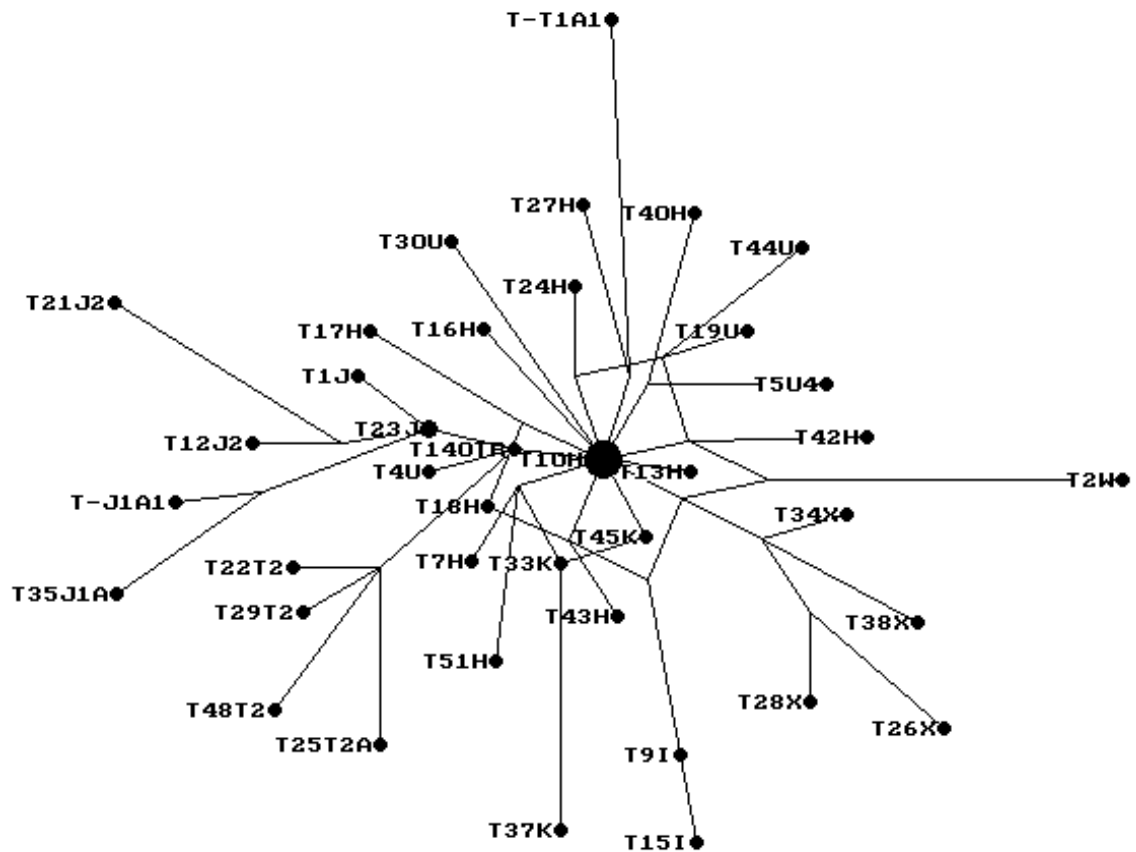
En la red media reducida del fragmento de 360pb construida para esta población (Fig. R-24), se observa también la diferenciación de las secuencias de los haplogrupos T, J y X. Sin embargo resulta extraño que la secuencia T-T1A1 (TOS32-T1a1) quede excluida del clúster formado por las secuencias restantes del haplogrupo T, contrariamente a lo observado en el árbol de N-J.

Un aspecto interesante de las redes, es la formación de reticulaciones que permiten observar relaciones filogenéticas alternativas en el mismo diagrama. En esta situación se encuentran las secuencias de los haplogrupos W e I, que podrían derivarse de la rama del haplogrupo X, pero que también pueden estar relacionadas con otras secuencias, principalmente del haplogrupo H. El conocimiento previo de las relaciones filogenéticas de estos haplogrupos nos da herramientas para considerar como más plausible la estructura que une a los haplogrupos X, W e I.

En una situación similar se encuentran las secuencias del haplogrupo K que también se encuentran involucradas en una relación reticular con secuencias del haplogrupo H, lo que haría difícil su distinción en el caso de que no se conociera el haplogrupo al que pertenecen. Por otra parte, estas reticulaciones apoyan la idea de que las secuencias del haplogrupo H son la que introducen un mayor “ruido de fondo” en el análisis. En este sentido es importante recordar que en el árbol filogenético también se perdía resolución entre las secuencias de los haplogrupos X, W e I debido a la presencia de varias secuencias del haplogrupo H (TOS16, TOS43 y TOS18).

La red del fragmento de 153pb (Fig. R-24) muestra características en concordancia con los resultados obtenidos para la población CAT. Las secuencias de los haplogrupos T, X, W e I continúan diferenciadas aunque al contrario de lo que ocurrió

con el árbol filogenético, la reducción del fragmento no mejora la agrupación de las secuencias W, X e I, señalando una desventaja de este método de análisis respecto a la reconstrucción filogenética de N-J a partir de las distancias de Tamura-Nei. Las secuencias del haplogrupo J pierden diferenciación, a excepción, nuevamente, de las secuencias del subgrupo J1a.



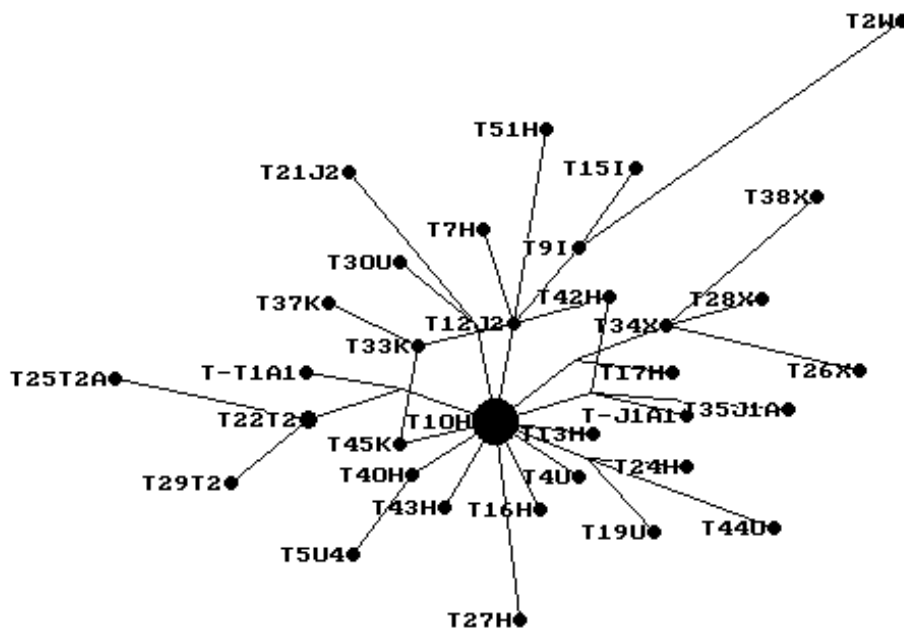


Figura R-24 Redes medias reducidas obtenidas para los fragmentos de 360 (arriba) y 153pb (abajo) de la población TOS. El tamaño de los círculos es proporcional a la frecuencia del haplotipo (ver Tabla R-49). El programa NETWORK sólo admite nombres de 7 caracteres en mayúscula, por lo que los nombres de los haplotipos han sido adaptados a esta limitación, substituyendo el prefijo TOS por la letra T. La secuencia T10H (TOS10-H) corresponde a la secuencia de referencia (CRS). La secuencia T-T1A1 corresponde a la secuencia TOS32-T1a1.

Los resultados del análisis de la población de Galura (GALU) muestran las mismas características que los resultados anteriores. En los árboles filogenéticos (Fig. R-25) los haplogrupos T y J resultan bien definidos para el árbol de 360pb y en el árbol de 153pb las secuencias del haplogrupo J pierden su definición pero no las del haplogrupo T.

En esta muestra no existen secuencias del subgrupo J1a, pero si del subgrupo U5, lo que resulta interesante pues en el análisis de las secuencias de la población CAT se observó que este subgrupo podría presentar una buena diferenciación en ambos segmentos. Los resultados de la población de Galura muestran que este subgrupo se diferencia bastante bien en el fragmento de 360pb y que no pierde del todo su diferenciación en el fragmento de 153pb, confirmando, en general, los resultados anteriores.

El problema lo presenta la secuencia G16-U5, que se sitúa en el grupo de las secuencias del haplogrupo T. Una inspección de las secuencias originales (Francalacci et al. 1999) nos muestra que este efecto puede ser debido a la sustitución 16304C que la presentan tanto la secuencia G16-U5 como las secuencias de algunos subgrupos del haplogrupo T. Los haplogrupos K e I no se diferencian quizá debido a que están representados por una única secuencia. En la población CAT, el haplogrupo K tampoco se diferenciaba y también estaba representado por un único haplotipo, aunque con una frecuencia mayor.

En la red media reducida del fragmento de 360pb calculada para esta población (Fig. R-26) se observa también la diferenciación de los haplogrupos T y J y del subgrupo U5. Asimismo, se observa la situación conflictiva de la secuencia G20-H con distintas posibilidades para explicar su estatus filogenético mostradas por las reticulaciones. Esta situación ya resultaba patente en los árboles filogenéticos, pues en el árbol de 360pb, esta secuencia se situaba en la bifurcación que separa los grupos T y J, al igual que lo indicado por una de las posibilidades que definen su situación en la red, en tanto que en el árbol de 153pb quedaba incluida en el subgrupo U5, como lo indica otra de las posibilidades de la red.

Analizando la red media reducida del fragmento de 153pb, es posible observar que el haplogrupo T continua diferenciado, al igual que el subgrupo U5, que conserva incluso la relación con la secuencia G16-U5, a diferencia de lo obtenido con el árbol de 153pb.

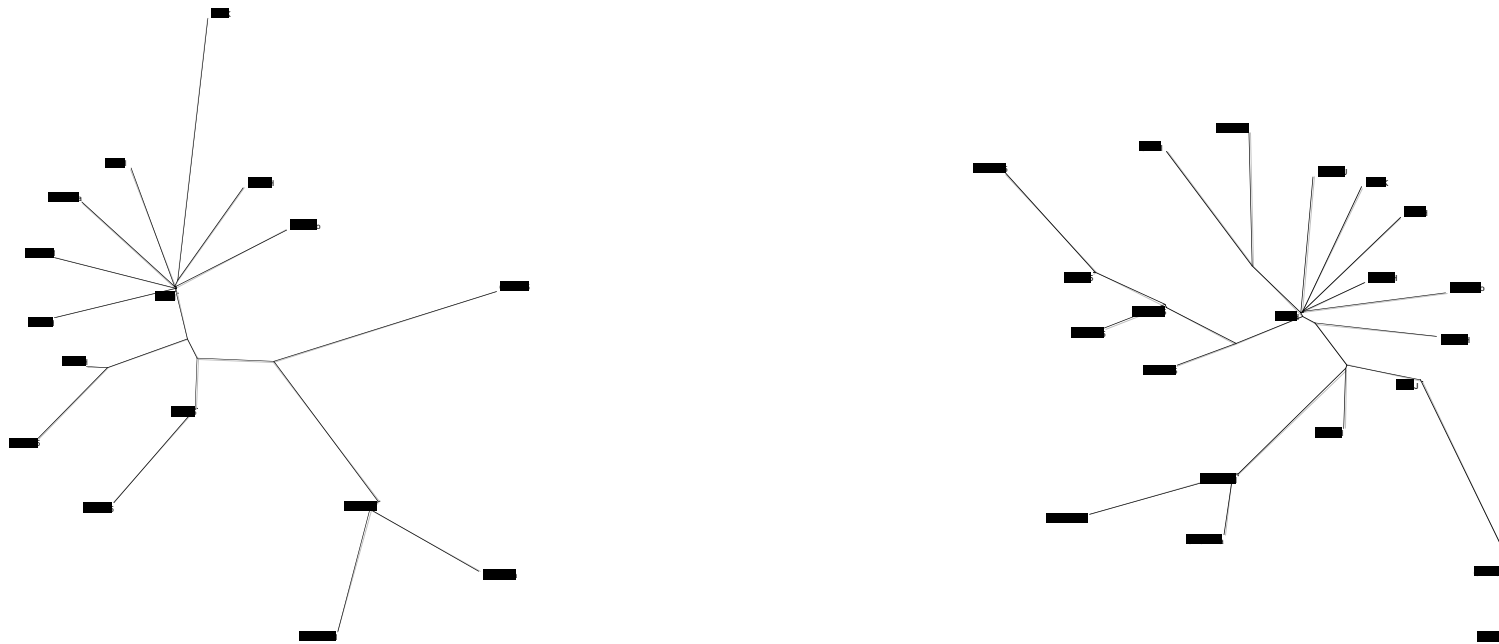


Figura R-25 Árboles filogenéticos construidos con el método de N-J a partir de la matriz de distancias de Tamura-Nei, para los fragmentos de 360 (izquierda) y 153pb (derecha) de la población GALU. Las secuencias idénticas fueron agrupadas según la Tabla R-49. La secuencia G4-H corresponde a la secuencia de referencia (CRS).

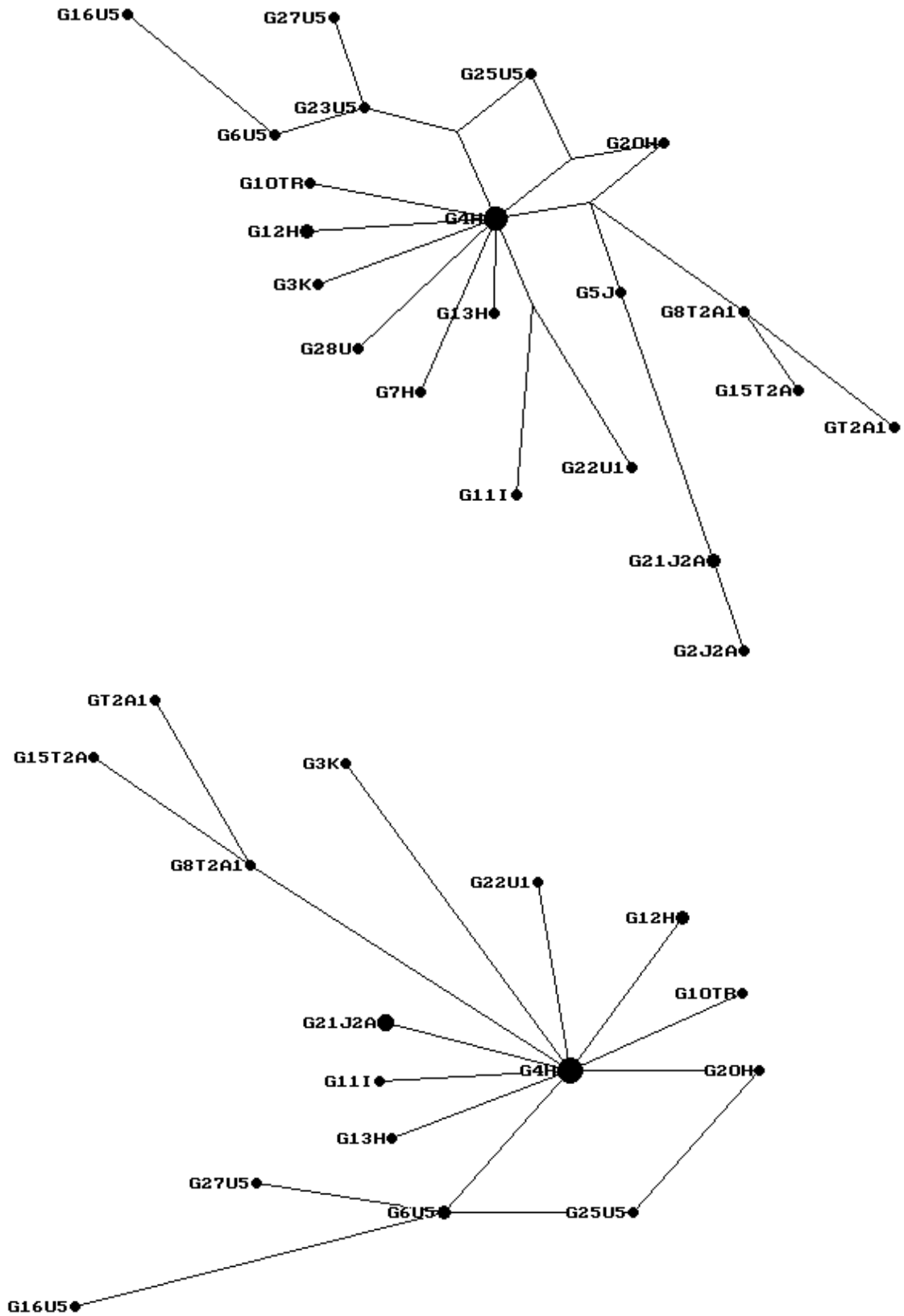


Figura R-26 Redes medias reducidas obtenidas para los fragmentos de 360 (arriba) y 153pb (abajo) de la población GALU. El tamaño de los círculos es proporcional a la frecuencia del haplotipo (ver Tabla R-49). La secuencia G4H (G4-H) corresponde a la secuencia de referencia (CRS).

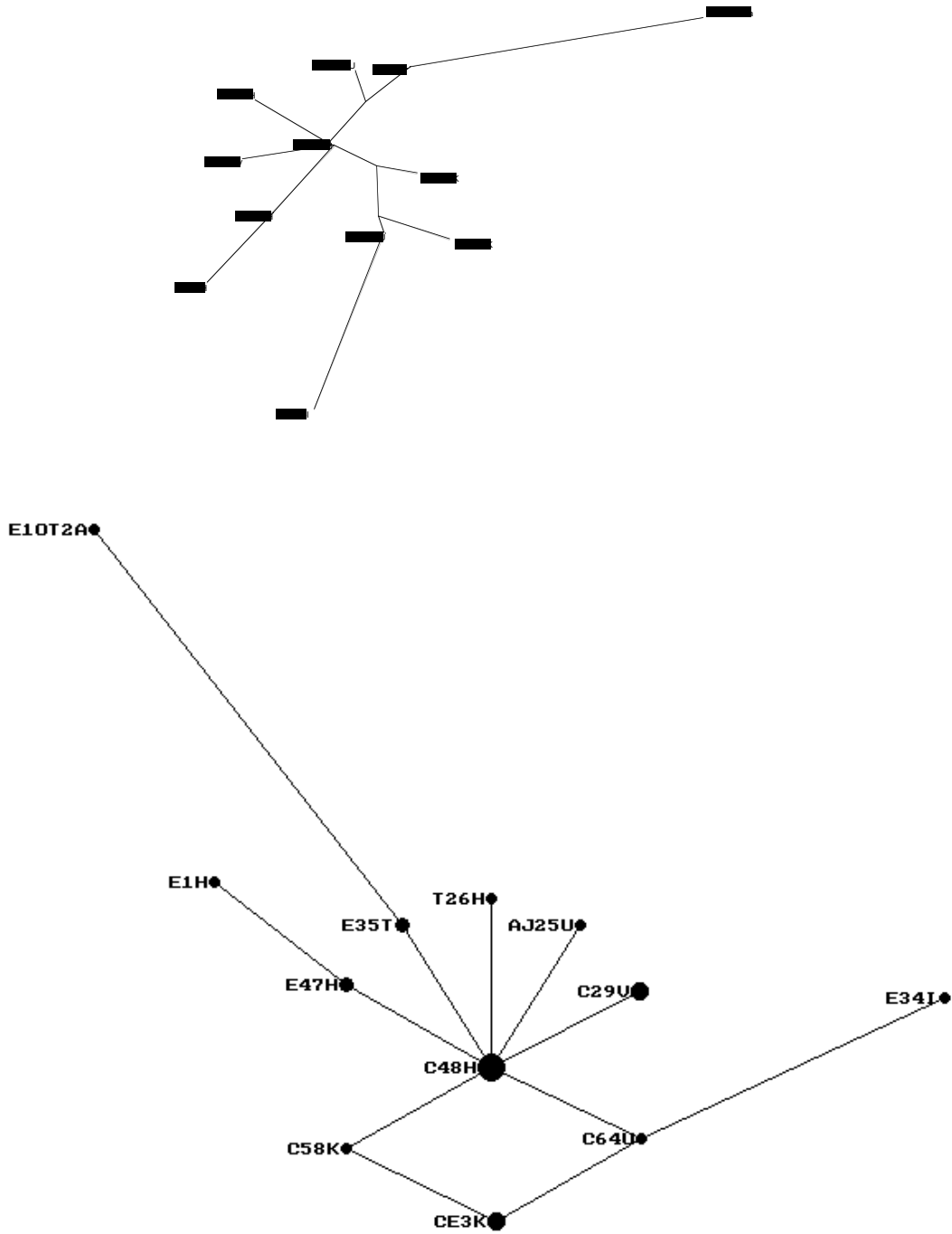


Figura R-27 Análisis filogenético de las secuencias de la Plaça Vella. Las secuencias idénticas fueron agrupadas según se muestra en la Tabla R-49. Arriba, árbol filogenético construido con el método de N-J a partir de la matriz de distancias de Tamura-Nei. Abajo, red media reducida en la que el tamaño de los círculos es proporcional a la frecuencia de los haplotipos. En la red se ha omitido el prefijo PV en los nombres de las secuencias. La secuencia PVC48-H (ó C48H en la red) es idéntica a la secuencia de referencia (CRS).

Por lo que respecta a la población de la Plaça Vella, al igual que para el resto de poblaciones, el árbol filogenético de N-J y la red media reducida presentan más o menos la misma estructura (Fig. R-27). Las secuencias del haplogrupo K presentan cierta diferenciación, pero se encuentran relacionadas con las secuencias PVC64-U y PVE34-I. Gran parte de esta estructuración se debe a la sustitución 16311C, pues la comparten las secuencias PVCE3-K, PVC64-U y PVE34-I. Se sabe que este sitio es inestable (Wakeley, 1993) y a pesar de haber repetido el análisis utilizando para él un peso equivalente al 10% del asignado a los sitios restantes durante la construcción de la red, continuó generando confusión. No obstante, el haplogrupo K es en realidad un subgrupo del haplogrupo U (Macaulay et al. 1999a), así que lo que resulta en verdad peculiar es la relación de estas secuencias con la secuencia del haplogrupo I. Sin embargo, el hecho de que sólo exista una secuencia del haplogrupo I puede ocasionar que su diferenciación no sea más evidente, aunque tanto en el árbol como en la red se sitúa en una posición relativamente distante. Las secuencias del haplogrupo V tampoco resultan diferenciadas probablemente debido a que están representadas por un único haplotipo (CAT29-V). Nuevamente, las secuencias del haplogrupo T sí presentan diferenciación, en tanto que las secuencias del haplogrupo H presentan la filogenia en forma de estrella similar a la observada en el resto de poblaciones analizadas.

Análisis filogenético interpoblacional

Distancias genéticas

En un primer análisis fue incluida la población subsahariana de Hadza, para comparar las distancias que presentan las poblaciones europeas entre sí, con la que presentan con esta población Africana. El resultado concuerda con lo obtenido por otros investigadores, ya que las poblaciones europeas resultan muy poco diferenciadas (Tabla R-50). La población CAT presenta un rango de diferencias con el resto de poblaciones europeas que va de 0.00000 (con BRI y TOS) hasta 0.02195 (con GALI). El árbol

filogenético construido a partir de la matriz de distancias de la Tabla R-50 muestra las poblaciones europeas en un grupo que se une con las poblaciones asiática y Hadza a través de la población de Turquía (Fig. R-28). En el conglomerado europeo se observa

	ASI	BRI	CAT	FIN	GALI	GALU	HAD	SAR	TOS	TRK
BRI	0.05917									
CAT	0.03955	0.00000								
FIN	0.02852	0.01024	0.01458							
GALI	0.08408	0.00196	0.02195	0.02028						
GALU	0.06308	0.00967	0.00505	0.02886	0.02543					
HAD	0.31725	0.61294	0.58232	0.64765	0.78611	0.67266				
SAR	0.05380	0.00235	0.00438	0.01077	0.01768	0.01730	0.60097			
TOS	0.03725	0.00000	0.00000	0.00834	0.00730	0.00251	0.53588	0.00073		
TRK	0.02035	0.00879	0.01211	0.00471	0.01956	0.02117	0.50291	0.01421	0.00000	
VAS	0.05799	0.00821	0.01788	0.01545	0.01110	0.01635	0.73457	0.01219	0.01173	0.03151

Tabla R-50 Distancias genéticas entre 11 poblaciones obtenidas mediante la transformación de Reynolds del estadístico F_{ST} . Calculadas para el fragmento de 360pb.

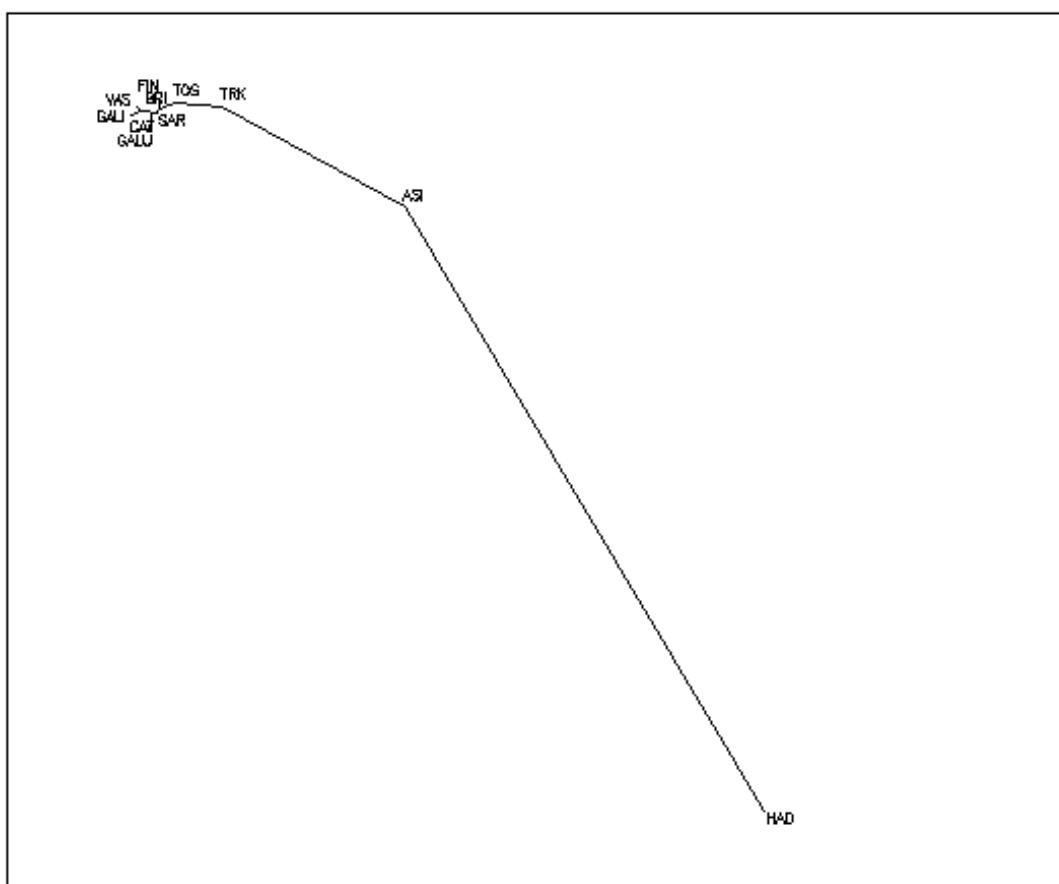


Figura R-28 Árbol filogenético construido con el método de N-J a partir de la matriz de distancias de la Tabla R-50. Se analizó el fragmento de 360pb de distintas poblaciones (descritas en la sección de Material y Métodos).

cierta diferenciación de las poblaciones VAS y GALI, así como de las poblaciones CAT y GALU que se sitúan en la misma rama. La estructura general del árbol muestra en las poblaciones europeas un gradiente de este a oeste, con la población Turca en un extremo y la de Galicia y País Vasco en el otro. No obstante, la pequeña rama que conduce a las poblaciones CAT y GALU, indica que dentro de este gradiente existen ciertas diferenciaciones.

Las distancias basadas en las transformaciones del estadístico F_{ST} son más adecuadas para poblaciones con distancias genéticas pequeñas (Reynolds et al. 1983; Slatkin, 1995). Por este motivo se construyó un árbol filogenético excluyendo la población Hadza. Este árbol (Fig. R-29) muestra nuevamente el gradiente este – oeste, pero se hace más patente la diferenciación que presentan algunas poblaciones. Concretamente, se observa que las poblaciones CAT y GALU se separan en una misma rama, confirmando lo observado en el árbol de la Figura R-28. Por su parte, las poblaciones Vasca y Gallega presentan una diferenciación individual, al igual que la población mixta de Cerdeña que muestra una diferenciación independiente que notablemente no está relacionada con la población Galuresa del norte de la isla (ver Material y Métodos).

Así pues, estos resultados nos muestran que en Europa existe, a pesar de la relativa homogeneidad, cierta estructura debida quizá a la historia demográfica particular de algunas poblaciones. Estas diferencias no han sido encontradas por otros investigadores utilizando distancias como la de Nei y Miller aplicada por Francalacci y col. (1996) o la distancia definida por las diferencias medias por parejas dentro y entre las poblaciones analizadas (ej. Bertranpetit et al. 1995), con las que incluso se han llegado a obtener distancias negativas entre algunas poblaciones. No obstante, la distancia aquí aplicada basada en el estadístico F_{ST} , no es del todo óptima cuando se analizan secuencias de DNA, porque dicho estadístico no hace un uso completo de la información contenida en los datos de secuencias (Slatkin, 1991). A pesar de que esta deficiencia puede ser disminuida aplicando una distancia apropiada entre los haplotipos, como la utilizada aquí basada en el modelo de Tamura-Nei con heterogeneidad de la tasa

de mutación (ver Material y Métodos), sería recomendable investigar, en trabajos posteriores, aproximaciones alternativas como el método cladístico de Slatkin y Maddison (1990).

Continuando con la tónica de este trabajo, se realizó la reconstrucción filogenética utilizando fragmentos más pequeños, para incluir más poblaciones y para observar el efecto de la reducción del fragmento analizado en el resultado de la reconstrucción. En primer término, se calculó la matriz de distancias entre todas las poblaciones analizadas, excepto la población HAD, utilizando el fragmento de 302pb.

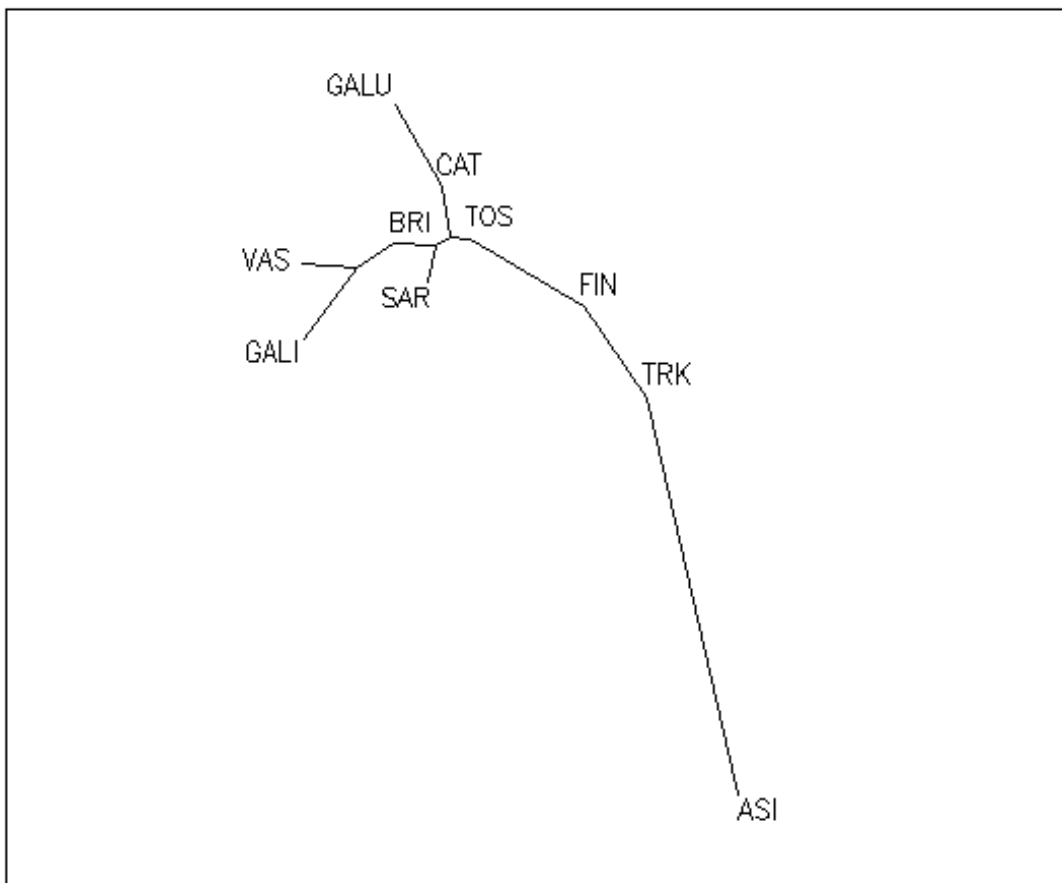


Figura R-29 Árbol filogenético de Neighbor-Joining del fragmento de 360pb, construido a partir de la matriz de distancias de la Tabla R-50, excluyendo la población Hadza.

El resultado tanto de la matriz de distancias (Tabla R-51) como del árbol filogenético (Fig. R-30a) es sorprendentemente similar al obtenido con el fragmento de 360pb, indicando que el fragmento de 302pb contiene casi toda la información

filogenética contenida en el fragmento de 360pb, al menos la que resulta relevante para la reconstrucción filogenética con los métodos aquí utilizados.

De hecho, dado que los índices de diversidad nucleotídica aumentan en los fragmentos de 302 y 153 pb, estos fragmentos contienen relativamente más información que el segmento grande y sólo queda la duda de si la calidad de la información es equiparable o si existe en el segmento grande alguna zona decisiva para la definición de la topología de los árboles filogenéticos. Como hemos visto, el segmento de 302pb no muestra signos de contener una menor calidad de información.

	ASI	BRI	CAT	FIN	GALI	GALU	SAR	TOS	TRK
BRI	0.05410								
CAT	0.03731	0.00016							
FIN	0.02476	0.01024	0.01426						
GALI	0.07925	0.00237	0.02134	0.02017					
GALU	0.06240	0.01087	0.00627	0.02993	0.02581				
SAR	0.05069	0.00221	0.00365	0.01116	0.01828	0.01830			
TOS	0.03451	0.00000	0.00000	0.00836	0.00664	0.00217	0.00077		
TRK	0.01700	0.00790	0.01167	0.00442	0.01905	0.02190	0.01480	0.00000	
VAS	0.05412	0.00853	0.01758	0.01527	0.01129	0.01664	0.01268	0.01125	0.03160

Tabla R-51 Distancias genéticas entre 10 poblaciones, obtenidas mediante la transformación de Reynolds del estadístico F_{ST} , calculadas para el fragmento de 302pb.

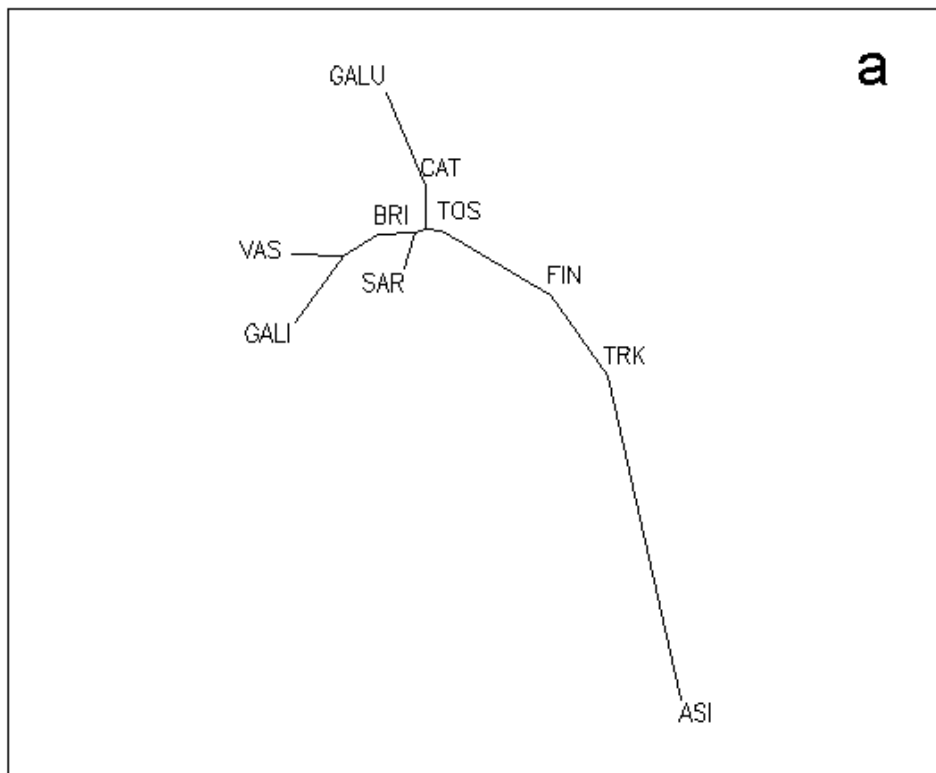
	ARG	BRI	CAT	FIN	GALI	GALU	SAR	TOS	TRK
BRI	0.11267								
CAT	0.11890	0.00016							
FIN	0.10959	0.01023	0.01426						
GALI	0.13307	0.00236	0.02134	0.02017					
GALU	0.13393	0.01087	0.00627	0.02993	0.02581				
SAR	0.08644	0.00220	0.00365	0.01116	0.01827	0.01830			
TOS	0.08808	0.00000	0.00000	0.00836	0.00663	0.00217	0.00077		
TRK	0.10290	0.00789	0.01167	0.00442	0.01904	0.02190	0.01480	0.00000	
VAS	0.11429	0.00852	0.01758	0.01527	0.01130	0.01664	0.01268	0.01124	0.03160

Tabla R-52 Distancias genéticas entre 10 poblaciones obtenidas mediante la transformación de Reynolds del estadístico F_{ST} , calculadas para el fragmento de 302 pb.

Por otra parte, el análisis de este segmento nos permite incluir la población beréber de Argelia. El árbol filogenético resultante (Fig. R-30b) construido a partir de la matriz de distancias de la Tabla R-52, muestra que esta población se distancia considerablemente de las poblaciones europeas, uniéndose al grupo por la rama de la

población Sarda. Es notable que esta unión no se realice a través de ninguna de las poblaciones ibéricas, lo que indica que la influencia de éstas en la población Beréber no ha sido considerable.

Para situar las relaciones de la población ARG con las poblaciones europeas en un contexto de mayor amplitud, se han calculado las distancias genéticas incluyendo las poblaciones ASI y HAD (Tabla R-53). El árbol filogenético resultante (Fig. R-31) muestra que la población ARG se sitúa en una rama que surge de un punto intermedio entre la población ASI y la población TRK. Este resultado está en completo acuerdo con la posibilidad destacada por Rando y col. (1998) de que la misma expansión ocurrida en el sur de Europa por la costa norte del mediterráneo, pudo haber seguido simultáneamente la ruta de la costa sur, finalizando en el norte de África. Es necesario investigar más sobre este punto incluyendo más poblaciones de la costa sur del mediterráneo.



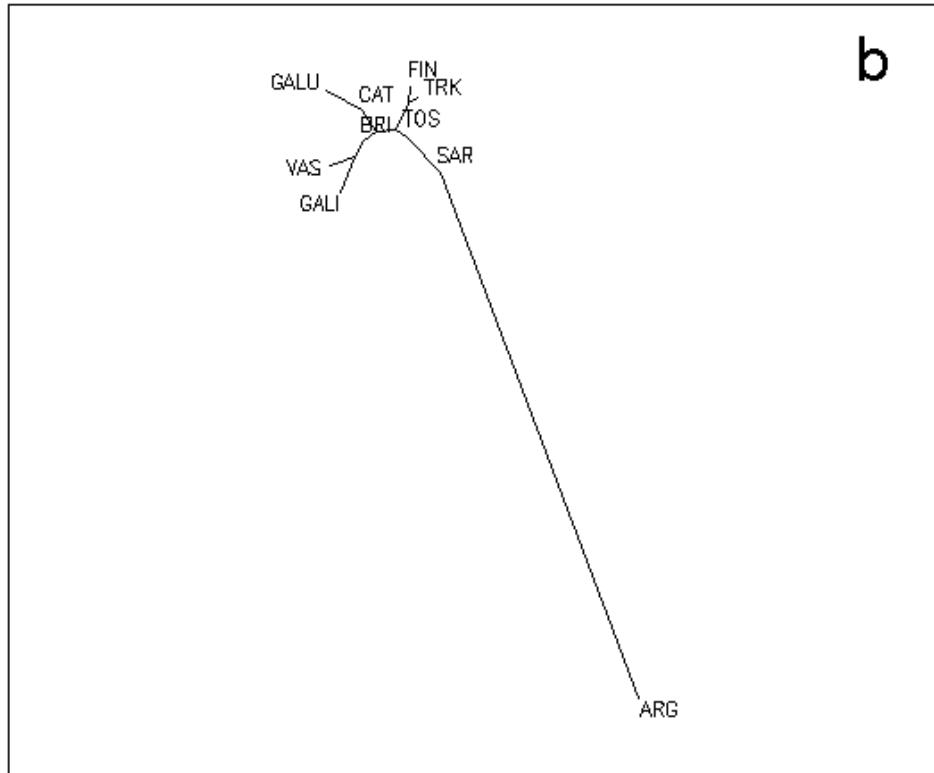


Figura R-30 Árboles filogenéticos construidos con el método de N-J del fragmento de 302pb; **(a)**, utilizando la matriz de distancias de la Tabla R-51; **(b)**, a partir de la matriz de la Tabla R-52.

	ARG	ASI	BRI	CAT	FIN	GALI	GALU	HAD	TOS	TRK
ASI	0.08120									
BRI	0.11267	0.05409								
CAT	0.11890	0.03731	0.00015							
FIN	0.10959	0.02477	0.01023	0.01426						
GALI	0.13307	0.07925	0.00237	0.02133	0.02017					
GALU	0.13393	0.06240	0.01086	0.00627	0.02993	0.02581				
HAD	0.46357	0.31056	0.60558	0.59403	0.63799	0.77819	0.67635			
TOS	0.08807	0.03451	0.00000	0.00000	0.00836	0.00663	0.00217	0.52314		
TRK	0.10290	0.01700	0.00789	0.01167	0.00442	0.01904	0.02190	0.49471	0.00000	
VAS	0.11429	0.05413	0.00852	0.01758	0.01527	0.01129	0.01664	0.72270	0.01124	0.03160

Tabla R-53 Distancias genéticas entre 11 poblaciones obtenidas mediante la transformación de Reynolds del estadístico F_{ST} ; calculadas para el fragmento de 302 pb.

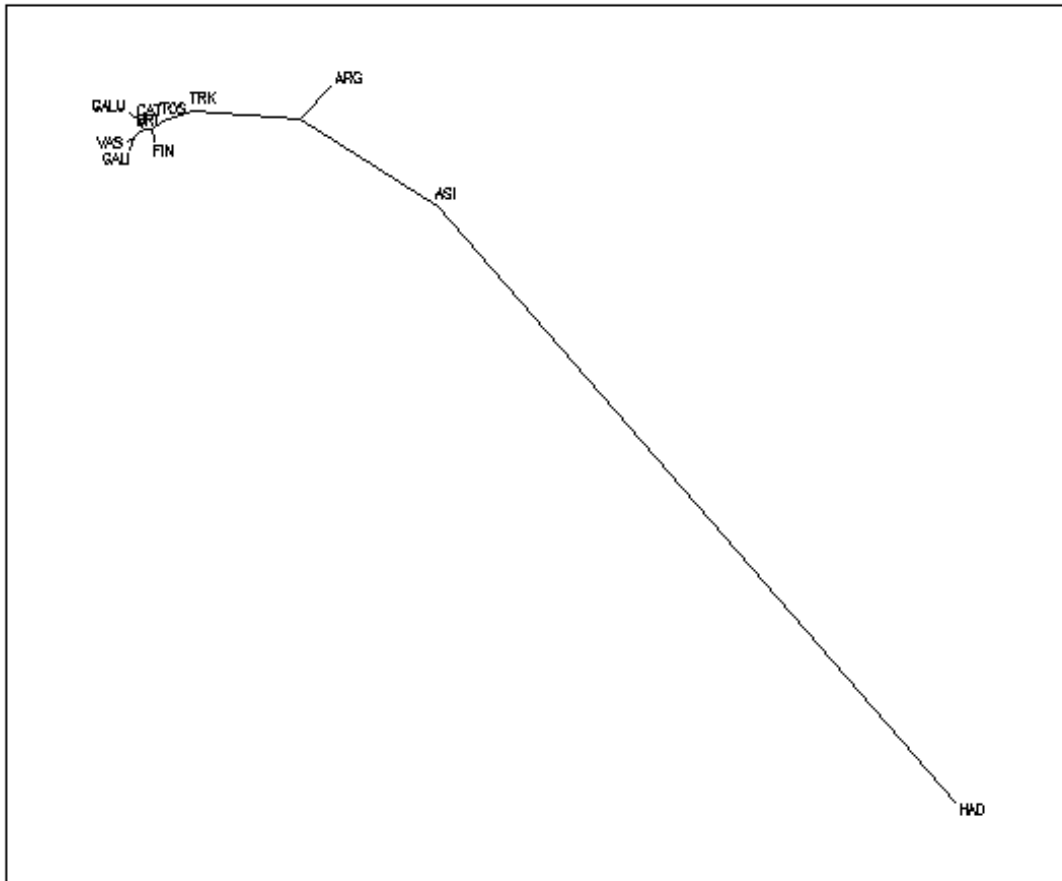


Figura R-31 Árbol filogenético sin raíz construido con el método de N-J del segmento de 302pb utilizando las distancias de la Tabla R-53.

Finalmente se analizó el segmento de 153pb para poder incluir los datos de la población de la Plaça Vella. Utilizando las 10 poblaciones europeas y la población asiática, el árbol filogenético resultante (Fig. R-32a), construido a partir de la matriz de distancias de la Tabla R-54, muestra únicamente dos cambios importantes respecto a los obtenidos con los fragmentos de 360 y 302 pb.

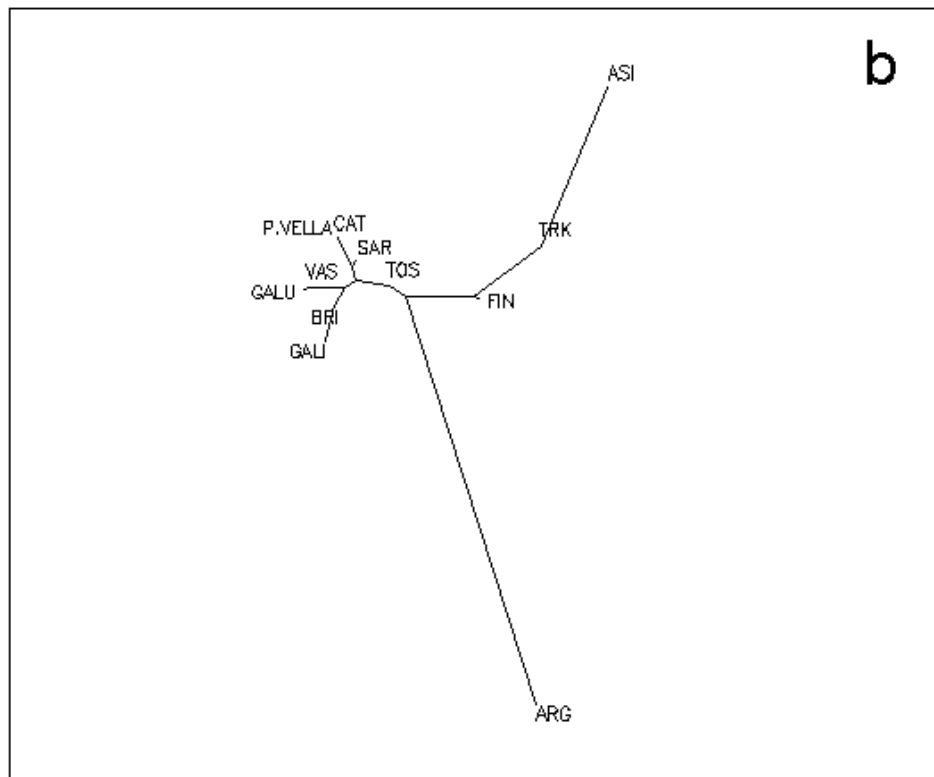
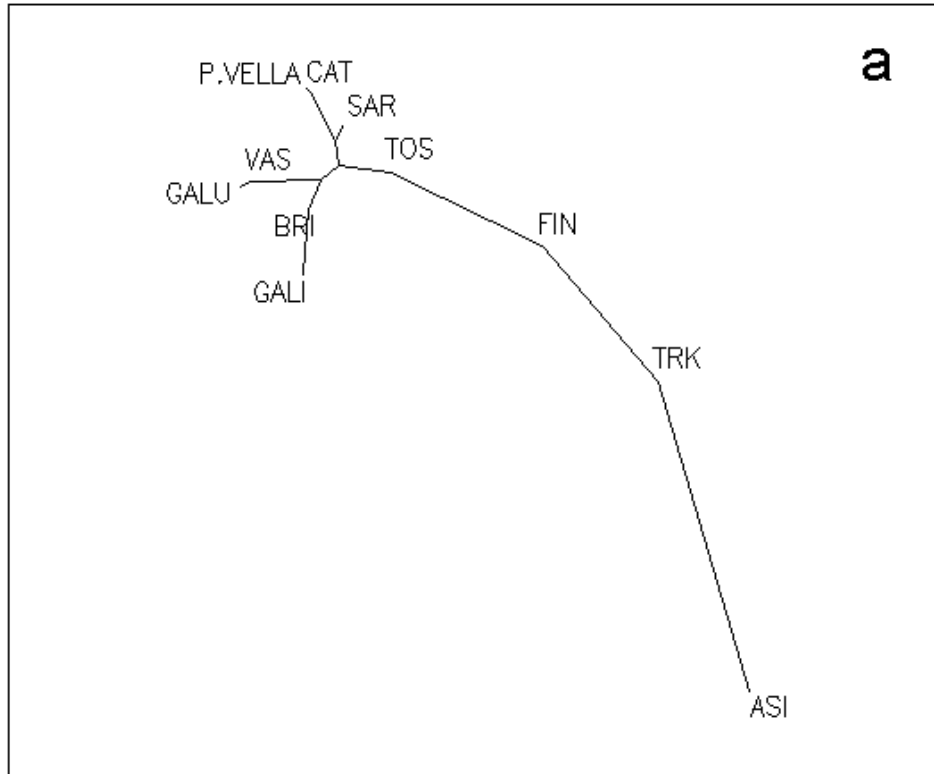


Figura R-32 Árboles filogenéticos sin raíz construidos con el método de N-J usando el fragmento de 153pb; **(a)**, utilizando la matriz de distancias de la Tabla R-54; **(b)**, a partir de las distancias de la Tabla R-55.

Estos cambios afectan a las poblaciones de Cerdeña y Galura. La primera se sitúa en la rama que conduce a las poblaciones Catalanas en tanto que la segunda se une con la población VAS. Es importante recordar que estas poblaciones demostraban cambios en el orden de diversidad al analizar el fragmento de 153pb, lo que explica este cambio de posición en el árbol al analizar este fragmento. Este resultado demuestra la importancia de contar con un marco de referencia de la estructura mitocondrial de las poblaciones antes de iniciar un análisis de DNA antiguo. Para el caso particular de las poblaciones GALU y SAR, recuperar el fragmento de 153pb no sería suficiente para un análisis de reconstrucción filogenética. Otra característica importante de este árbol, lo constituye la posición de la población de la Plaça Vella, que se sitúa en la misma rama que la población actual de Cataluña con muy poca diferenciación. Aunque el análisis de las frecuencias de haplogrupos no mostraba esta cercana relación, el contenido haplotípico indica que las muestras están cercanamente relacionadas. En otras palabras, que sus linajes mitocondriales están filogenéticamente cercanos. Este resultado alcanza mayor relevancia en tanto que el análisis de la variabilidad nucleotídica en la población actual de Cataluña no mostró inconsistencias entre los distintos tamaños de segmentos analizados; indicando que para esta población es válido el análisis del fragmento de 153pb aquí utilizado.

Al igual que para el fragmento de 302pb, se ha incluido la población ARG para observar su relación con las poblaciones europeas cuando se analiza el fragmento de 153pb. El árbol obtenido (Fig. R-32b) a partir de la matriz de distancias de la Tabla R-55, muestra que la población de Argelia se une con las poblaciones europeas en un punto cercano a la población Toscana, sin embargo, al incluir la población Hadza (Tabla R-56; Figura R-33) se observa que la población beréber se une por el mismo punto por el que se une si se analiza el segmento de 302pb, entre las poblaciones ASI y TRK.

	ASI	BRI	CAT	FIN	GALI	GALU	P.VELLA	SAR	TOS	TRK
BRI	0.07695									
CAT	0.05565	0.00532								
FIN	0.03923	0.02076	0.02512							
GALI	0.09925	0.00000	0.02699	0.02463						
GALU	0.07345	0.00514	0.00847	0.02654	0.01720					
P.VELLA	0.06426	0.00000	0.00000	0.02993	0.01721	0.03243				
SAR	0.08747	0.00066	0.00702	0.02639	0.01307	0.01253	0.00558			
TOS	0.04443	0.00000	0.00000	0.01273	0.00478	0.00418	0.00346	0.00011		
TRK	0.01113	0.01542	0.03042	0.00274	0.02091	0.03577	0.02607	0.02106	0.00211	
VAS	0.07219	0.00638	0.01132	0.02476	0.01194	0.00075	0.02315	0.01380	0.00774	0.03097

Tabla R-54 Distancias genéticas entre 11 poblaciones obtenidas mediante la transformación de Reynolds del estadístico F_{ST} ; calculadas para el fragmento de 153 pb.

	ARG	ASI	BRI	CAT	FIN	GALI	GALU	P.VELLA	SAR	TOS	TRK
ASI	0.09793										
BRI	0.08618	0.07695									
CAT	0.09400	0.05566	0.00533								
FIN	0.09972	0.03925	0.02075	0.02512							
GALI	0.09339	0.09928	0.00000	0.02700	0.02464						
GALU	0.08537	0.07347	0.00515	0.00846	0.02654	0.01721					
P.VELLA	0.08379	0.06427	0.00000	0.00000	0.02992	0.01720	0.03242				
SAR	0.08190	0.08750	0.00065	0.00703	0.02640	0.01307	0.01252	0.00557			
TOS	0.06100	0.04444	0.00000	0.00000	0.01274	0.00478	0.00418	0.00345	0.00012		
TRK	0.07305	0.01114	0.01542	0.03042	0.00275	0.02092	0.03577	0.02607	0.02107	0.00212	
VAS	0.07837	0.07222	0.00638	0.01132	0.02476	0.01194	0.00074	0.02314	0.01381	0.00775	0.03097

Tabla R-55 Distancias genéticas entre 12 poblaciones obtenidas mediante la transformación de Reynolds del estadístico F_{ST} ; calculadas para el fragmento de 153 pb.

Como se había expresado anteriormente, para clarificar las relaciones entre esta población y las poblaciones europeas es necesario el análisis de más poblaciones. Por el momento, estos resultados son compatibles con los análisis llevados a cabo por Rando y col. (1998).

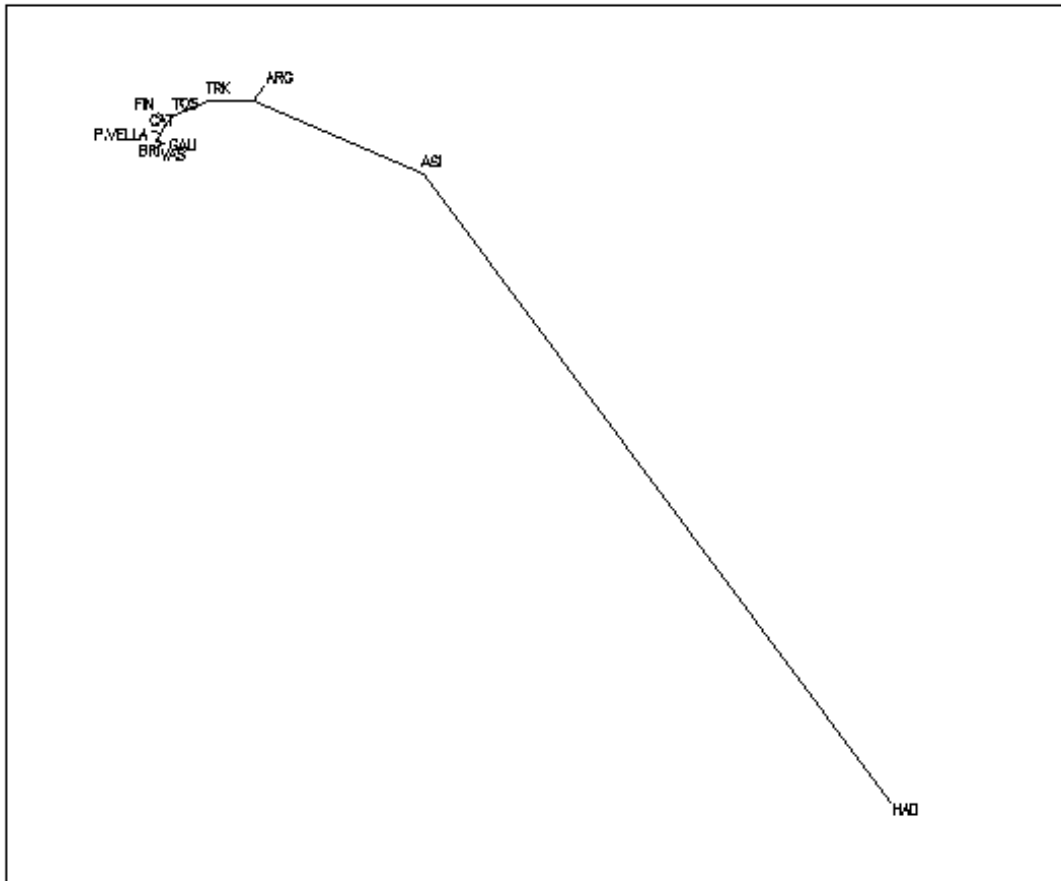


Figura R-33 Árbol filogenético sin raíz construido con el método de N-J usando el fragmento de 153pb a partir de la matriz de distancias de la Tabla R-56.

	ARG	ASI	BRI	CAT	FIN	GALI	HAD	P.VELLA	TOS	TRK
ASI	0.09794									
BRI	0.08618	0.07696								
CAT	0.09398	0.05565	0.00533							
FIN	0.09972	0.03927	0.02075	0.02511						
GALI	0.09339	0.09933	0.00000	0.02702	0.02465					
HAD	0.64904	0.43284	0.79912	0.74497	0.80984	1.04318				
P.VELLA	0.08378	0.06429	0.00000	0.00000	0.02992	0.01721	0.85144			
TOS	0.06100	0.04445	0.00000	0.00000	0.01275	0.00479	0.67991	0.00345		
TRK	0.07305	0.01115	0.01542	0.03041	0.00276	0.02092	0.62406	0.02607	0.00213	
VAS	0.07837	0.07224	0.00638	0.01132	0.02478	0.01196	0.91707	0.02314	0.00776	0.03098

Tabla R-56 Distancias genéticas entre 11 poblaciones obtenidas mediante la transformación de Reynolds del estadístico F_{ST} , calculadas para el fragmento de 153 pb.

Como nota final, cabe destacar que en los tres árboles construidos con el fragmento de 153pb, las poblaciones P.VELLA y CAT, aparecen siempre como un grupo monofilético. Sin embargo, es necesario considerar también que dada la naturaleza exploratoria de este trabajo no se ha realizado un análisis exhaustivo de la robustez de los

árboles aquí presentados; análisis que inevitablemente debemos dejar para futuras investigaciones.

Relaciones entre los haplotipos de las poblaciones

Como se ha observado en el análisis intrapoblacional, las secuencias de algunos haplogrupos pueden introducir un importante ruido de fondo si se analizan junto con las secuencias de otros haplogrupos. Para evitar estas confusiones se realizó el análisis separando previamente las secuencias obtenidas de la base de datos de acuerdo al haplogrupo al que pertenecen, según el análisis de restricción o por medio de la correspondencia secuencia-haplogrupo. La mayoría de haplogrupos presentan substituciones diagnósticas bien definidas, sin embargo, algunas secuencias de los haplogrupos H y U no pueden ser distinguidas. Por tanto, el conjunto de datos del haplogrupo H puede contener cierto número de secuencias pertenecientes al haplogrupo U, aunque en baja proporción.

Por otra parte, el conjunto correspondiente a las secuencias del haplogrupo U podría carecer de algunas secuencias (las incluidas en el haplogrupo H), sin embargo es poco probable que contenga secuencias del haplogrupo H, ya que todos los subgrupos del haplogrupo U presentan substituciones características. En cualquier caso, las secuencias en las que la asignación resultaba dudosa, fueron señaladas con un signo de interrogación (?).

En este análisis se utilizó el fragmento de 302pb, para poder incluir los datos de la población de Argelia. En total se analizaron 588 secuencias procedentes de 10 poblaciones, entre las que se encontraron 254 haplotipos distintos para este fragmento (302pb).

En la Figura R-34 se presenta el árbol filogenético del haplogrupo H. La estructura del árbol concuerda con lo observado por otros investigadores, como la forma de estrella y la presencia de diversos subgrupos en cuya raíz se suele observar un

haplotipo frecuente. Richards y col. (1998) observaron subgrupos derivados de los haplotipos 16129, 16291, 16304, 16311 y 16362 (transiciones). En el árbol de la Figura R-34 es posible identificar los grupos de los haplotipos 16129 (*BCSK), 16291 (*BFBKV), 16304 (*BCISTKV) y 16311 (*AISKV). Además el haplotipo caracterizado por la transición 16172 encontrado en dos individuos de la población de Galicia (representados por GALI34) podría también ser la base de un subgrupo diferenciado. Las transiciones que caracterizan estos subgrupos podrían ser de las más antiguas del haplogrupo H, no obstante, de acuerdo con Richards y col. (1998), por el momento no es posible nombrar estos subgrupos hasta que no se tenga más información, además de que los sitios 16129 y 16311 han sido identificados como sitios con una tasa de mutación alta (Richards et al. 1998).

De las 7 secuencias del haplogrupo H encontradas en la población CAT, 6 son diferentes para el segmento de 302pb. La secuencia CAT21 resultó ser única y se encuentra situada en un grupo formado por ramas profundas, asociada más cercanamente a una secuencia de Galura (G12-H). En la bifurcación que separa estas dos secuencias se encuentra el haplotipo *ISV-H caracterizado por la transición 16093 y compartido por las poblaciones de Galicia, Cerdeña y País Vasco. En el mismo grupo pero en una rama un poco más distante se encuentra otra de las secuencias de la población catalana (CAT58), representada por el haplotipo *BCSK. Otra de las secuencias (CAT15) forma parte del haplotipo compartido *BCFS que se encuentra en otro grupo de ramas profundas. Asimismo, la secuencia CAT20 está representada por el haplotipo compartido *BCISTKV, que se encuentra en una rama que surge del grupo principal. Dos de las secuencias (CAT1 y CAT60) son idénticas a la secuencia de referencia (CRS), que es la más común y que se encuentra en todas las poblaciones analizadas, incluyendo la población de Argelia. Finalmente, la secuencia CAT63 está representada por el haplotipo *CI, que no es distinguible en el árbol por que se encuentra en una de las múltiples ramas cortas que surgen del núcleo central y su nombre queda cubierto por el del haplotipo GALI105.

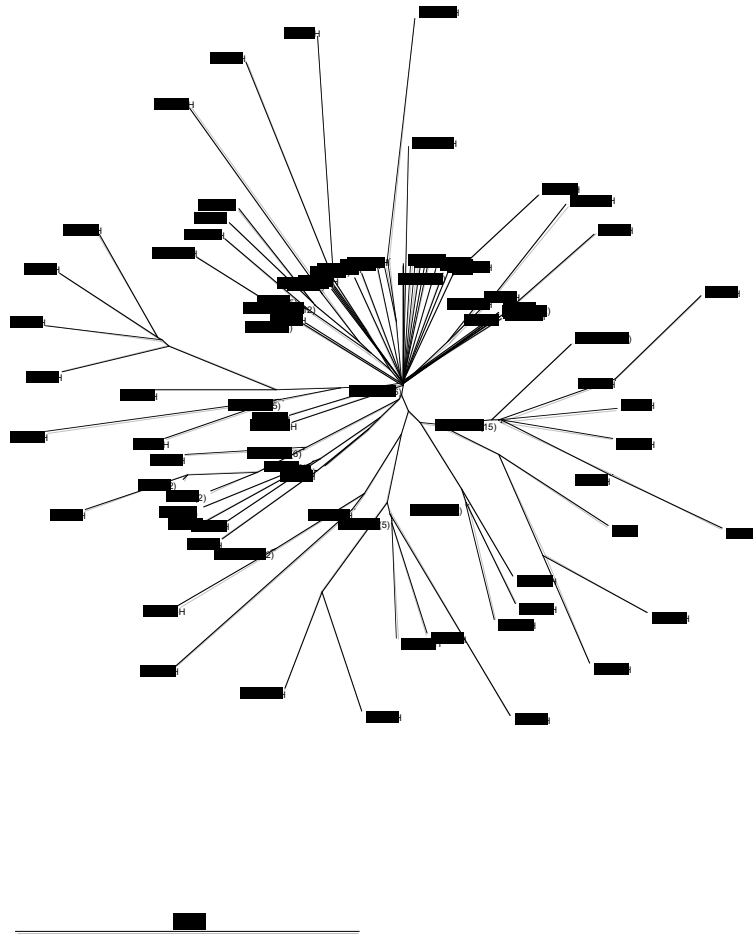


Figura R-34 Árbol filogenético de N-J de las secuencias del haplogrupo H construido a partir de las distancias de Tamura-Nei. El análisis comprende 93 haplotipos distintos encontrados en 251 individuos procedentes de 10 poblaciones. La frecuencia observada de los haplotipos se indica mediante un número entre paréntesis. Los haplotipos compartidos entre poblaciones están marcados con un asterisco * y la frecuencia de estos haplotipos en cada población se presenta en la Tabla R-57.

En general, no parece haber una diferenciación clara de las secuencias de la población catalana, pero si llama la atención el hecho de que ésta no esté representada por ningún haplotipo en los dos grupos que presentan una estructura más definida. Estos dos grupos pueden subdividirse en 5 subgrupos, cuatro de ellos identificados por los haplotipos 16311 (*AISKV), 16172 (*GALI34), 16291 (*BFKV) y 16278 (*GALI118); en tanto que el grupo formado por los haplotipos G20, TOS51 y TRK18, está probablemente caracterizado por la transición 16256, ya que la presentan las tres muestras. El número de secuencias de la población CAT (n=25) es relativamente bajo, lo que podría explicar esta falta de representación en estos grupos. No obstante, la muestra de Galura, con un número de secuencias similar (n=27), si está representada aunque sólo

por un haplotipo (G20). Sería necesario ampliar la muestra de la población CAT, para poder obtener conclusiones definitivas en esta cuestión. Sin embargo, cabe hacer notar que las secuencias del haplogrupo H de la población antigua de la Plaça Vella, tampoco presentan secuencias que contengan sustituciones en los sitios que definen estos grupos.

Pob.	N	*CRS	*AB	*AISKV	*BCISTKV	*BCFS	*BFKV	*BT	*BCSK	*BF	*BS	*CI	*IV	*ISV	*US
ARG	85	7	4	7	-	-	1	-	-	-	-	-	-	-	-
BRI	100	14	1	-	2	2	-	1	3	1	1	-	-	-	-
CAT	25	2	-	-	1	1	-	-	1	-	-	1	-	-	-
FIN	47	6	-	-	-	1	2	-	-	1	-	-	-	-	-
GALI	92	25	-	2	2	-	-	-	-	-	-	1	1	1	-
GALU	27	5	-	-	-	-	-	-	-	-	-	-	-	-	1
SAR	69	17	-	3	1	1	-	-	1	-	1	-	-	1	3
TOS	48	9	-	-	1	-	-	1	-	-	-	-	-	-	-
TRK	45	2	-	1	2	-	1	-	1	-	-	-	-	-	-
VAS	45	9	-	2	3	-	1	-	-	-	-	-	1	1	-
Total		96	5	15	12	5	5	2	6	2	2	2	2	3	4

Tabla R-57 Frecuencia de los haplotipos comunes del haplogrupo H que presenta cada una de las poblaciones analizadas. La nomenclatura utilizada se describe en la sección de Material y Métodos.

El haplogrupo V, probablemente derivado del haplogrupo H (Torrioni et al. 1998b), también presenta una filogenia en forma de estrella (Figura R-35) debido quizá a que participó en la misma expansión post-glacial que el haplogrupo H (Richards et al. 1998). Como se ha expuesto en el marco de referencia, el Norte de Africa, Cataluña y Escandinavia (en la población Saami) son las tres áreas en las que se pudo haber originado este haplogrupo (Torrioni et al. 1998b) y por este motivo se incluyeron las secuencias de la población Saami (Sajantila et al. 1995) en la construcción de este árbol (ver Material y Métodos). El haplotipo *ABCFIVM se sitúa en el centro de la filogenia (Figura R-35); está caracterizado por una única transición en la posición 16298 y es, con mucho, el más frecuente (55/79). En general existe muy poca variabilidad en este haplogrupo, pues sólo hemos encontrado 16 secuencias diferentes en los 79 individuos analizados. Después del haplotipo central, los siguientes haplotipos más comunes presentan una frecuencia mucho menor (5/79), pero resulta interesante que uno de ellos esté compartido únicamente por las poblaciones Saami y Argelina, indicando que se trata

de una variante antigua o que ha surgido independientemente en ambas poblaciones. Los dos haplotipos más distantes (SSA2 y SSA14) provienen de la población Saami pero es difícil discernir si se trata de tipos antiguos ya que esta población ha sufrido un drástico efecto fundador y de deriva (Sajantila et al. 1995). Una de las secuencias de la población catalana (CAT75) es idéntica a la secuencia más común (*ABCFIVM), en tanto que la otra (CAT3) es única. En la muestra de la Plaça Vella se encontraron 3 individuos del haplogrupo V, pero los tres presentan la misma secuencia, idéntica al haplotipo *ABCFIVM.

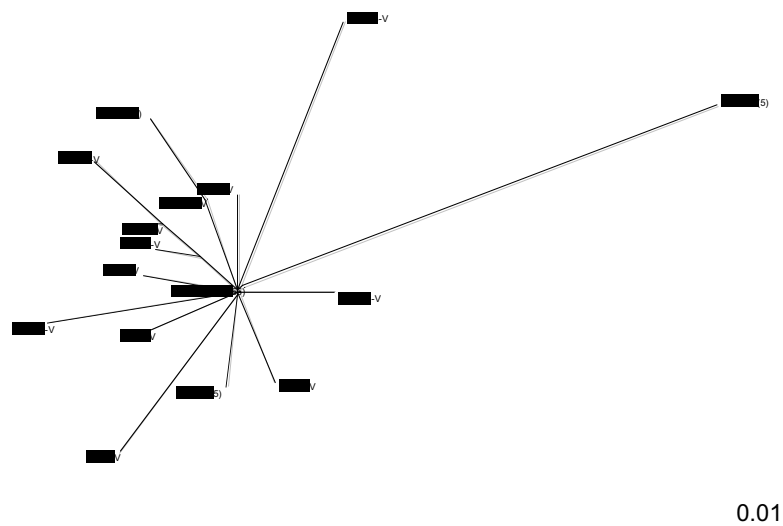


Figura R-35 Árbol filogenético de N-J de las secuencias del haplogrupo V construido a partir de las distancias de Tamura-Nei. El análisis comprende 16 haplotipos distintos encontrados en 79 individuos procedentes de 7 poblaciones. La frecuencia observada de los haplotipos se indica mediante un número entre paréntesis. Los haplotipos compartidos entre poblaciones están marcados con un asterisco *. Las frecuencias de estos haplotipos en las distintas poblaciones son las siguientes: *AM-V1 (ARG 1, FSA 4); *ABCFIVM-V (ARG 4, BRI 2, CAT 1, FIN 2, GALI 2, VAS 3 y SSA 41).

El árbol de las secuencias del haplogrupo K presenta también una filogenia en forma de estrella (Figura R-36), que podría deberse a que este haplogrupo haya participado también en la expansión reflejada en los haplogrupos H y V, propuesta por algunos investigadores (Richards et al. 1998). Este haplogrupo en realidad es un subhaplogrupo del U, pero dada su filogenia en forma de estrella hemos preferido analizarlo por separado. En el árbol se observa un subgrupo (K1) que podría indicar que este haplogrupo tiene una antigüedad mayor que la del haplogrupo V pero mucho menor que la del H.

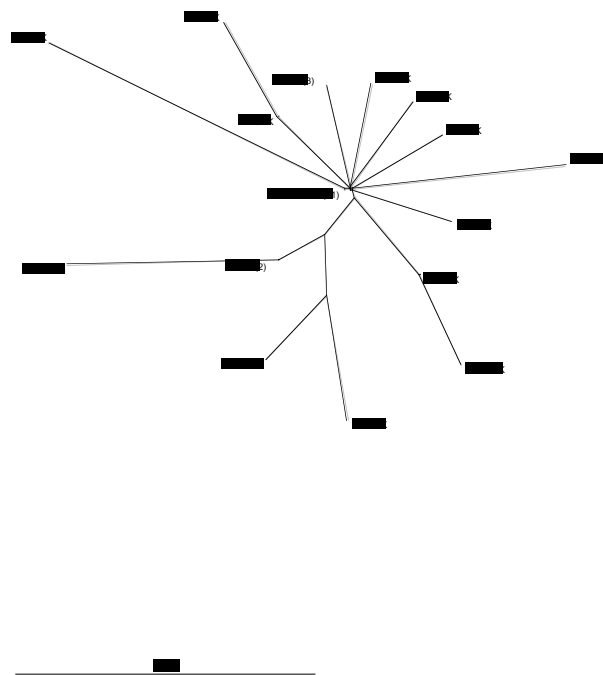


Figura R-36 Árbol filogenético de N-J de las secuencias del haplogrupo K construido a partir de las distancias de Tamura-Nei. El análisis comprende 16 haplotipos distintos encontrados en 29 individuos procedentes de 9 poblaciones. La frecuencia observada de los haplotipos se indica mediante un número entre paréntesis. Los haplotipos compartidos entre poblaciones están marcados con un asterisco *. Las frecuencias de estos haplotipos en las distintas poblaciones son las siguientes: *BCFIUST-K (BRI 2, CAT 3, FIN 1, GALI 2, GALU 1, SAR 1 y TOS 1); *IS-K1 (GALI 1, SAR 1); *ITV-K (GALI 1, TOS 1 y VAS 1).

La secuencia más frecuente, situada en el centro de la filogenia (*BCFIUST), está caracterizada por el motivo 16224C–16311C, en tanto que el subgrupo K1 está caracterizado además por la transición 16093. Sin embargo, la secuencia BRI56 incluida

en este subgrupo en la reconstrucción filogenética, no presenta esta sustitución, lo que podría ser explicado por un evento de retromutación, o por la presencia de otras sustituciones conflictivas que produzcan un sesgo en la reconstrucción filogenética. Sin embargo, hacen falta más datos para llegar a una conclusión clara. Las tres muestras encontradas en la población catalana actual pertenecientes al haplogrupo K presentan una secuencia idéntica al haplotipo *BCFIUST, y en este sentido, la población catalana no destaca especialmente en lo que respecta a las secuencias que presenta de este haplogrupo. En la población antigua de la Plaça Vella se encontraron tres individuos con este haplogrupo, pero sólo dos haplotipos diferentes. Uno de ellos, presentado por dos individuos (PV CE-39 y PV C-13), es idéntico al haplotipo *BCFIUST, y el otro (PV C-58b) es igual al *ITV presentado por las poblaciones de Galicia, Toscana y País Vasco. Este resultado contrasta con el observado para el haplogrupo V, en el que la población catalana presentaba por lo menos un haplotipo exclusivo (CAT3).

El árbol obtenido con las secuencias del haplogrupo U (Fig. R-37) presenta una estructura muy distinta a la observada en los árboles anteriores. En él son distinguibles sin ambigüedad los subgrupos U3, U4, U5 y U6. El subgrupo U1 también resulta diferenciado, aunque relacionado con la única secuencia encontrada del subgrupo U2 (FINI101). El centro de la filogenia lo ocupa la secuencia G28, que para este segmento (302pb) es idéntica a la secuencia de referencia (CRS). Los haplotipos TOS4 y TOS30 pertenecen al haplogrupo U según el análisis con enzimas de restricción (Torrioni et al. 1996) pero no presentan ninguna sustitución característica de algún subgrupo, por lo tanto es normal que en la reconstrucción filogenética no hayan sido incluidos en ninguno de ellos.

Dado que algunas secuencias de las poblaciones que no han sido caracterizadas con enzimas de restricción podrían encontrarse en esta situación, y por lo tanto haber sido incluidas en el conjunto del haplogrupo H, es probable que el centro de la filogenia del haplogrupo U esté subrepresentado y debería de mostrar una forma de estrella producida por múltiples ramificaciones de secuencias derivadas del haplotipo G28.

Un punto interesante lo representa el subgrupo en cuyo extremo se encuentra la secuencia G22, ya que lo constituyen una serie de secuencias que podrían pertenecer al

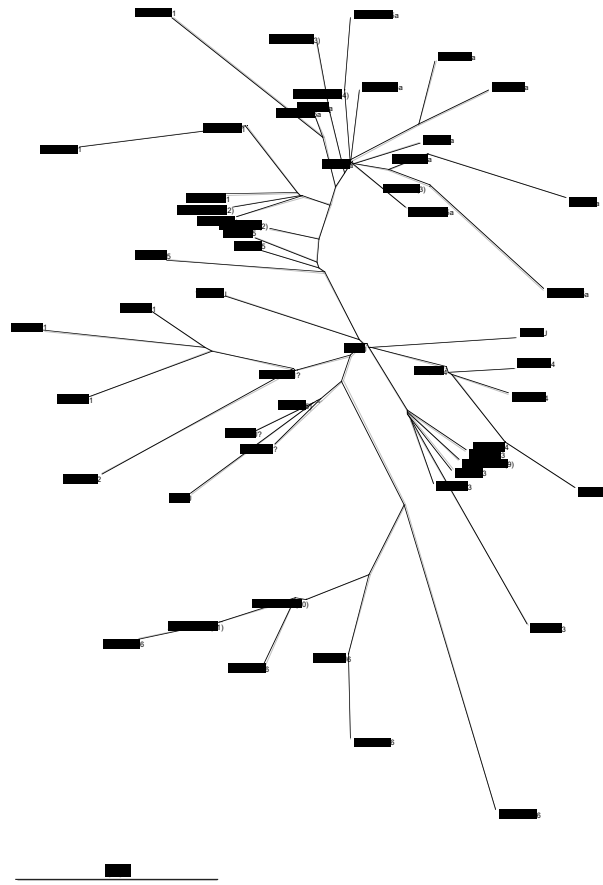


Figura R-37 Árbol filogenético de N-J de las secuencias del haplogrupo U construido a partir de las distancias de Tamura-Nei. El análisis comprende 54 haplotipos distintos encontrados en 94 individuos procedentes de 10 poblaciones. La frecuencia observada de los haplotipos se indica mediante un número entre paréntesis. Los haplotipos compartidos entre poblaciones están marcados con un asterisco *. Las frecuencias de estos haplotipos en las distintas poblaciones son las siguientes: *FU-U5a1a (FIN 1, GALU 1); *FIU-U5a (FIN 1, GALI 1 y GALU 1); *IV-U? (GALI 1, VAS 1).

haplogrupo U, pero que no presentaban todas las sustituciones para ser catalogadas en algún subgrupo concreto. El hecho de que estén relacionadas con la secuencia G22, que está caracterizada con enzimas de restricción, indica que es muy probable que estas secuencias pertenezcan al haplogrupo U.

El subgrupo U6, está constituido casi exclusivamente por individuos de Argelia, su edad podría ser similar a la del subgrupo U5, por lo que podría constituir un grupo derivado de un antecesor común en el Medio Este (Macaulay et al. 1999a). Las

secuencias de dos individuos de Galicia (GALI12 y GALI13) pertenecen a este grupo, de acuerdo a la clasificación previa y a la reconstrucción filogenética, no obstante, es posible que su origen se deba a migración relativamente reciente. Richards y col. (1998) han estimado alrededor del 1% de inmigración de tipos mitocondriales norteafricanos y asiáticos en el acervo mitocondrial europeo. El grupo U3 está compuesto exclusivamente por individuos Turcos y Argelinos, en los que el haplotipo (ARG167) es único pero con una elevada frecuencia (9 individuos). Estos tipos mitocondriales podrían estar relacionados también con la migración desde el Medio Este hacia el norte de Africa en tiempos preneolíticos propuesta por algunos investigadores (Rando et al. 1998; Macaulay et al. 1999a).

Tres de las secuencias de la población catalana actual pertenecen al haplogrupo U. Una de ellas (CAT64) se sitúa en el subgrupo U4, en concordancia con su clasificación previa. En un estudio previo más completo, las secuencias de este subgrupo han sido encontradas en el noroeste de Europa, en poblaciones de la costa mediterránea y en Bulgaria (Richards et al. 1998), por lo que el haplotipo ARG38 encontrado en Argelia podría representar una migración relativamente reciente desde la península ibérica o desde alguna población del mediterráneo hacia el norte de Africa. Las secuencias CAT4 y CAT34 pertenecen al grupo U5, que es el más complejo y uno de los más antiguos (Richards et al. 1998). Está compuesto por dos subgrupos principales, el U5a y el U5a1. La secuencia CAT4, fue clasificada como U5 pues no presenta substituciones de ninguno de estos subgrupos. La reconstrucción filogenética la sitúa en la base de todo el grupo, junto a otras secuencias U5 como SAR11 y BRI93, de acuerdo con lo esperado. La secuencia CAT34 pertenece al subgrupo U5a, y se sitúa en un grupo formado por las secuencias G16, GALI44 y *FIU, concretamente en la rama que conduce a la secuencia de la población Galuresa (G16). Las secuencias del haplogrupo U encontradas en la población de la Plaça Vella, no presentan substituciones característica de ninguno de los subgrupos, por lo que es probable que se sitúen en el centro de la filogenia de forma equiparable a los haplotipos TOS4 y TOS30.

En ninguna de las poblaciones analizadas se encontraron secuencias del subgrupo U5b, un subgrupo definido por Richards y col. (1998), caracterizado por el motivo 16189C–16270T y que de acuerdo con estos autores se distribuye ampliamente en el Oeste y centro de Europa.

De acuerdo con Richards y col. (1998) el haplogrupo J presenta una distribución geográfica muy sorprendente, debido a la presencia de diversos haplotipos antecesores. Según su análisis estos linajes fueron introducidos en Europa durante el neolítico y representan el único aporte genético del Medio Este durante este período. Se compone de dos grupos principales, el J1 y el J2. A su vez, el J1 está formado por los subgrupos J1a1 y J1a2 (J1a y J1b en la nomenclatura de Richards et al. 1998). En la base del J1 se han encontrado haplotipos en poblaciones del Medio Este, y los haplotipos del subgrupo J1a1 han sido encontrados raramente en el sur de la península ibérica (Côrte-Real et. al. 1996), en tanto que los del subgrupo J1a2 sólo han sido encontrados en Gran Bretaña y en Italia. En el árbol que hemos construido con las secuencias de este haplogrupo (Fig. R-38) se observan claramente estos subgrupos, sólo que en el correspondiente al J1a2 está incluida también una secuencia de Galicia (*BI) y en el J1a1 está incluido también un haplotipo del País Vasco (VASH26). En el centro de la filogenia se encuentra el haplotipo más común *ABFIUTK que presenta el motivo 16069T–16126C, el más básico de este haplogrupo.

La única secuencia del haplogrupo J obtenida de la población catalana actual, fue asignada al subgrupo J2a y en la filogenia se encuentra situada en el grupo correspondiente, representada por el haplotipo *CS. Es notable que en esta rama sólo se sitúen secuencias de la isla de Cerdeña, aparte de la secuencia catalana, sugiriendo un movimiento migratorio por la costa mediterránea, en acuerdo con el patrón migratorio propuesto por Richards y col. (1998). En otros estudios se han encontrado secuencias de este subgrupo en Turquía, Italia, Cerdeña e Iberia, aunque también han sido localizadas en Islandia (Richards et al. 1998). Por su estructura filogenética, este subgrupo podría ser más reciente que el J1, sin embargo, son pocas las secuencias analizadas como para tener una visión clara.

En cualquier caso, sería muy interesante intentar localizar este haplogrupo en muestras antiguas preneolíticas, para contrastar la hipótesis de Richards y col. (1998) acerca de su introducción neolítica. Sobretudo considerando que el subgrupo J1 muestra características compatibles con una antigüedad considerablemente mayor que la del J2, por lo que podría pensarse que existía en Europa en tiempos preneolíticos.

Las tres secuencias del haplogrupo J encontradas en la población de la Plaça Vella, son idénticas a la secuencia de referencia (CRS), debido a la reducción del fragmento. Es probable que no pertenezcan a ningún subgrupo y por lo tanto se situarían en el centro de la filogenia, con el haplotipo más común.

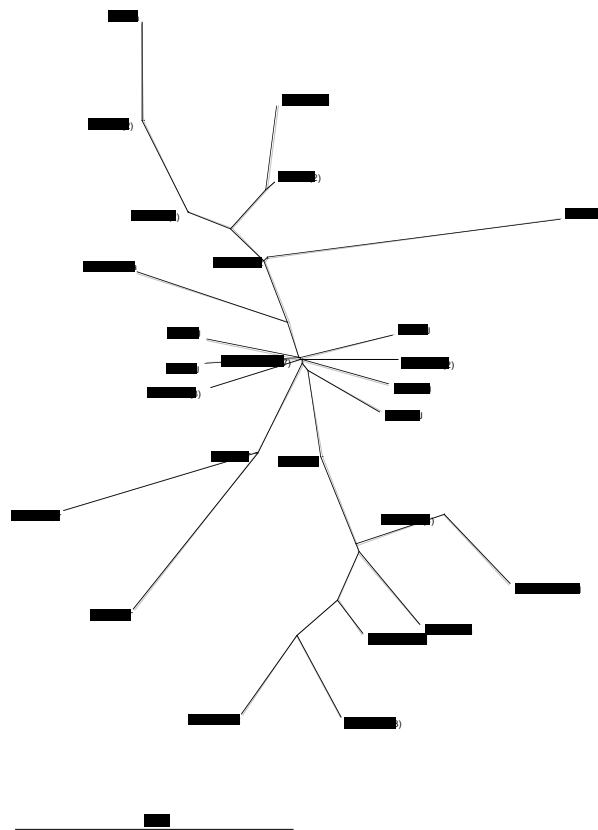


Figura R-38 Árbol filogenético de N-J de las secuencias del haplogrupo J construido a partir de las distancias de Tamura-Nei. El análisis comprende 26 haplotipos distintos encontrados en 53 individuos procedentes de 10 poblaciones. La frecuencia observada de los haplotipos se indica mediante un número entre paréntesis. Los haplotipos compartidos entre poblaciones están marcados con un asterisco *. Las frecuencias de estos haplotipos en las distintas poblaciones son las siguientes: *ABFIUTK-J (ARG 1, BRI 6, FIN 1, GALI 3, GALU 1, TOS 2 y TRK 3); *BT-J1a1 (BRI 3, TOS 1); *BI-J1a2.1 (BRI 2, GALI 1); *CS-J2a (CAT 1, SAR 1); *TK-J2 (TOS 1, TRK 1).

El haplogrupo T tiene un origen común con el haplogrupo J en el Medio Este, a partir de una secuencia ancestral (JT) caracterizada por la transición 16126 (Macaulay et al. 1999a). Es mucho más antiguo que el J y pudo haber formado parte de la primera migración Este – Oeste durante el Paleolítico superior temprano (Richards et al. 1998). De acuerdo con Francalacci y col. (1999) está compuesto por dos subgrupos principales, e 1 y e 2, y T 1 y T 2.

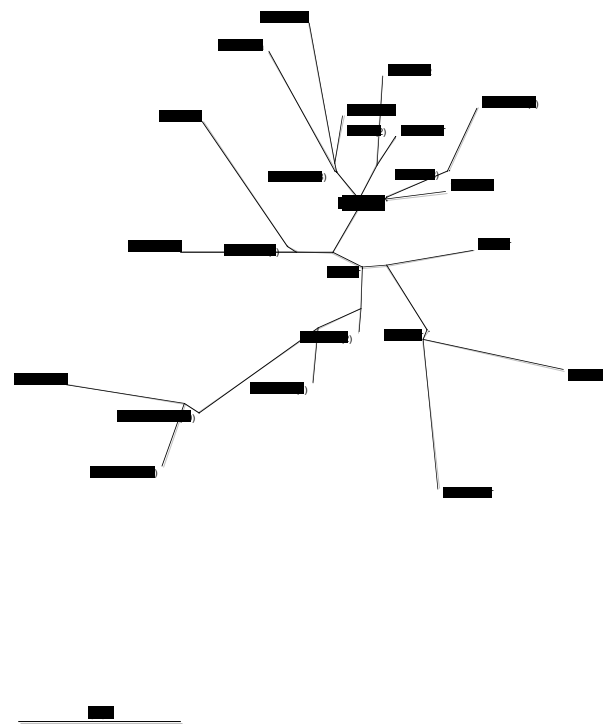


Figura R-39 Árbol filogenético de N-J de las secuencias del haplogrupo T construido a partir de las distancias de Tamura-Nei. El análisis comprende 26 haplotipos distintos encontrados en 49 individuos procedentes de 10 poblaciones. La frecuencia observada de los haplotipos se indica mediante un número entre paréntesis. Los haplotipos compartidos entre poblaciones están marcados con un asterisco *. Las frecuencias de estos haplotipos en las distintas poblaciones son las siguientes: *ABCFK-T1a1 (ARG 4, BRI 2, CAT 1, FIN 1 y TRK 2); *BFIU-T2a (BRI 2, FIN 3, GALI 2 y GALU 1); *BU-T2a1 (BRI 1, GALU 1); *SV-T (SAR 1, VAS 1); *TK-T2 (TOS 1, TRK 1).

aunque Richards y col. (1998) sólo identifican el subgrupo T1. El árbol obtenido con nuestros datos (Fig. R-39) coincide con la clasificación previa de las secuencias, llevada a cabo con el criterio de Francalacci y col. (1999), ya que es posible observar los dos grupos bien diferenciados. Incluso, dentro de cada grupo es posible distinguir subagrupaciones que se corresponden con la nomenclatura propuesta por estos autores.

En la raíz del subgrupo T1a1 se encuentra el haplotipo compartido *ABCFK en el que está incluida una de las secuencias encontradas en la población catalana actual (CAT16-T1a1). Este haplotipo podría estar relacionado con un patón migratorio que seguiría la costa mediterránea hasta las poblaciones de la zona atlántica en vista de que dos de las secuencias que se derivan de él proceden de las poblaciones de Cerdeña (SAR15) y Toscana (TOS32). Por otra parte, este haplotipo se encontró con mayor frecuencia en la población de Argelia (4/10), y sería interesante investigar si se trata de una aportación reciente o si proviene de las migraciones primigenias ocurridas durante el Paleolítico. Otra de las secuencias catalanas (CAT5) carece de substituciones características de alguno de estos subgrupos, por lo que es normal su situación en el árbol partiendo del centro de la filogenia. La secuencia CAT24, pertenece al grupo T2 y se sitúa junto a otras secuencias de este grupo, como las procedentes de Turquía y Toscana (TRK38, *TK y TOS22).

En la raíz del subgrupo T2a se encuentra un haplotipo frecuente compartido por varias poblaciones (*BFIU), del que se derivan secuencias de tres poblaciones, una de ellas, la CAT10, es la cuarta secuencia del haplogrupo T obtenida de la población catalana actual. Por otra parte, en la raíz del subgrupo T2a1, se encuentra el haplogrupo compartido *BU, con sólo dos haplotipos derivados pertenecientes a las poblaciones de Cerdeña (SAR36) y Galura (G17). Las secuencias JT fueron incluidas tanto en el árbol del haplogrupo J como del T, dado que están relacionadas con los dos haplogrupos. De hecho, en un árbol construido con los haplotipos de los dos haplogrupos (no mostrado), las secuencias JT (especialmente FIN90) se sitúan en el punto por el que se unen las filogenias de los dos haplogrupos.

De la población antigua de la Plaça Vella se obtuvieron dos secuencias que pertenecen al haplogrupo T. Una de ellas (PV E-35) resultaría idéntica para este segmento (153pb) a la secuencia SAR5, situada en el centro de la filogenia. La otra (PV E-10) es igual para este segmento al haplotipo CAT10, situado en el subgrupo T2a, por lo que podría tratarse de una variante exclusiva de la población Catalana.

Un aspecto interesante del árbol que hemos construido con las secuencias del haplogrupo T, es que no muestra signos de contener una filogenia en forma de estrella en ninguna de sus zonas. Si este haplogrupo fue introducido en el Paleolítico superior, el efecto de la hipotética expansión poblacional post-glacial debería de ser notorio, a menos que su frecuencia fuese considerablemente más baja que la observada en las poblaciones actuales. Este punto es particularmente interesante, ya que los cambios en la frecuencia de los distintos haplogrupos podrían mostrar un panorama muy distinto al inferido por Richards y col. (1998) sobre la historia demográfica en Europa.

Por ejemplo, un aumento importante en la frecuencia del haplogrupo H, y en menor medida de los haplogrupos K y V en detrimento de otros haplogrupos, podría explicar su diversificación sin implicar necesariamente una expansión poblacional. Richards y col. (1998), basándose en simulaciones, demostraron que la frecuencia del haplogrupo J no ha cambiado considerablemente desde el neolítico, sin embargo, es necesario investigar también la posibilidad de cambios en la frecuencia de otros haplogrupos probando escenarios de crecimiento poblacional desde el Paleolítico. Sobretudo si se considera que teóricamente, los linajes descendientes de un único tipo mitocondrial, con el tiempo reemplazarán a los descendientes de otros linajes (ver por ej. Vigilant et al. 1991). En este sentido resulta interesante preguntarse si el haplogrupo H se está expandiendo dentro de la población (aumentando su frecuencia) para eventualmente reemplazar a los haplogrupos restantes. Por este motivo, resulta también relevante utilizar métodos más robustos para inferir la antigüedad relativa de los haplogrupos, como el método de máxima verosimilitud de Weiss y von Haeseler (1998).

En la Figura R-40 se presenta el árbol de los haplogrupos I, X y W, que han sido analizados conjuntamente debido a su proximidad filogenética y a que se encuentran con baja frecuencia en las poblaciones hasta ahora estudiadas. En el centro de la filogenia no se observa ningún haplotipo, pero sin duda debe tratarse de una secuencia con la transición 16223, ya que es característica de los tres haplogrupos. En general, las secuencias están agrupadas de acuerdo al haplogrupo que les corresponde según la asignación previa, excepto en tres casos que no se sitúan dentro de ningún grupo. Uno de

ellos (TRK50-I), situado cerca del haplogrupo W, presenta las dos substituciones típicas del I, pero también presenta las dos substituciones del haplogrupo W. La secuencia GALI133-I presenta un patrón similar, ya que contiene las substituciones del I, pero también las substituciones características del haplogrupo X, lo que puede explicar su situación en una rama independiente cerca de la raíz del haplogrupo X.

No descartamos que estas secuencias puedan estar relacionadas con secuencias del grupo L y un análisis con enzimas de restricción podría clarificar su situación. El haplotipo CAT19 de la población catalana, fue caracterizado con enzimas de restricción como perteneciente al haplogrupo X, sin embargo, en el árbol se sitúa fuera de este grupo, en una posición relativamente distante. Si se analiza la secuencia es posible observar que carece de una de las substituciones características del haplogrupo X, la 16278 T y que otra de ellas, la 16391C, no queda incluida en el fragmento de 302pb analizado. Así, su situación en el árbol puede ser un artefacto producido por estas dos circunstancias. La posición 16278 presenta una tasa de mutación relativamente rápida (Richards et al. 1998), así que es posible que en el linaje de este haplotipo se haya producido una retromutación en este sitio.

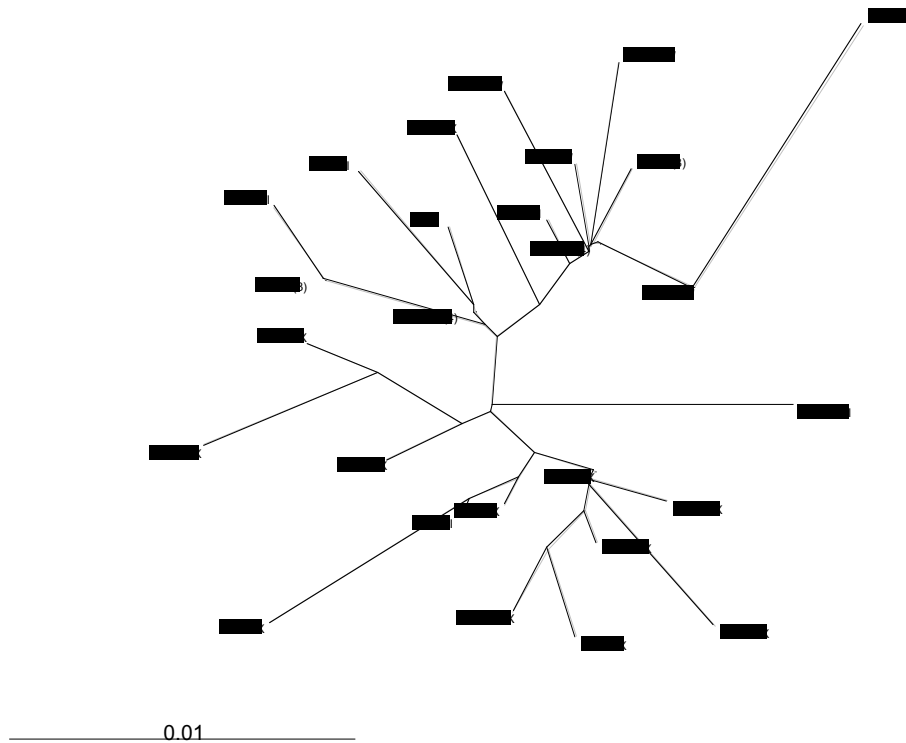


Figura R-40 Árbol filogenético de N-J de las secuencias de los haplogrupos I, X y W construido a partir de las distancias de Tamura-Nei. El análisis comprende 27 haplotipos distintos encontrados en 37 individuos procedentes de 9 poblaciones. La frecuencia observada de los haplotipos se indica mediante un número entre paréntesis. Los haplotipos compartidos entre poblaciones están marcados con un asterisco *. Las frecuencias de estos haplotipos en las distintas poblaciones son las siguientes: *BFT-I (BRI 1, FIN 1 y TOS 1); *FI-W (FIN 2, GALI 1); *FSK-W (FIN 2, SAR 1, TRK 1).

En las muestras de la Plaça Vella fue encontrado un individuo perteneciente al haplogrupo I (PV E-34). Como su secuencia ha resultado ser única, se ha realizado el análisis de los haplogrupos I, X y W utilizando el fragmento de 153pb. Este análisis presenta el problema de que se excluye el sitio 16129, relevante para la caracterización del haplogrupo I. Por lo que algunas secuencias pertenecientes a distintos haplogrupos resultan idénticas. En la Figura R-41 se muestra el árbol filogenético obtenido.

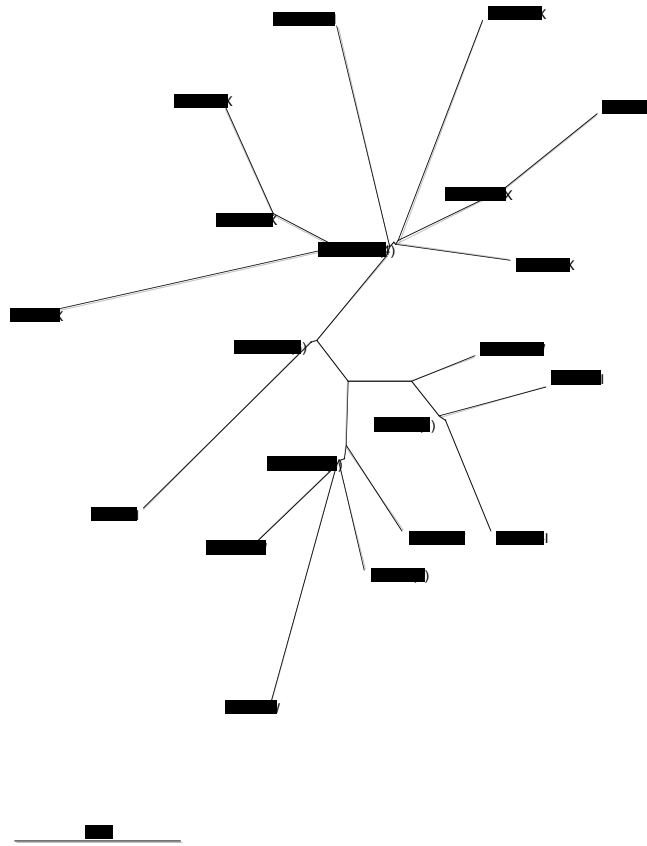


Figura R-41 Árbol filogenético de N-J de las secuencias de 153pb de los haplogrupos I, X y W construido a partir de las distancias de Tamura-Nei. El análisis comprende 20 haplotipos distintos encontrados en 38 individuos procedentes de 10 poblaciones. La frecuencia observada de los haplotipos se indica mediante un número entre paréntesis. Los haplotipos compartidos entre poblaciones están marcados con un asterisco *. Las frecuencias de estos haplotipos en las distintas poblaciones son las siguientes: *BFT-I (BRI 2, FIN 1 y TOS 1); *FI-W (FIN 2, GALI 1); *FSK-WI (FIN 2, SAR 1, TRK 3); *BTVK-X (BRI 1, TOS 1, VAS 1 y TRK 1); *CFU-IX (CAT 1, FIN 4, GALU 1).

La mayoría de secuencias conservan su diferenciación, lo que permite distinguir los tres haplogrupos. La secuencia de la población de la Plaça Vella (PV E-34) se sitúa en un grupo formado por otras dos secuencias del haplogrupo I, una de ellas de la población Toscana (TOS15) y otra correspondiente a un haplotipo compartido por las poblaciones Británica, Finlandesa y Toscana (*BFT). Este resultado descarta en cierta medida la posibilidad de que la secuencia antigua sea un artefacto de amplificación.

El haplogrupo I presenta una frecuencia extremadamente baja en la población europea actual, y sería interesante constatar si en poblaciones antiguas su presencia también era escasa. El hecho de haberlo encontrado en una muestra pequeña, abre la

posibilidad de que su frecuencia haya sido mayor en tiempos relativamente recientes. En la muestra de 92 individuos de la población Catalana actual, caracterizados con enzimas de restricción, no fue encontrado ninguno perteneciente a este haplogrupo.

DISCUSIÓN

DNA ANTIGUO

Consideraciones metodológicas

Criterios generales

El presente trabajo es un estudio piloto que tiene la finalidad de valorar la viabilidad de llevar a cabo estudios de DNA antiguo a nivel poblacional en poblaciones europeas. Por lo tanto, se ha procurado realizar más un estudio extensivo que intensivo, en el sentido de que se ha preferido analizar el mayor número posible de individuos, antes que la corroboración exhaustiva del resultado de un número de individuos menor. La lógica subyacente en esta aproximación, es la idea de que la variabilidad a nivel poblacional puede ser un criterio único de autenticidad y esta variabilidad no puede ser valorada con un número muy limitado de individuos. Así, se ha preferido la secuenciación del fragmento del D-loop en individuos diferentes antes que repetirla sistemáticamente en los mismos individuos. No obstante, teniendo en cuenta la importancia de la repetibilidad de los resultados, se ha pensado en un método alternativo de corroboración, como la caracterización de marcadores genéticamente ligados. A lo largo de la presente discusión se valorará la efectividad de esta aproximación.

Otro criterio general seguido durante el desarrollo de este trabajo, fue la minimización del riesgo de contaminación, por lo que se eligieron los procedimientos que implicaban menos pasos y se evitaron los pasos considerados innecesarios.

Por ejemplo, se evitó la cuantificación del DNA obtenido, ya que como se ha visto en el Marco Teórico, la información que se puede obtener con métodos convencionales no es tan precisa y en cambio requiere la apertura de los tubos para obtener una alícuota para la medición. Se ha propuesto que la cuantificación con un

método fiable como la PCR competitiva, es esencial para descartar que un resultado de DNA antiguo provenga de un evento de contaminación, no obstante, otros investigadores han demostrado que esta cuantificación no aporta la información adecuada para este fin (Kolman y Tuross, 2000).

También se evitó la dilución del DNA obtenido, pues implica la apertura de los tubos y agregar agua que puede representar una fuente de contaminación. En cambio, lo que se hizo fue optimizar la relación entre la cantidad de tejido inicial y la cantidad final de agua en la que quedaba diluido el DNA en el paso final de la extracción. Después sólo se abrían los tubos para obtener la alícuota para las amplificaciones. El número de intentos de amplificación que se realizaron con cada extracto también estuvo condicionado por este criterio. En general, se evitó la repetición de muchos intentos, pues se consideró que repetir los intentos de amplificación puede conducir finalmente a la amplificación de DNA contaminante. No obstante, en pocos casos se llegó a un máximo de 6 intentos debido a que se consideraba importante la obtención de un resultado para corroborar otro, o porque se sospechaba que podría haber amplificación, sobretodo si en un intento anterior se había observado alguna “sombra” en el gel de agarosa que pudiera corresponder con la banda buscada; pero en promedio se realizaron dos intentos por extracto.

También con la idea de minimizar el riesgo de contaminación, se evitó la purificación de las amplificaciones antes de reamplificarlas. De acuerdo con algunos autores, este paso es uno de los más propensos a la contaminación (Kolman y Tuross, 2000). Tampoco se recurrió a la reamplificación sistemática de todas las muestras amplificadas, sino solamente de aquellas en las que se había observado una banda bien definida del tamaño esperado, aunque fuese muy tenue.

Por todos estos motivos, cuando se observan los resultados *a posteriori*, puede parecer extraño que para unos individuos se hayan obtenido más extractos o que se hayan hecho más amplificaciones que para otros, sin embargo, este fenómeno está justificado por la naturaleza del trabajo. De hecho, estos y otros factores condicionan en exceso el diseño experimental, como se discute a continuación.

Diseño experimental

Tradicionalmente, en los estudios de DNA antiguo el diseño experimental ha ocupado un segundo plano, debido en parte al costo de los análisis y al valor intrínseco de las muestras, pero quizá también a la premura por conseguir resultados positivos, pues a partir de éstos es posible conseguir financiación para nuevos estudios. Así, se han hecho muy pocos trabajos en los que se intente evaluar, por un lado, la influencia que diferentes variables pueden tener en la preservación del DNA (como el tipo de suelo, la temperatura media a la que han estado expuestos los restos, etc.) y por otro, las variables que pueden influir en la extracción y amplificación del DNA preservado (como los protocolos, condiciones de amplificación etc.). Esta falta de diseño experimental se ve claramente reflejada en la completa ausencia de análisis estadísticos sobre los resultados obtenidos en los pocos trabajos en los que se han comparado algunas variables, como la influencia de la antigüedad de los restos en la preservación del DNA.

Desafortunadamente, esta tesis no escapa por completo a esta tendencia. Por ejemplo, a pesar de que en un principio se utilizaron diversos protocolos, no fue posible analizar un número representativo de muestras para cada uno de ellos, pues al empezar a obtener resultados positivos con el protocolo UAB-2, los demás protocolos fueron relegados. Sin embargo, se ha intentado recuperar la mayor cantidad posible de información, no sólo en cuanto al protocolo de extracción se refiere, sino también en cuanto al número de muestras de las que se extrajo DNA y el número de alícuotas analizadas de cada extracto, así como del tipo de Polimerasa utilizada en la amplificación. Estos datos, aunque a veces incompletos, han servido para suplir hasta cierto punto la falta de diseño experimental que caracteriza los trabajos de DNA antiguo.

Por otra parte, el diseño experimental también se ve limitado por la particularidad que supone la preservación del DNA antiguo. En el presente trabajo se repetía la extracción (o la amplificación) preferentemente con muestras de individuos que previamente habían mostrado algún resultado positivo o algún indicio de contener DNA, con la finalidad de corroborar el resultado en el primer caso y de evitar el gasto

innecesario que supone analizar una muestra en la que existen menos posibilidades de encontrar DNA, en el segundo. Esto a fin de cuentas ejerce un sesgo en el muestreo, en el sentido de que para algunos individuos no se analiza el mismo número de muestras que para otros. Pero tampoco es posible definir un número representativo de muestras de cada individuo, ya que las condiciones de preservación del DNA pueden variar de individuo a individuo a pesar de que pertenezcan a la misma necrópolis. De hecho, las condiciones de preservación varían incluso entre los dientes de un mismo individuo, como lo indican los resultados obtenidos en este trabajo, en los que se puede observar que en muchos casos no fue posible recuperar DNA de distintas piezas de un mismo sujeto (ver por ejemplo Tabla R-11).

Otro aspecto a tener en cuenta cuando se aplican análisis estadísticos para evaluar las diferencias de la amplificación en relación a distintas variables, es que se debe asumir la existencia de DNA en tejidos antiguos y que tanto su preservación como su recuperación son variables aleatorias; siendo el objetivo del análisis la determinación de las variables que pueden influir significativamente en ellas. Sin embargo, no se sabe nada de la distribución que siguen las variables “preservación” y “recuperación” y por lo tanto se debe tener precaución a la hora de elegir los métodos de análisis estadísticos. Afortunadamente las pruebas no paramétricas como la χ^2 de Pearson y el estadístico exacto de Fisher, han sido concebidas para analizar este tipo de variables.

Variables que influyen en la preservación

Dentro de las variables analizadas, la población de procedencia de las muestras y el tipo de tejido utilizado son las que pueden tener una influencia importante en la preservación del DNA. En realidad, la población de procedencia compendia un conjunto de variables independientes entre sí, que pueden tener un efecto individual y/o conjunto, como son la antigüedad, las condiciones climáticas a las que ha estado sometida la necrópolis a lo largo del tiempo y el tipo de suelo en el que se encuentra el yacimiento (pH, % de materia orgánica etc.).

De las poblaciones analizadas, únicamente son comparables las poblaciones de s'Illot des Porros (IP) y de la Plaça Vella (PV), en cuanto al número de individuos y de muestras analizadas se refiere. Los resultados indicaron una eficiencia significativamente menor en la recuperación de DNA endógeno en la necrópolis de IP en comparación con la obtenida en PV. Existen diversas hipótesis para explicar esta diferencia. Por un lado, la más sencilla y directa, apunta a la posibilidad de que el DNA esté más degradado en la población de IP, debido a la antigüedad o a factores climáticos, ya que la población de IP es más antigua y se encuentra situada en un islote marino, en el que las condiciones de humedad son evidentemente extremas. Otra hipótesis postularía que las diferencias se deben a que los protocolos están optimizados para las muestras de la Plaça Vella y por tanto las muestras procedentes de cualquier otra población no responderían de la misma forma si los factores tafonómicos no son equivalentes. No obstante, los resultados preliminares obtenidos en este trabajo no indican diferencias substanciales en cuanto a los inhibidores encontrados en una u otra población (ver resultados de los espectros, Pag. 298).

Por otra parte, resultaría interesante investigar porqué sólo algunos individuos del mismo yacimiento contienen DNA amplificable, ya que en general todos ellos han estado sometidos a las mismas condiciones de preservación. Es obvio que las causas de este fenómeno deben ser buscadas en las particularidades de cada individuo. Por ejemplo, si en verdad las primeras etapas son decisivas para la preservación del DNA, la época del año en la que falleció y fue enterrado el individuo podría tener un papel decisivo. Otro factor a tener en cuenta es la constitución del individuo, ya que el porcentaje de grasa y agua en los tejidos podrían acelerar o ralentizar la autólisis. Asimismo, la ubicación del individuo en el yacimiento podría tener cierta influencia, ya que en ocasiones pueden darse diferencias tafonómicas importantes en las distintas zonas de un mismo yacimiento. Otra hipótesis para explicar este fenómeno, sería la posibilidad de que el DNA obtenido provenga de eventos de contaminación relacionados con las muestras de los individuos que dan resultados positivos. Esta posibilidad no puede ser descartada *a priori*, sin embargo, el análisis de los resultados puede ayudar a evaluar su probabilidad. Por

ejemplo, si esta hipótesis fuese cierta, sería difícil explicar porqué se presentan diferencias en la eficiencia de la amplificación entre distintas poblaciones si los individuos han sido tratados con los mismos procedimientos, con los mismos reactivos y de forma indiscriminada.

En lo referente a la diferencia observada con los diferentes tipos de diente utilizados, en el presente trabajo se encontró cierta desventaja en la utilización de molares para la extracción de DNA. Esto definitivamente tiene que estar relacionado con las condiciones de preservación en los diferentes tipos de piezas dentales y para explicar este fenómeno sería necesario investigar sobre la composición de los tejidos internos en cada uno de ellos y su relación con la degradación del DNA. Por el momento, la observación que podemos aportar es el hecho de que la raíz de los molares está dividida, lo que ocasiona que el tamaño de la cavidad pulpar sea menor con respecto a las demás piezas dentales, contrariamente a lo que podría parecer. Si se demostrara que existe una correlación entre la amplificación y el tamaño de la cavidad pulpar, entonces esto indicaría que lo importante es la cantidad y la calidad del material utilizado. Al tener una menor cavidad pulpar, en los molares el tejido que se colecta es mayoritariamente dentina y esmalte de la corona, en tanto que en otras piezas lo que se colecta es principalmente tejido necrótico de la cavidad pulpar (lo que podría haber sido pulpa dentaria) y en menor medida, dentina de las paredes internas que podrían contener el DNA liberado del tejido de la pulpa dentaria. Si esta hipótesis es cierta, utilizar toda la pieza dental no sería lo más adecuado, ya que la relación DNA/fracción mineral-contaminantes sería más desfavorable en la medida en la que se utilice más dentina de las capas intermedias y externas o esmalte.

Aunque la extrema baja eficiencia encontrada en las muestras de hueso puede estar relacionada con la falta de idoneidad del protocolo, este resultado concuerda con la observación de la mayoría de investigadores que han comparado extractos de huesos y dientes, quienes han detectado DNA de mejor calidad en piezas dentales. Esto hace que todavía sea más interesante investigar sobre la composición de los dientes y su relación con la preservación del DNA. ¿Es el alto contenido de hidroxapatita el responsable de la

mejor preservación? ¿o se debe a que en los dientes la “momificación celular” es más rápida y/o efectiva? Otra hipótesis podría plantear que la preservación del DNA en huesos y dientes es la misma y que la diferencia en la eficiencia de la amplificación se debe a otros factores, como una mayor presencia de sustancias inhibitorias del PCR en los huesos, o a un inhibidor de naturaleza diferente al que se presenta en piezas dentales. La porosidad del tejido óseo lo hace más propenso a absorber todo tipo de sustancias del suelo y de hecho, los extractos de hueso que hemos obtenido en el laboratorio son invariablemente más intensos en cuanto al “color marrón”. Además, el análisis espectrofotométrico de los inhibidores en extractos de huesos, ha revelado ciertas diferencias con el encontrado en los dientes (Pag. 299).

Por otra parte, el hecho de que los huesos puedan ser más propensos a la contaminación debido a la porosidad de su tejido, los sitúa en clara desventaja en relación a las piezas dentales. Diversos investigadores que han caracterizado el mtDNA extraído de huesos, han encontrado la presencia de distintas moléculas procedentes de fuentes naturales, la mayoría de ellas atribuibles a contaminación durante el manejo de los especímenes (Richards y Sykes, 1995; Handt et. al. 1996; Kolman y Tuross, 2000). En este contexto, el hecho de utilizar únicamente la parte interna del diente y no toda la pieza, podría explicar porqué en nuestros resultados se han encontrado relativamente pocos casos con distintos mtDNAs, ya que el hecho de descartar la parte externa del diente, podría evitar en gran medida la amplificación de DNA contaminante.

Variables que influyen en la recuperación del DNA

Una vez asumida la posibilidad de que el DNA sea preservado en tejidos antiguos y como parte del proceso de optimización de los protocolos de extracción, es necesario evaluar las variables que pueden influir en la eficiencia de su recuperación. Al analizar los resultados de la primera fase de este trabajo (la amplificación de los fragmentos de la región de control) se observó que no existían diferencias importantes entre los protocolos, ni entre las polimerasas utilizadas y que existió un incremento significativo

en la eficiencia con la utilización del paso del “almacenamiento frío”. Sin embargo, en la segunda fase, durante la amplificación de las regiones codificantes en la que el “almacenamiento frío” fue un factor constante, se vieron incrementadas las diferencias entre los resultados de algunas variables, tanto en las involucradas en la preservación (como el tipo de tejido) como en la recuperación (como el protocolo de extracción). En conjunto, estos datos indican que el principal factor que disminuye la eficiencia de la recuperación del DNA preservado es la presencia de inhibidores. Más aún, estas sustancias pueden incluso “enmascarar” los resultados al analizar las variables que influyen en la preservación del DNA, ya que puede disminuir las diferencias que presentan, por ejemplo, los diferentes tipos de dientes.

Sin embargo, también hay que reconocer que existe un límite respecto a los beneficios que se pueden obtener con la mejora de los protocolos, pues por muy notable que ésta sea, si no existe DNA en las muestras, será imposible su recuperación, como en el caso de las muestras de cabello. Lo mismo podría decirse de las muestras en las que existe una menor probabilidad de preservación de DNA como en la población de s'Illot des Porros. En cualquier caso, únicamente con investigaciones más completas que incluyan estudios de prospección molecular se podrá distinguir de manera fiable si la baja eficiencia en la recuperación de DNA endógeno en un conjunto de muestras se debe a que contienen poco o ningún DNA amplificable, a la presencia excesiva de sustancias inhibitorias o a inhibidores diferentes de los que se presentan en otro tipo de muestras de las que sí es posible obtener DNA. Por el momento, para las muestras analizadas en este trabajo, el paso del “almacenamiento frío” ha demostrado ampliamente su utilidad, pues ha incrementado considerablemente la eficiencia de la amplificación sin aumentar el riesgo de contaminación.

El tamaño de los fragmentos amplificados es una variable que influye en la eficiencia de la amplificación, pero en realidad está más relacionado con la preservación del DNA. En general nuestros resultados apoyan la idea de que los fragmentos más pequeños tienen más probabilidades de ser conservados intactos. Sin embargo, a partir de estos resultados no se puede decir que los fragmentos grandes no existan o que se

preserven con menor frecuencia, ya que puede ser que la limitación en la amplificación no se deba a la fragmentación sino a otro tipo de daño molecular que impida la replicación. Obviamente, mientras más grande sea el fragmento existe mayor probabilidad de que contenga más sitios dañados. Por este motivo hemos enfatizado la palabra “intactos”.

Criterios de autenticidad

Desde los inicios de la disciplina del DNA antiguo, la autenticidad de los resultados obtenidos ha sido el principal tema de discusión, ya que es de vital importancia demostrar que el DNA obtenido provenga de los restos estudiados y no de fuentes externas. Esta tarea, que en los estudios con DNA contemporáneo es sencilla, en los estudios de DNA antiguo se ha visto complicada hasta el límite, sobretodo en los análisis de restos humanos. El origen de esta dificultad viene dada por la relación DNA endógeno/DNA exógeno, pues en la medida en la que el DNA endógeno sea más escaso, el DNA exógeno cobrará mayor relevancia. Obviamente, la dificultad aumenta cuando no es posible detectar el DNA contaminante, o cuando no es posible diferenciarlo del DNA endógeno.

La detección del DNA contaminante no es tan sencilla como puede parecer y la diferenciación entre DNA endógeno y DNA exógeno es realmente difícil cuando se estudian poblaciones humanas del mismo origen geográfico que el de los investigadores. Por estos motivos la autenticación del DNA extraído debe basarse en un conjunto de criterios que nos permitan excluir la posibilidad de que el DNA provenga de fuentes externas.

Muchos investigadores han comprobado que los criterios de exclusión de primer nivel (por así llamarlos) no son suficientes. A pesar de tomar precauciones como la limpieza de las muestras, la separación de las áreas de trabajo y el uso de guantes y campanas de aire filtrado, éstas no han sido suficientes para evitar la contaminación de las muestras. Más aún, los blancos experimentales (blancos de extracción y de amplificación) tampoco han demostrado ser útiles en la detección del DNA contaminante, pues a pesar de no haber detectado DNA en los blancos, diversos investigadores han detectado DNA contaminante en algunas muestras (Richards y Sykes, 1995; Handt et al. 1996; Krings et al. 1997; Stone y Stoneking, 1998; Kolman y Tuross, 2000). Este fenómeno es de esperar si la contaminación se produce durante la

manipulación previa de las muestras, pues los blancos no tienen la capacidad de detectar este tipo de contaminación. Sin embargo, si la contaminación se produce durante la extracción o la amplificación, los blancos deberían mostrarla. El fallo en la detección de este tipo de contaminación, puede atribuirse al efecto *carrier*, pero también a la posibilidad de que la contaminación haya sido puntual, es decir, que haya ocurrido en una única muestra y que por tanto no esté relacionada con los reactivos ni con el procedimiento general.

Los resultados obtenidos en este trabajo apoyan esta hipótesis pues el hecho de que la secuencia de los blancos secuenciados sea diferente a la secuencia de las muestras, ilustra la posibilidad de que la contaminación en el blanco sea puntual y no que se encuentre en los reactivos como teóricamente se espera. El problema es que si puede darse una contaminación puntual en un blanco, lo mismo puede ocurrir con cualquier muestra, que al amplificar en presencia de un blanco limpio nos haría obtener un falso positivo. La situación se agrava si consideramos que en un grupo de extracción se incluyen más muestras que blancos (en relación de 5 a 1 en nuestro caso), por lo que las contaminaciones puntuales tienen más probabilidad de ocurrir en las muestras que en los blancos. Por otra parte, esta posibilidad también nos hace pensar que cuando se obtiene un blanco contaminado no necesariamente se debe asumir que estén contaminadas las muestras, por lo que la fiabilidad de los blancos se ve aún más comprometida, sin olvidar también que la presencia de una banda en el blanco puede deberse a un artefacto de PCR como se puntualizó en la sección de resultados. Sin embargo, la información obtenida de los blancos puede ser más relevante si se incluye dentro de un marco de criterios más amplio.

Por ejemplo, la relación entre la eficiencia y el porcentaje de blancos contaminados puede ayudar en la evaluación de la posibilidad de que los resultados obtenidos provengan de contaminaciones puntuales o se deban al efecto *carrier*. En general es deseable que la eficiencia supere la proporción de blancos contaminados, pues si asumimos que la contaminación es una cuestión de probabilidades debemos asegurarnos que la eficiencia de la amplificación esté por encima de la probabilidad de

contaminación, estimada con la proporción de blancos contaminados. En otras palabras, el valor de la relación eficiencia / contaminación (E/C) debería ser superior a 1.

Relación eficiencia / contaminación

En cuanto al análisis de los fragmentos de la región de control, el porcentaje de blancos de extracción contaminados asciende a un 15.55% (7 de 45); en tanto que en el 50% (22 de 44) de los grupos de extracción realizados, se obtuvo por lo menos una muestra con amplificación positiva en presencia de blancos limpios, lo que nos da un valor de E/C= 3.21. La diferencia que existe entre los porcentajes de contaminación y de eficiencia es estadísticamente significativa ($\chi^2= 12.02$; g.l.= 1; $P=0.00053$), lo que indica que los procesos por los cuales se obtienen amplificaciones en los blancos, no son los mismos que los que actúan en la amplificación de las muestras.

En los blancos, el único proceso plausible son los eventos de contaminación durante el análisis molecular, pero en las muestras es necesario buscar una explicación alternativa para el aumento del porcentaje de amplificación. Existen dos posibilidades, por un lado, que el DNA provenga de contaminaciones ocurridas antes del análisis en el laboratorio (durante la excavación o durante el estudio morfológico); y por otro, que en las muestras exista DNA endógeno. De cualquier modo, no se debe pasar por alto que estos factores explicarían el incremento, pero podría existir un porcentaje de muestras amplificadas en grupos con blancos limpios, en las que el DNA provenga de eventos de contaminación ocurridos durante el análisis, que no serían detectados debido a fenómenos tales como el efecto *carrier*, o las contaminaciones puntuales. En este sentido, resulta interesante preguntarse si la proporción de blancos contaminados (15.55% en este caso), puede ser un indicador del porcentaje de muestras, respecto al total, cuya amplificación proviene de una contaminación no detectada.

Es importante aclarar que este análisis se ha realizado con el conjunto de datos obtenido de todos los protocolos, sin embargo, la relación efectiva de E/C para este

trabajo es en realidad mayor, pues no tiene sentido incluir los protocolos con los que no se obtuvo ningún resultado positivo. Con el protocolo más utilizado, el UAB-2, se obtuvo un porcentaje de contaminación del 7.40% (2 de 27) en los blancos de extracción analizados, en tanto que en un 55.55% (15 de 27) de los grupos de extracción se obtuvo por lo menos un resultado positivo ($E/C = 7.5$). Obviamente, esta diferencia también es significativa ($\chi^2 = 14.51$; g.l. = 1; $P = 0.000139$).

Por otra parte, esta relación disminuye de manera preocupante cuando se analizan los resultados de la contaminación durante la amplificación, ya que el 16.04% (13 de 81) de los grupos de amplificación revelaron estar contaminados, en tanto que tan sólo el 27.16% (22 de 81) mostraron al menos un resultado positivo ($E/C = 1.69$). Sin embargo, este resultado también está sesgado, tanto por las muestras en las que no existía DNA endógeno (de cabello) ó en las que existía muy poco DNA (de hueso), como por los resultados de los protocolos de extracción que no funcionaron. Si se analiza esta relación en las muestras de dientes, se observa que la diferencia entre la proporción de contaminación de los grupos de amplificación (16.92%; 11 de 65) y su eficiencia (32.30%; 21 de 65) es significativa ($\chi^2 = 4.145$; g.l. = 1; $P = 0.041$); y que el valor de E/C es más alto (1.91). Esta comparación, pone de manifiesto que la relación C/E alcanza su máxima utilidad como criterio de autenticación en las muestras que contienen DNA amplificable y para los protocolos que de verdad funcionan. En las muestras en las que el DNA es demasiado escaso, esta proporción será muy desfavorable, haciendo más difícil la distinción entre DNA contaminante y DNA endógeno.

También es importante destacar que esta relación aumenta cuando se amplifican fragmentos más cortos. En este estudio, durante la recuperación de los fragmentos de las regiones codificantes no se detectó ningún blanco de extracción contaminado, sin embargo, la eficiencia sí resultó considerablemente aumentada. Lo mismo puede decirse en relación a la contaminación detectada durante la amplificación, ya que se detectaron 3 blancos contaminados (4.76%) de un total de 48 y en cambio la eficiencia resultó ser de 80.04% (385 de 481 alícuotas) con una diferencia estadísticamente significativa (prueba exacta de Fisher bilateral, $P < 0.0001$) y un valor de E/C de 16.81.

Desafortunadamente, en la literatura no existen datos de referencia adecuados en cuanto a la relación E/C. Sin embargo, a medida que se realicen y publiquen más estudios en los que se utilice este criterio, se obtendrá un panorama más claro en cuanto a los valores medios que se deben esperar para determinado tipo de poblaciones y para cada paso del análisis.

Así por ejemplo, para las muestras dentales de la población de la Plaça Vella el valor de esta relación es de 5.5 para la extracción y de 3.14 para la amplificación si se analizan fragmentos grandes (>231 pb) y de 14.84 para la amplificación de fragmentos más cortos (<166 pb); en tanto que para la población de IP es de tan sólo 1 y 0.33 para la extracción y amplificación de fragmentos grandes, pero de 16.14 para la amplificación de los fragmentos cortos. Para las poblaciones con valores muy bajos, las precauciones deberían ser todavía más extremas y los métodos de comprobación más exhaustivos, si cabe.

Asimismo, esta relación puede ayudar a evaluar de manera objetiva la utilidad de los protocolos. Siempre será preferible el protocolo con el valor más alto de E/C. En la Tabla D-1 se muestran estos valores para los protocolos UAB-1, UAB-2, CINVESTAV y MINI, utilizados para la recuperación de los fragmentos de la región de control a partir de piezas dentales.

Protocolos	E/C
UAB-1	2
UAB-2	15.02
CINVESTAV ^a	^b
MINI	1

Tabla D-1 Valores de E/C para distintos protocolos; ^a comprende las tres versiones del protocolo; ^b no es posible calcularlo debido a que no existen grupos de extracción contaminados, pero sería el valor más alto de todos.

Estos resultados demuestran que los dos mejores protocolos serían el UAB-2 y el CINVESTAV, aunque no hay que olvidar que excepto para el protocolo UAB-2, con el resto de protocolos se realizaron muy pocos grupos de extracción.

Contaminación no detectada

La relación E/C nos indica de manera general si en las muestras de la población en estudio existe DNA endógeno y si los protocolos que utilizamos son adecuados. Además nos puede dar un nivel aproximado de la contaminación no detectada que pudo haber ocurrido durante el análisis molecular. Sin embargo, no nos indica que proporción del incremento de la amplificación en las muestras cuyos blancos no resultaron contaminados se debe a la existencia de DNA endógeno y que proporción puede deberse a la contaminación previa al análisis molecular.

En teoría, la contaminación previa puede ser detectada a través del análisis de extracciones independientes de los mismos individuos y para detectar la contaminación ocurrida durante el análisis, es necesario considerar las posibles fuentes. Por un lado, la contaminación puede provenir del investigador que lleva a cabo la extracción y amplificación, en cuyo caso, contar con la secuencia del investigador, facilita en gran medida su detección. Sin embargo, la contaminación ocurrida durante el análisis también puede proceder de DNA extraído previamente o por contaminación cruzada entre las muestras. No obstante, tal como se expone en el Marco Teórico, el mayor riesgo lo representa la contaminación cruzada de amplicones, también llamada contaminación por *carryover*. Este tipo de contaminación tiene la particularidad de que involucra un fragmento específico de DNA y por tanto facilita su identificación, ya que un fragmento discreto no mostrará los marcadores ligados de otras regiones del genoma.

En nuestros resultados, la contaminación de las muestras con el DNA del investigador queda descartada por el hecho de que ninguna de las secuencias obtenidas

resultó ser idéntica a su DNA. Para detectar los otros tipos de contaminaciones se utilizó el criterio de la concordancia secuencia-haplogrupo, que será discutido a continuación.

Concordancia secuencia-haplogrupo

A través de esta concordancia es posible excluir la posibilidad de que la secuencia obtenida y la caracterización del haplogrupo de una muestra se deban a contaminación por *carryover*. La posibilidad de error con este criterio esta dada por la probabilidad de que existan eventos independientes de contaminación cruzada con distintos segmentos y que sean los requeridos para caracterizar un haplogrupo concreto. Aunque no es posible calcular esta probabilidad, no es difícil imaginar que si no existe una intención previa, sería de una magnitud ínfima, pues en dado caso, todos los amplicones tendrían la misma probabilidad de contaminar, produciendo resultados discordantes.

Sin embargo, una contaminación previa ocurrida con DNA total procedente de una sola fuente, no puede ser detectada analizando la concordancia en el mismo extracto, por lo que es necesaria la corroboración del resultado con un segundo extracto, ya sea analizando la secuencia o caracterizando el haplogrupo. El resultado del análisis de los distintos extractos de los individuos PV C-56 y PV C-6 podrían indicar una contaminación con DNA total. No se debe pasar por alto que un evento de contaminación posterior podría invalidar un resultado auténtico previo, por lo que de ser posible es necesario analizar más de dos extractos en los casos contradictorios. Un buen ejemplo de esto lo representan los extractos del individuo PV C-13.

Asimismo, existe otra forma de contaminación que no puede ser detectada con el criterio de la concordancia: si una muestra es contaminada con DNA total que pertenezca al mismo haplogrupo que la secuencia previamente obtenida, se obtendría una concordancia positiva, a pesar de estar analizando dos fuentes distintas de DNA (en este caso podría estar el individuo PV CE-39). Por lo tanto, es muy importante estimar la probabilidad con la que puede ocurrir este evento para saber hasta que punto son fiables los resultados. Esta probabilidad dependerá, por un lado, del riesgo de contaminación y

por otro, del riesgo de que el DNA contaminante pertenezca al mismo haplogrupo que el DNA del que se obtuvo la secuencia. Esta última posibilidad puede ser estimada a través de la probabilidad de que dos individuos elegidos al azar pertenezcan al mismo haplogrupo.

Usando los datos de la Tabla R-29, que muestra la distribución de los haplogrupos en la población de la Plaça Vella, es posible calcular esta probabilidad. La diversidad genética calculada con estos datos indica que la probabilidad de que dos individuos elegidos al azar presenten diferente haplogrupo es de 0.85 (ver Tabla R-33), por lo que la probabilidad de que estos individuos pertenezcan al mismo haplogrupo sería de $1-0.85=0.15$, o lo que es lo mismo, del 15%. Existe pues, una probabilidad del 85% de que dos DNAs elegidos al azar presenten diferente haplogrupo, que serían detectados con el criterio de la concordancia en caso de que contaminasen una muestra. En nuestro proceso de autenticación, 3 de 31 individuos mostraron contradicción en la concordancia secuencia-haplogrupo y si se asume que estos 3 individuos representan el 85% de la contaminación ocurrida, el 15% restante, que no sería detectada debido a que involucra dos DNAs con el mismo haplogrupo, correspondería a 0.52 muestras. En este caso, estas 3.52 muestras serían el resultado total de los procesos de contaminación en las 31 muestras analizadas, lo que corresponde a un 11.35%.

Es necesario abrir aquí un paréntesis para hacer notar que estas contaminaciones no han sido detectadas con los blancos y que podrían estar relacionadas con el porcentaje de contaminación de blancos contaminados si se asume que este porcentaje refleja los procesos de contaminación puntual; hay que recordar que este porcentaje estaría entre el 15.55% de la estimación global y el 7.4% estimado a partir del protocolo con el que se analizó el mayor número de muestras (valor medio de 11.47%).

Continuando con nuestra estimación, podemos decir que una muestra tiene la probabilidad de ser contaminada con un DNA perteneciente a un mismo haplogrupo con una probabilidad de 0.017, ya que este es el resultado de la intersección entre la probabilidad de contaminación (11.35%) y la probabilidad de que el DNA contaminante

pertenezca al mismo haplogrupo que el DNA original (15%). Esto quiere decir que en nuestros datos existe menos del 0.2% de probabilidad de aceptar como auténtico un DNA que no presenta contradicciones entre la secuencia y el haplogrupo, un nivel bastante aceptable para estudios de DNA antiguo, pues en 100 muestras analizadas difícilmente se tendría una muestra en esta situación. Por otra parte, este análisis demuestra que para el proceso de autenticación es más relevante la proporción de contaminaciones detectadas que el riesgo de que una muestra sea contaminada con un DNA del mismo haplogrupo.

Por otra parte, cabe hacer notar que en las poblaciones con una menor diversidad de haplogrupos (como las nativas americanas) el riesgo de este error será mayor, pues la probabilidad de encontrar dos DNAs mitocondriales pertenecientes al mismo haplogrupo es más alta.

Otro posible error con esta aproximación está dado por la posibilidad de que un individuo haya sufrido una mutación en uno de los sitios diagnósticos del haplogrupo mostrando una discordancia aparente, lo que podría llevarnos a descartar un resultado auténtico. No obstante, un individuo en esta situación debería mostrar concordancia en la designación del haplogrupo con dos extractos independientes, lo que ayudaría a discernir si en verdad existe contaminación o se trata de un polimorfismo propio del individuo o incluso un error de la *Taq*. El individuo PV C-72, es un claro ejemplo de esta posibilidad. Este individuo presenta la sustitución 16294T, que puede ser característica del haplogrupo T, sin embargo, el análisis con enzimas de restricción en dos extractos, mostró que este individuo pertenece al haplogrupo H. Por lo tanto en este individuo la secuencia puede ser una excepción a la correspondencia entre secuencia y haplogrupo, o bien, puede tratarse de un error de la *Taq*. Para descartar esta última posibilidad sería necesario la repetición de la secuencia. Lo importante, sin embargo, es el hecho de que caracterizando el haplogrupo con dos extractos independientes se puede evitar el error de descartar un resultado auténtico.

Por otra parte, estas consideraciones demuestran que el criterio de la concordancia alcanza su máximo nivel de certeza con las muestras en las que se cumple la

correspondencia, pues como se expuso en la sección de material y métodos, la probabilidad de que se produzcan los errores necesarios para obtener la correspondencia es completamente negligible. En cambio, si una muestra no presenta la concordancia, no necesariamente será por un evento de contaminación, aunque la probabilidad de que lo sea es bastante alta. Para minimizar este error, son necesarios análisis posteriores, como en el caso del individuo PV C-72.

Ahora bien, ¿qué porcentaje de muestras puede estar en esta situación? Esta no es una pregunta trivial, pues en realidad, el criterio de la concordancia se basa en la presunción de que la correspondencia ocurre en cerca del 100% de las muestras y por lo tanto que las excepciones serán muy pocas. Por el momento no se han hecho estimaciones con significación estadística para saber qué porcentaje de muestras cumple con la correspondencia y qué porcentaje no lo hace.

En el artículo de Torroni y col. (1996), en el que se analizó la correspondencia en 49 muestras, es posible encontrar un individuo (38) con el polimorfismo 16294T, que no pertenece al haplogrupo T, no obstante, sí presenta otras substituciones características del haplogrupo al que en realidad pertenece según los sitios de restricción que contiene. En un estudio previo con población senegalesa, Graven y col. (1995) encontraron tres posibles excepciones al analizar 119 individuos. Stone y Stoneking (1998), encontraron que un individuo de los 52 de los que se contaba con la secuencia y el haplogrupo y que pertenecía al haplogrupo A, presentaba una secuencia relacionada con secuencias caucásicas. No obstante, los autores han hecho notar que la ganancia del sitio *HaeIII*, que caracteriza e haplogrupo A, ha sido encontrada por otros autores (Cann et al. 1987a) en dos individuos caucásicos. Por lo tanto, parece plausible que las inconsistencias están relacionadas con casos aislados e involucrando sitios inestables.

Por ejemplo, en los datos de Torroni y col. (1996) se observan 3 individuos (de 49) que presentan un estatus contradictorio en la posición diagnóstica 73, del segmento hipervariable II, en tanto que Salas y col. (1998) informan de la detección de 7 individuos de 35 que contienen una substitución diferente a la esperada en esta posición,

poniendo en duda la validez de este sitio para la definición del haplogrupo de una muestra, a pesar de que en su estudio no se realizó la determinación del haplogrupo con enzimas de restricción. De igual forma, Izagirre y De la Rúa (1999) han detectado inconsistencias en esta posición, aunque no especifican en cuantas muestras encontraron la contradicción ni sobre cuantos individuos está basado este análisis. Por otra parte, en el extenso trabajo publicado por Macaulay y col. (1999a) no se hace especial referencia a la posibilidad de que no se cumpla la concordancia secuencia-haplogrupo, lo que puede indicar que la consideran mínima. Finalmente, como se muestra en la sección de resultados, en las secuencias de DNA de la población catalana actual, no se han detectado inconsistencias en el segmento analizado, aunque éste no incluye la posición 73.

En cualquier caso, resulta de vital importancia continuar analizando muestras utilizando los dos métodos (secuenciación y enzimas de restricción), para verificar hasta que punto se cumple la correspondencia, detectar los sitios más estables y establecer intervalos de confianza estadística. Por el momento, podemos decir que nuestros resultados se ajustan bastante a lo predicho por la concordancia, presentando problemas únicamente las muestras que han demostrado contener algún tipo de contaminación, y en el caso particular del individuo PV-C72, que no escapa de manera determinante con lo previsto, en vista de que se ha encontrado por lo menos un caso similar en DNA moderno (Torroni et al. 1996).

Finalmente, es necesario destacar también otra característica del criterio de la concordancia que puede estar sujeta a error. En este trabajo se definieron los haplogrupos analizando 9 sitios de restricción, no obstante, algunos haplogrupos se distinguen entre sí únicamente por uno de estos sitios. Esto dificultaría en un momento dado, distinguir una contaminación por *carryover*, ya que un amplicón con el polimorfismo que diferencia un haplogrupo de otro, podría contaminar una muestra y no producir resultados contradictorios en el resto de los sitios polimórficos. Tal es el caso de los haplogrupos X e I que se distinguen solamente por el sitio 10394 *DdeI*. Por tanto, es recomendable que en estudios posteriores se utilice un número mayor de sitios de restricción, incluyendo por lo menos dos que sean específicos para cada haplogrupo. Por otra parte, en el análisis

realizado por Macaulay y col. (1999a), se ha descubierto que algunos de los sitios que se han utilizado en este trabajo son algo inestables, lo que hace todavía más recomendable la utilización de sitios adicionales.

Por otro lado, a pesar de los inconvenientes descritos, el criterio de la concordancia ha demostrado ser una herramienta única para la autenticación del DNA extraído. En este trabajo se han detectado contaminaciones que no fueron detectadas con los blancos y en el caso del individuo IPC1AA, se detectó una contaminación que no habría sido detectada en el caso de utilizar sólo un método de caracterización, ya sea la secuenciación o la determinación del haplogrupo.

Otra de las virtudes de este método, es que nos permite reconocer tipos mitocondriales que a pesar de presentar la misma secuencia son diferentes, y a la inversa, es posible diferenciar DNAs que pertenecen a un mismo haplogrupo. Por ejemplo, de los 7 individuos que presentaron una secuencia idéntica a la de referencia, tres pertenecen al haplogrupo H, 3 al J y uno a “otro”, lo que indica que por lo menos existen 3 DNAs diferentes en estos 7 individuos. Por otra parte, en los 8 individuos que pertenecen al haplogrupo H existen 5 secuencias diferentes, lo que nos indica que en estos individuos existen por lo menos 5 tipos mitocondriales distintos. Este hecho es de la mayor relevancia para el proceso de autenticación, pues cuando se analiza únicamente el haplogrupo, existe el riesgo de estar amplificando un solo DNA contaminante y no habría manera de distinguirlo si no se conoce la secuencia particular de cada uno de ellos. Lo mismo puede ocurrir con la secuenciación, ya que podría amplificarse el mismo DNA contaminante a partir de distintas muestras sin posibilidad de detección, pero si dos secuencias iguales pertenecen a diferente haplogrupo, entonces el DNA no procede de la misma fuente. Utilizando los resultados conjuntos de secuencia y haplogrupo, es posible distinguir por lo menos 15 DNAs de procedencia diferente en los 24 individuos de la Plaça Vella cuyo DNA fue autenticado. Este dato resulta importante a la hora de valorar la posibilidad de que el DNA obtenido provenga de fuentes naturales que hayan podido contaminar las muestras antes del análisis molecular.

Análisis de la variabilidad

Como se ha visto en el análisis filogenético, la variabilidad encontrada en la población de la Plaça Vella se encuentra dentro del rango esperado para una población europea. Este hecho puede ser utilizado como un criterio de autenticidad, no obstante, es necesario hacer algunas consideraciones en cuanto a la utilización de este criterio dentro del proceso de autenticación. La principal objeción que puede hacerse es la cuestión de si una población europea antigua debe mostrar el mismo grado de variabilidad que el de una población natural actual. En respuesta a este punto, es imprescindible recordar que el acervo del mtDNA europeo fue establecido probablemente en el paleolítico y que las poblaciones europeas presentan una homogeneidad inusual en cuanto a la variabilidad genética mitocondrial, tal como se ha expuesto en el Marco de Referencia. Por lo tanto, es bastante razonable esperar que la variabilidad que existe entre diferentes poblaciones sea similar, aunque no idéntica. Esta distinción puede ser la clave en la pretensión de utilizar esta característica del DNA obtenido como criterio de autenticidad. En este sentido, es de fundamental importancia definir un marco de referencia comparativo que permita realizar una comparación válida. En este trabajo se han utilizado diversas muestras de poblaciones europeas y se ha definido un rango de variabilidad utilizando distintos índices. Por ejemplo, para el segmento de 153pb, el índice de diversidad genética varía de 0.7387 a 0.9768 y el índice de diversidad nucleotídica de 0.01262 a 0.02478. Los valores respectivos encontrados en la Plaça Vella son de 0.8949 y 0.01459, por lo que puede considerarse que esta muestra presenta una variabilidad dentro del rango esperado para una población natural europea.

Desafortunadamente no nos es posible estimar con qué probabilidad o con que verosimilitud una población concreta mostrará una variabilidad comprendida en este rango, pero tomando en cuenta que las poblaciones estudiadas son muy diversas en cuanto a su origen étnico y geográfico, podemos esperar que esta probabilidad sea alta. Por otra parte, la esencia de este criterio consiste en brindar un dato comparativo y el otro marco de referencia estaría dado por la probabilidad o la verosimilitud de obtener una

variabilidad similar originada por eventos de contaminación. Esta probabilidad es mucho más difícil de estimar, sin embargo, es una probabilidad que será menor en la medida en la que la población en estudio sea amplia y el tamaño de la muestra sea importante.

Estudios posteriores en cuanto a la variabilidad que puede generarse a través de eventos de contaminación podrían ayudar a obtener un panorama más claro. Estos estudios podrían llevarse a cabo amplificando segmentos de DNA específicos de la especie humana en restos animales, preferentemente en restos arqueológicos que hayan pasado por procesos equivalentes a los que son sometidos los restos humanos.

Razón de verosimilitud

Bajo la premisa de que el DNA antiguo existe y de que es posible recuperarlo, la principal hipótesis a probar es que el DNA obtenido sea auténtico. En este sentido, debemos estimar por un lado, la probabilidad o verosimilitud de que el DNA extraído sea endógeno y por otro, la probabilidad o verosimilitud de que este DNA sea exógeno o de origen quimérico.

En apoyo del origen endógeno podemos decir que la muestra obtenida ha pasado cinco niveles de autenticación, resumidos en la Fig. D-1 y expuestos a continuación:

- La contaminación generalizada de los reactivos fue excluida descartando todos los grupos de extracción y amplificación que mostraron una banda del tamaño esperado en el gel de agarosa. La posibilidad de tener una contaminación importante no detectada también quedó excluida mediante las secuencias obtenidas de las poblaciones control, ya que la mayoría de ellas resultaron ser únicas, no compartidas con individuos de la población en estudio.
- El análisis de la relación eficiencia-contaminación, demuestra que la mayor parte de los resultados no provienen de contaminaciones puntuales.

- La contaminación debida a *carryover* puede ser detectada y descartada mediante la concordancia entre la secuencia de la región de control y la caracterización del haplogrupo en el mismo extracto. Como se ha expuesto, sería muy poco probable que contaminaciones subsecuentes de amplicones fuesen completamente concordantes, ya que deben provenir de segmentos diferentes de la molécula del mtDNA. Es importante notar que con la repetición de la secuenciación directa de la misma muestra amplificada no es posible descartar este tipo de contaminación.

- Analizando extractos independientes del mismo individuo, se obtuvo un cuarto nivel de autenticidad, mediante la concordancia intra-individual entre la secuencia y el haplogrupo. Los individuos con resultados discrepantes fueron excluidos. La contaminación con mtDNA íntegro procedente de los reactivos o de las personas que han manipulado las muestras en los distintos pasos, es improbable, debido a que los extractos independientes de diferentes piezas dentales del mismo individuo dieron resultados coincidentes.

- Después de eliminar todos los individuos que no cumplieron con los criterios previos, la muestra presenta una distribución de haplogrupos compatible, en general, con la que presentan otras poblaciones europeas. Asimismo, la variabilidad encontrada en las secuencias de la región de control es la esperada en poblaciones europeas. De esta forma queda excluida la posibilidad de contaminación proveniente de las personas que han manipulado las muestras. Este tipo de contaminación, produciría un patrón más homogéneo.

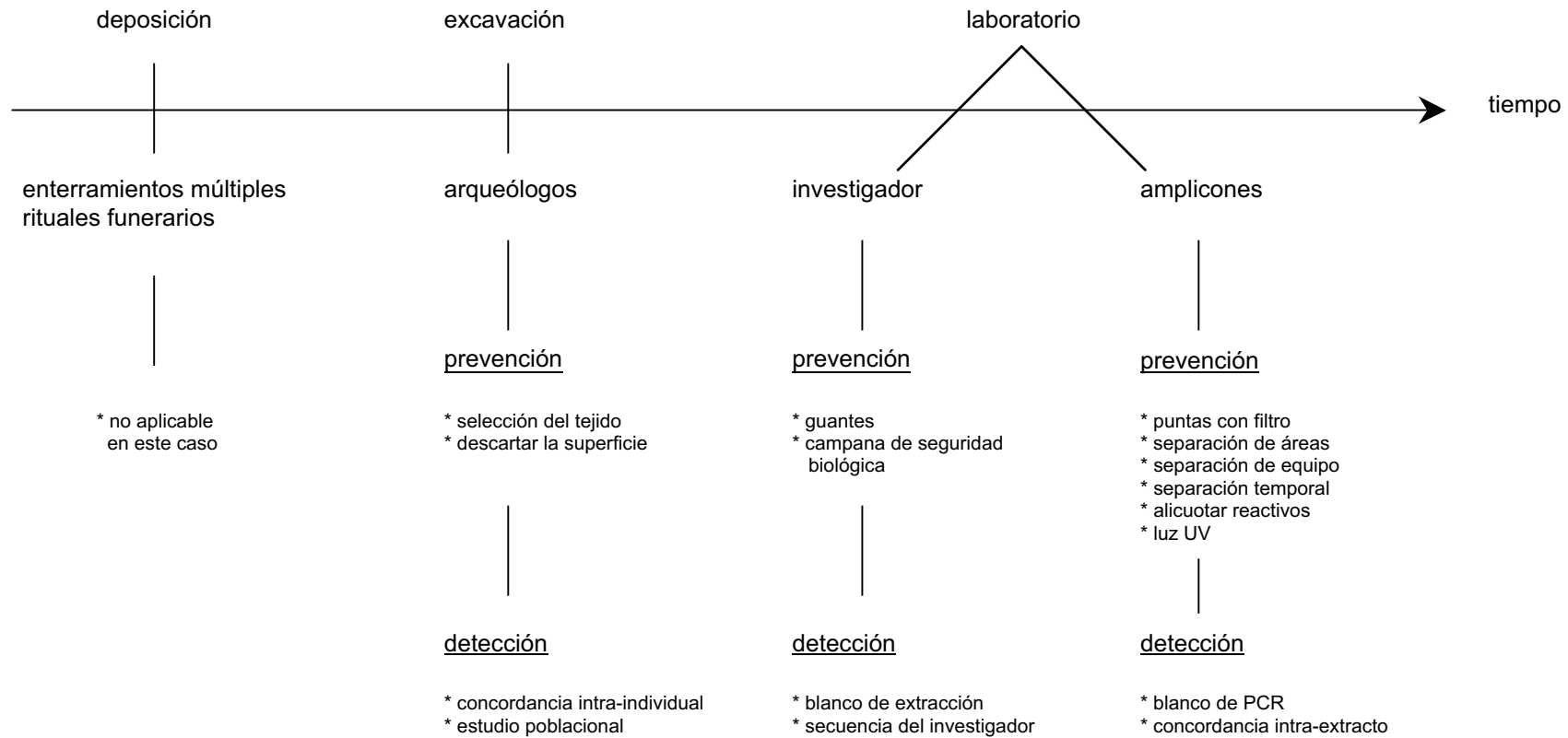


Figura D-1 Posibles fuentes de contaminación en distintos procesos, su prevención y detección. Modificado de Francalacci (1995).

Por otra parte, los datos finalmente obtenidos podrían ser interpretados en términos de contaminación únicamente bajo los siguientes supuestos:

- Solamente 24 de los 52 individuos estudiados fueron contaminados con al menos 15 fuentes diferentes de mtDNA.
- La parte interna de todos los dientes analizados de cada individuo fue contaminada con un sólo mtDNA de estos 15.
- La contaminación se produjo de tal forma que la variabilidad resultante presenta todas las características esperadas en una población natural.

En nuestra opinión, este escenario presenta características poco razonables, que sólo podrían producirse en caso de que existiera una intencionalidad previa, ya que el tejido elegido para la extracción del DNA, el tejido de la cavidad pulpar, es el menos propenso a la contaminación con DNA exógeno. Por lo tanto, consideramos que es posible evaluar la autenticidad del DNA mitocondrial obtenido de muestras que cumplan todos los criterios antes expuestos y concluimos que existe una muy alta probabilidad de que el mtDNA presentado en este trabajo sea auténtico.

El “almacenamiento frío” y los inhibidores del PCR

El efecto del frío

Los datos obtenidos en este trabajo demuestran que el almacenamiento de los extractos a 4°C, como paso previo al PCR, incrementa significativamente la eficiencia de la amplificación, independientemente de los fenómenos que ocurran durante este periodo de almacenamiento frío. Como primera hipótesis de trabajo se ha asumido que con este procedimiento ocurre una separación de las sustancias inhibitorias del DNA, sin embargo, el incremento en la eficiencia también podría ser explicado en el caso de que el frío modificara de alguna manera las sustancias contaminantes de tal forma que les haga perder su poder inhibitorio sin perder su solubilidad. No obstante, las diferencias en cuanto a la cantidad de residuos que se detectan cuando se aplica el “almacenamiento frío” respecto a la cantidad que es detectada cuando no se aplica este paso, apoyan la hipótesis de la pérdida de la solubilidad. Por otra parte, existe la posibilidad de que la precipitación de las sustancias se deba únicamente al tiempo transcurrido durante el almacenamiento, en cuyo caso, el frío no tendría ningún efecto. Montiel y col. (1997) descartaron esta posibilidad sometiendo los extractos a diferentes tiempos de centrifugación con el fin de promover la precipitación de las sustancias inhibitorias, sin observar ningún efecto positivo. Parece por tanto, que el efecto de la baja temperatura tiene un papel decisivo en el aumento de la eficiencia observada después de este tratamiento.

Caracterización de los residuos

Las consideraciones anteriormente expuestas apoyan la hipótesis de que el frío promueve la separación de las sustancias inhibitorias reduciendo su solubilidad y bajo estas condiciones es razonable esperar que los residuos sean polímeros polares, ya que

este tipo de compuestos podría formar micelas o unirse a las paredes del tubo de polipropileno por virtud de sus cargas libres. Por otra parte, el hecho de que en el protocolo de Persson (1992) y en el protocolo UAB-2 se realice una diálisis centrífuga con membranas de corte de peso molecular de 30,000 Da, indica que el peso molecular de estas sustancias debe ser igual o mayor de 30,000 Da.

Los espectros de los residuos de huesos y dientes, demuestran que existen diferencias en la composición de las sustancias que se extraen de cada tipo de muestra. Sin embargo, también es cierto que en ambos casos, las sustancias presentan características similares, lo que nos ha llevado a considerar la posibilidad de que las sustancias inhibitorias formen parte de un conjunto de sustancias relacionadas. De las sustancias propuestas como posibles inhibidores, los productos Maillard y los ácidos húmicos y fúlvicos, son en realidad conjuntos de sustancias que comparten características comunes. En efecto, los ácidos húmicos y fúlvicos no presentan una forma química definida pues son un conjunto de compuestos poliméricos que comparten características comunes y están definidos únicamente por los procesos de extracción utilizados (Stevenson, 1982, p. 36-53). Por otra parte, los productos Maillard también pueden presentar variaciones, pues surgen de la condensación de los aminoácidos con los azúcares reductores y no hay que olvidar que existen 20 aminoácidos distintos. Los compuestos porfirínicos, por su parte, parecen ser un poco más homogéneos en su composición y en sus características espectrofotométricas.

Mediante el análisis comparativo, se ha descartado la posibilidad de que los residuos de los extractos sean compuestos porfirínicos, pues no presentan ninguna de las características previamente publicadas. Más aún, el compuesto porfirínico analizado en nuestro laboratorio presenta características muy diferentes a las encontradas en los residuos.

La gran variedad de respuestas espectrofotométricas que presentan los productos Maillard, no nos permite descartar la posibilidad de que los residuos pertenezcan a esta “familia” de compuestos. Es necesario investigar más sobre las propiedades de estos

compuestos, no sólo sobre sus propiedades quimioluminiscentes, sino también sobre su posible efecto en la reacción de PCR. Desde un principio se ha hipotetizado que la inhibición que presentan los extractos de DNA antiguo se debe a la presencia de los productos Maillard, sin embargo, no se ha investigado si efectivamente estas sustancias tienen propiedades inhibitorias. En este sentido, sería deseable añadir concentraciones conocidas de estos compuestos en reacciones de PCR estandarizadas para observar su poder inhibitorio, en caso de que lo tengan.

Los ácidos fúlvicos y húmicos presentaron características espectrofotométricas similares a las encontradas en los residuos de los extractos. Particularmente, resulta sorprendente la presencia de los picos invertidos en los espectros de absorción exactamente en las mismas longitudes de onda que los observados en los residuos. Por otra parte, aunque la similitud no fue exacta en los espectros de emisión, las características observadas en los ácidos dejan abierta la posibilidad de que alguna sustancia de las que forman el conjunto de estos compuestos corresponda a la encontrada en los extractos de DNA, tanto en huesos como en dientes. Puntualizando un poco más, la evidencia obtenida en el análisis comparativo es más robusta en favor de los ácidos húmicos, ya que éstos presentaron picos de absorción a 240 y 289nm al igual que los residuos; y al ser analizados en baja concentración, presentaron un pico de emisión alrededor de los 450nm (excitando a 390nm), que podría corresponder con el observado en los residuos.

Por otra parte, el poder inhibitorio de los ácidos fúlvicos y húmicos fue probado en nuestro laboratorio añadiendo diferentes concentraciones de ambos ácidos en PCR's con DNA moderno (230ng) de alta pureza (1.9). La inhibición fue total con 100ng de ácidos húmicos (en reacciones de 50 µl), en tanto que 200ng de ácidos fúlvicos produjeron tan sólo una inhibición moderada (Fig. D-2), indicando que los ácidos húmicos tienen un mayor poder inhibitorio. En este punto, cabe mencionar que la distinción puede ser importante para la optimización de los protocolos de extracción de DNA, ya que los ácidos húmicos se caracterizan por ser solubles únicamente a pH > 8, en tanto que los fúlvicos pueden solubilizarse a un pH más bajo, de tal forma que el control del pH de los tampones puede ser útil en la eliminación de estas sustancias en alguna fase de la

extracción. Esta es una vía de investigación importante que queda abierta para estudios posteriores.

No obstante, a pesar de las evidencias descritas, no podemos ser concluyentes en cuanto a la identidad del inhibidor, ya que son necesarios otros métodos analíticos de mayor precisión, como la pirólisis y la cromatografía de gas/espectrometría de masa, para una caracterización inequívoca de estas sustancias. Además, no hay que olvidar los estudios en los que se han detectado inhibidores del PCR que no provienen del suelo, como los de Waite y col. (1997) y Scholz y col. (1998), referidos en el Marco Teórico. Aunque tampoco debemos pasar por alto que los ácidos húmicos, al igual que los productos Maillard, son productos de la degradación orgánica y que por lo tanto existe la posibilidad de que puedan ser producidos *in situ*, lo que explicaría convenientemente la presencia de estas sustancias en el interior de los dientes, pues la dureza de sus tejidos hace improbable que provengan del suelo. Cabe pues preguntarse si las sustancias inhibitorias encontradas en los extractos de piezas dentales pueden representar “precursores” de sustancias tales como los ácidos húmicos o fúlvicos o eventualmente, productos de la reacción Maillard.

Ácidos húmicos y “almacenamiento frío”

Dejando de lado la cuestión de la caracterización de las sustancias inhibitorias, uno de los aspectos más interesantes de nuestras investigaciones es el efecto del almacenamiento en nevera sobre las sustancias inhibitorias. En este sentido, es importante preguntarse si la acción del frío también es eficiente con sustancias de reconocido poder inhibitorio bajo condiciones controladas. Los ácidos húmicos suelen estar presentes en las extracciones de DNA de bacterias del suelo (Tsai y Olson, 1992; Tebbe y Vahjen, 1993; Chandler et al. 1997); presentan características muy similares a las que presenta el inhibidor de los extractos de restos antiguos y algunos autores han comprobado directamente su poder inhibitorio en el PCR, añadiendo cantidades conocidas

de estas sustancias en la mezcla de reacción (Schwieger y Tebbe, 1997). Además, como se ha mencionado anteriormente, en nuestro laboratorio también se ha comprobado la acción inhibitoria de los ácidos húmicos.

Parecía pues interesante corroborar la acción del frío sobre los ácidos húmicos y su papel inhibitorio durante el PCR. Con este fin se agregaron distintas concentraciones de ácidos húmicos (2, 2.8, 3.6, 5.2, 6.8 y 8.4 $\mu\text{g/ml}$) a distintas alícuotas de un extracto de DNA moderno de elevada concentración y pureza, se realizó una primera reacción de PCR, se guardaron las soluciones a 4°C durante 3 días y finalmente se realizó una segunda reacción de PCR. El resultado fue que en la primera ronda de PCR, se observó inhibición a partir de la solución que contenía 3.6 $\mu\text{g/ml}$ de ácidos húmicos (Fig. D-3), en tanto que en la segunda ronda, en la que se utilizaron las soluciones que habían sido almacenadas a 4°C, no se observó inhibición ni siquiera en la amplificación de la muestra que contenía 8.4 $\mu\text{g/ml}$ (Fig. D-4). Resultados similares fueron obtenidos utilizando DNA de Pasteurela hemolítica lo que indica que tanto el poder inhibitorio de los ácidos húmicos como la acción benéfica del almacenamiento frío, son independientes del tipo y secuencia del DNA amplificado y de la secuencia de los *primers* utilizados.

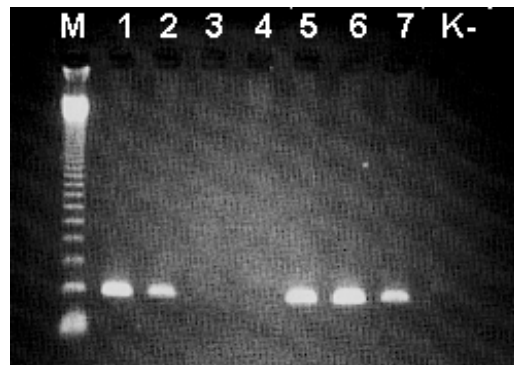


Figura D-2 Poder inhibitorio de los ácidos húmicos y fúlvicos al ser añadidos en reacciones de PCR de 50 μl . Carriles: M= múltiplos de 123 pb (Sigma) 1= Control positivo 2= 50 ng de ácidos húmicos 3= 100 ng de ácidos húmicos 4= 200 ng de ácidos húmicos 5= 50 ng de ácidos fúlvicos 6= 100 ng de ácidos fúlvicos 7= 200 ng de ácidos fúlvicos K-= blanco de PCR

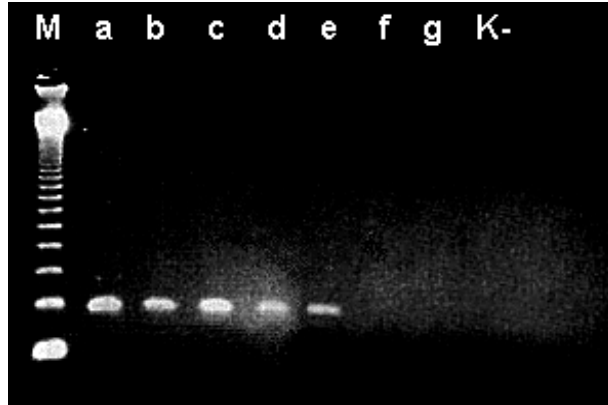


Figura D-3 Amplificación de las soluciones de DNA y ácidos húmicos a distintas concentraciones antes del “almacenamiento frío”. Carriles: M= múltiplos de 123 pb (Sigma) a= Control positivo b= 2 $\mu\text{g/ml}$ c= 2.8 $\mu\text{g/ml}$ d= 3.6 $\mu\text{g/ml}$ e= 5.2 $\mu\text{g/ml}$ f= 6.8 $\mu\text{g/ml}$ g= 8.4 $\mu\text{g/ml}$ K-= blanco de PCR

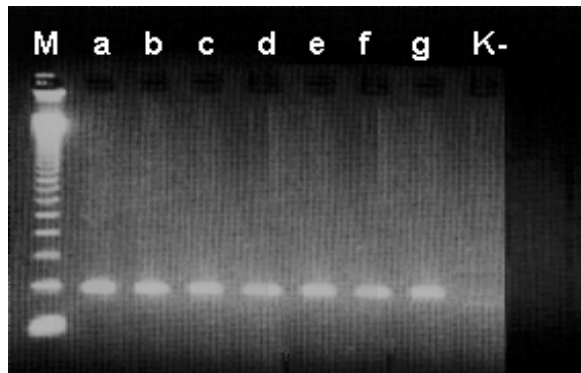


Figura D-4 Amplificación de las soluciones de DNA y ácidos húmicos después de 3 días a 4°C. Carriles: M= múltiplos de 123 pb (Sigma) a= Control positivo b= 2 $\mu\text{g/ml}$ c= 2.8 $\mu\text{g/ml}$ d= 3.6 $\mu\text{g/ml}$ e= 5.2 $\mu\text{g/ml}$ f= 6.8 $\mu\text{g/ml}$ g= 8.4 $\mu\text{g/ml}$ K-= blanco de PCR

Límites del “almacenamiento frío”

A través de los resultados obtenidos en este trabajo y tomando en cuenta los puntos discutidos en esta sección, es posible definir los límites del método del “almacenamiento frío”. Existen ciertas condiciones previas para poder observar el beneficio de este método:

- en la muestra debe existir DNA endógeno
- los métodos de extracción deben estar optimizados

-
- la concentración de las sustancias inhibitorias debe estar dentro de un rango determinado

Este último punto es esencial, ya que es muy probable que exista un umbral de inhibición, y la cantidad que puede ser absorbida por las paredes del tubo debe tener un límite. En este sentido es fundamental la optimización de la cantidad de agua final en la que queda disuelto el extracto, de tal forma que la precipitación de una cierta cantidad de sustancias permita superar el umbral inhibitorio. Algunos autores recomiendan la dilución de la muestra para reducir la concentración de los inhibidores, sin embargo, de esta forma se diluye también el DNA, a veces por debajo del nivel de detección, pero también puede conllevar los problemas asociados cuando existe una muy baja concentración de DNA endógeno, potenciando fenómenos como el Jumping PCR, o permitiendo que las moléculas contaminantes cobren mayor relevancia. Diluir el extracto hasta cierto punto y aplicar el “almacenamiento frío” puede ser un método de optimización ideal bajo estas condiciones, pues permitiría una dilución moderada de la muestra de tal forma que la concentración de las sustancias inhibitorias quede dentro del rango de acción del “almacenamiento frío”, pero que por otra parte conserve una concentración adecuada del DNA extraído.

Es importante notar que no es posible definir un protocolo general para este método, pues dependerá de las características particulares de las muestras en estudio, por lo que para cada tipo de muestras será necesario realizar una optimización.

La gran ventaja de este método es que nos permite mejorar la eficiencia de la amplificación sin introducir ningún paso que involucre la manipulación de las muestras o la necesidad de añadir reactivos o soluciones adicionales que puedan representar fuentes de contaminación.

Mecanismo inhibitorio de las moléculas polares

Los ácidos húmicos y fúlvicos presentan una estructura química compleja con muchas posiciones coordinantes que pueden interactuar con cationes fácilmente. Estas posiciones pueden ser usadas para coordinar con otras sustancias en el suelo, o incluso con el DNA, en presencia de cationes polivalentes. Dicha propiedad podría explicar la afinidad que presenta el DNA por los ácidos húmicos descrita por algunos autores (Crecchio y Stotzky, 1998). Estas consideraciones nos conducen inevitablemente a hipotetizar sobre el mecanismo de acción de las sustancias inhibitorias durante el PCR, el cual podría involucrar su unión con los cationes de magnesio que se añaden en la reacción para utilizarlos como “puentes” de enlace con el DNA.

RECONSTRUCCIÓN FILOGENÉTICA

El análisis filogenético realizado en este trabajo nos permite obtener conclusiones en dos ámbitos principales. Uno de ellos concierne al punto de vista metodológico abordando cuestiones como las diferencias entre los resultados obtenidos con la distribución de los haplogrupos y los resultados obtenidos a partir del análisis de las secuencias de la región de control, o como lo relativo al contenido de información que poseen distintos segmentos del fragmento hipervariable I. El otro ámbito está relacionado con la reconstrucción filogenética en sí misma, de la que se pueden obtener conclusiones sobre la historia demográfica de la población Catalana y su relación con el resto de poblaciones europeas, en el contexto de la historia común que las une.

Consideraciones metodológicas

Haplogrupos vs. Secuencias

La reconstrucción filogenética a partir de la frecuencia de los haplogrupos produjo árboles en los que las poblaciones con altas frecuencias del haplogrupo H aparecían cercanamente relacionadas y considerablemente separadas del resto; como en el caso de Galicia y Barbagia. Dado que el aumento en la frecuencia de los haplogrupos, puede ser muy sensible a la deriva genética, estos árboles representan en realidad la diferenciación o similitud de las poblaciones, pero no sus relaciones filogenéticas. En este contexto parece más apropiado utilizar la prueba exacta de diferenciación poblacional para comparar las poblaciones con este tipo de datos. La prueba exacta nos indica si las poblaciones son diferentes o no, sin que la diferencia o similitud implique una relación filogenética. Con

este análisis hemos encontrado que las poblaciones más diferenciadas son las de Barbagia y Galicia seguidas de la del País Vasco.

A pesar de esto, la presencia o ausencia de determinados haplogrupos en las poblaciones si puede darnos cierta información acerca de sus relaciones filogenéticas, por ejemplo, la presencia de muestras del haplogrupo V en la población de Barbagia, evidencia su relación con poblaciones ibéricas mediterráneas y atlánticas (Francalacci et al. 1999).

También se debe considerar que las distancias genéticas utilizadas están concebidas para analizar conjuntamente la frecuencia de alelos en varios locus. Al reducir toda la información de los haplogrupos considerándolos alelos de un mismo locus no sólo se pierde información, sino que ésta se analiza de una forma poco adecuada. Por estos motivos, no recomendamos la utilización de estos métodos si no se realiza un análisis con enzimas de restricción de alta resolución que caracterice las muestras a nivel de haplotipo y no sólo de haplogrupo. Existen métodos de reconstrucción filogenética apropiados para este tipo de datos (ver documentación del programa Phylip). No obstante, en muestras antiguas, el análisis de alta resolución puede complicarse al extremo, ya que requiere la amplificación de todo el mtDNA en fragmentos solapados, para lo cual sería necesario amplificar alrededor de 100 fragmentos de unos 200 pb cada uno.

En contraste, la reconstrucción filogenética de las poblaciones mediante la transformación del estadístico F_{ST} a partir de las secuencias de la región de control ha resultado ser un método más completo y más adecuado. Con éste método se valoran las relaciones de cada uno de los haplotipos que conforman cada población con los que presentan las otras poblaciones; siendo posible utilizar un sofisticado modelo, desarrollado específicamente para la región de control del mtDNA humano, en la evaluación de estas relaciones. La transformación de Reynolds del estadístico F_{ST} estimado mediante la distancia de Tamura-Nei, no produce distancias negativas entre las poblaciones, contrariamente a lo obtenido con otros métodos. Además, los árboles de Neighbor-Joining construidos a partir de esta distancia entre poblaciones, están en

perfecto acuerdo con estudios previos y muestran una mejor diferenciación de las poblaciones europeas.

Por otra parte, la información contenida en la frecuencia de haplogrupos también es insuficiente a la hora de valorar la diversidad genética, lo que es claramente observable a través de la comparación del índice de diversidad calculado a partir de las frecuencias, con el obtenido a partir de las secuencias de la región de control. El índice de diversidad genética estimado a partir de la frecuencia de los haplogrupos en las poblaciones europeas cae en un rango que va de 0.5707 de la población de Barbagia, hasta 0.8478 de la población de la Plaça Vella; en tanto que el mismo índice estimado a partir de las secuencias de la región de control presenta valores más altos, en un rango menor, que va de 0.9295 de Galicia a 0.9939 de Turquía. En la Tabla D-2 se observa que el cambio en el índice de diversidad no sólo es cuantitativo, sino también cualitativo, ya que las poblaciones de Toscana y Cataluña presentan un orden invertido en la jerarquía de la diversidad. Mientras que con la frecuencia de haplogrupos la población Toscana presenta mayor diversidad que la Catalana; con la estimación a partir de las secuencias, la población Catalana tiene un índice mayor que el de la población Toscana.

Población	N	Diversidad genética	
		Haplogrupos	Secuencias
GALI	92	0.5824	0.9295
VAS	45	0.6980	0.9485
CAT	90	0.7538	0.9867
TOS	48	0.7863	0.9672

Tabla D-2 Índice de diversidad genética (Nei, 1987) estimado a partir de la frecuencia de haplogrupos y de la frecuencia de haplotipos (fragmento de 360pb). La estimación se realizó utilizando las mismas muestras. La población de la Plaça Vella no se incluye pues para la estimación del índice de diversidad a partir de la secuencia en esta población se utilizó el fragmento de 153pb.

Tamaño de los fragmentos

El análisis de fragmentos de distintos tamaños, nos ha permitido valorar la efectividad del fragmento pequeño en la inferencia de la historia demográfica y en la reconstrucción filogenética. Este análisis resulta fundamental teniendo en cuenta que en los estudios de DNA antiguo los fragmentos cortos son recuperados con mayor eficiencia.

A través de los índices de diversidad observamos que el contenido de información a nivel nucleotídico en el fragmento de 153pb comprendido entre las posiciones 16210 y 16362 de la región de control, es relativamente mayor al que contienen los fragmentos más grandes. Asimismo, se ha detectado poca diferencia entre los fragmentos de 302 y 360 pares de bases. No obstante, a medida que se reduce el fragmento analizado, se observa una pérdida de información haplotípica, lo que tiene como consecuencia que el fragmento de 153pb sea poco adecuado para determinados tipos de análisis.

A través de la comparación de los resultados obtenidos con los distintos tamaños de segmento, es posible observar hasta que punto se ve alterada la calidad de la información. En general se ha observado que para la mayoría de poblaciones los resultados de los índices de diversidad son consistentes analizando los tres segmentos, no obstante, algunas poblaciones muestran características diferentes dependiendo del segmento analizado. Por ejemplo, la población BRI muestra resultados contradictorios en cuanto al índice de diversidad genética según el segmento analizado, en tanto que las poblaciones de Galura y Cerdeña muestran inconsistencias en los resultados del índice de diversidad nucleotídica estimado a partir de los segmentos de distintos tamaños. Estas dos poblaciones, Galura y Cerdeña, también mostraron inconsistencias en cuanto a la diferencia media por parejas.

Las consecuencias de esta pérdida de información pueden verse en los resultados tanto de la inferencia de la historia demográfica como de la reconstrucción filogenética. Por ejemplo, el orden cronológico de la expansión poblacional cambia para las poblaciones de

Galura y Cerdeña, en tanto que en los árboles filogenéticos la situación de estas poblaciones también cambia de acuerdo al tamaño del fragmento analizado. El hecho de que estos cambios no involucren al resto de poblaciones indica que el fragmento de 153pb aquí analizado es en general un segmento adecuado para estudios de DNA antiguo. No obstante, resulta imprescindible contar con información previa de la población en estudio para observar si puede haber sesgo en los resultados como en el caso de las poblaciones de Galura y Cerdeña, para las cuales no sería suficiente recuperar el fragmento de 153pb según los resultados aquí expuestos.

En cuanto a la reconstrucción filogenética intrapoblacional, nuestros resultados muestran que el análisis de algunos haplogrupos se ve desfavorecido con el fragmento de 153pb, en tanto que para otros no resulta substancialmente alterado. En este sentido, es importante tener en cuenta que si se pretende realizar un análisis intrapoblacional con DNA antiguo será necesario contar con la información del haplogrupo al que pertenece la muestra, caracterizado con enzimas de restricción, para evitar confusiones en la interpretación de los resultados.

En general, nuestros análisis indican que recuperar el fragmento de 302 pb puede aportar virtualmente la misma información que el fragmento de 360pb. Para DNA antiguo, esto puede realizarse diseñando *primers* para fragmentos solapados de alrededor de 180pb cada uno. En caso de que esto no sea posible, recuperar el fragmento de 153pb y caracterizar las muestras mediante enzimas de restricción puede brindar información suficiente.

Árboles filogenéticos vs. redes medias

El conocimiento previo del haplogrupo al que pertenecen las secuencias analizadas en la reconstrucción filogenética intrapoblacional, representa un método objetivo de evaluación de los métodos de reconstrucción filogenética. Con los datos de las poblaciones de las que se conocía la caracterización de los haplogrupos y la secuencia en las mismas

muestras, se realizó la reconstrucción filogenética con dos métodos. Por un lado, la estimación de una distancia genética y la construcción de un árbol con el método de Neighbor-Joining; y por otro, la construcción de una red media reducida. En general, los resultados no mostraron diferencias importantes entre los dos métodos, sin embargo, en algunos casos el árbol filogenético reflejaba mejor las relaciones de las secuencias, de acuerdo con su haplogrupo, en comparación con la red media correspondiente. Particularmente, esto fue evidente con algunas muestras del subgrupo U5 de la población Catalana (CAT) y de los subgrupos U5 y T1a1 de la población Toscana. Así mismo, el método de las redes medias se mostró más sensible a la reducción del fragmento, como lo muestra la pérdida de definición de los haplogrupos W, X e I en la red de 153pb de la población Toscana. Por el contrario, el resultado fue ligeramente mejor con el método de las redes medias únicamente al analizar el fragmento de 153pb de las muestras de Galura, ya que las muestras del subgrupo U5 resultaron mejor diferenciadas respecto a lo obtenido con el árbol filogenético.

Por estos motivos consideramos que el método del árbol filogenético resulta por lo menos tan adecuado como el método de las redes medias reducidas en el análisis de las secuencias del mtDNA. Sin embargo, es importante destacar que la elección de la distancia genética puede resultar decisiva en la reconstrucción del árbol filogenético. En nuestro análisis, utilizamos la distancia de Tamura y Nei (1993), que está basada en el modelo evolutivo más complejo de los que hasta el momento se han desarrollado, al punto de que otras distancias pueden considerarse casos especiales o simplificados del modelo de Tamura-Nei (ver Weiss y von Haeseler, 1998). Además hemos utilizado la corrección de la distribución gamma de la tasa de mutación, ya que diversos estudios indican que en el D-loop existen sitios con diferentes tasas de mutación. La distribución gamma está definida por el parámetro a que es el inverso del coeficiente de variación. La elección de este parámetro es de fundamental importancia, sin embargo, diversos autores han obtenido diferentes valores para la región de control debido a que han analizado segmentos distintos. Wakeley (1993), estimó para un fragmento de 250pb (16130-16379) un valor de $a=0.47$, en tanto que Tamura y Nei (1993) obtuvieron un valor de $a=0.11$ al analizar conjuntamente los dos segmentos hipervariables de la región de control. En nuestro caso

hemos utilizado el valor obtenido por Wakeley por tratarse del valor estimado para un fragmento que se encuentra en la misma zona que los segmentos aquí analizados.

Aunque hemos utilizado el mismo valor para los tres fragmentos, nuestros estudios preliminares indican que este valor cambia substancialmente para cada uno de los segmentos, lo que está en concordancia con lo observado en los resultados de los índices de diversidad nucleotídica que indicaban tasas de mutación diferentes para cada segmento. Los valores del parámetro a de la distribución gamma estimados para fragmentos de distintos tamaños a partir de 7 muestras de poblaciones europeas se presentan en la Tabla D-3. Este parámetro debe ser calculado a partir de muestras de todo el mundo, para evitar que la estimación esté influida por los procesos demográficos (Weiss y von Haeseler, 1998), por este motivo consideramos nuestro análisis como preliminar. Es importante la confirmación de estos valores, con el fin de poder utilizar el valor adecuado para cada fragmento, consiguiendo así que la reconstrucción filogenética sea más precisa.

	Fragmento (pb)		
	360	302	153
a	0.234	0.31	0.86

Tabla D-3 Promedio del parámetro a de la distribución gamma estimado a partir de 7 poblaciones europeas para fragmentos de distintos tamaños.

Por otra parte, una ventaja que puede tener el método de las redes medias, es el hecho de que permite incluir la información de la frecuencia de los haplotipos para presentarla de forma gráfica a través del tamaño de los círculos que definen los nodos en la red. Sin embargo, como hemos hecho en el análisis interpoblacional esta información también puede ser incluida en los árboles filogenéticos, aunque el resultado no sea tan evidente visualmente.

Correlación secuencia-haplogrupo

La correlación secuencia-haplogrupo no sólo resultó fundamental en la autenticación del mtDNA antiguo, sino también se ha revelado muy útil en el análisis filogenético. Contar con la información del haplogrupo ayuda en la interpretación de los resultados de la reconstrucción tanto intra como interpoblacional. En el análisis intrapoblacional esta información ayuda en la interpretación de los árboles filogenéticos y en la elección de las rutas más plausibles en las redes medias, pues permite identificar las secuencias mal clasificadas. En el análisis interpoblacional esta correlación facilita el estudio de las relaciones entre los haplotipos de distintas poblaciones al permitir separar las secuencias de acuerdo al haplogrupo al que pertenecen, evitando que se produzca el “ruido de fondo” y permitiendo recuperar la información contenida en la estructura genética de cada uno de los haplogrupos. Esta aproximación puede ayudar a dilucidar distintos procesos ocurridos en la historia de las poblaciones, sobretodo si los distintos haplogrupos se han visto involucrados en procesos demográficos diferentes.

Las relaciones filogenéticas de las secuencias clasificadas por haplogrupos fueron examinadas construyendo un árbol de Neighbor-Joining a partir de la distancia de Tamura-Nei con distribución gamma. En este caso, se utilizó el parámetro a de la distribución gamma estimado para cada uno de los haplogrupos examinados, ya que a diferencia de la estimación a nivel poblacional, la estimación de este parámetro a nivel de haplogrupo no puede hacerse con muestras de la población mundial, pues los haplogrupos son específicos de cada continente. Para el cálculo de este parámetro se utilizaron todas las secuencias de cada haplogrupo encontradas en las 10 poblaciones analizadas. Los valores del parámetro a resultaron muy diversos (Tabla D-4) y estas diferencias constituyen una evidencia indirecta de que los haplogrupos tienen historias diferentes, ya sea relacionadas con procesos migratorios o con procesos demográficos diferentes sufridos por las poblaciones europeas.

Esto nos demuestra que en el acervo mitocondrial de las poblaciones existe mucha información y es necesario encontrar los métodos para poder descifrarla adecuadamente. Algunos autores (Simoni et al. 2000b) han propuesto la eliminación, del banco de datos, de la información que origina el ruido de fondo en los estudios de mtDNA. No obstante, una alternativa a la eliminación de caracteres, puede ser la correcta separación y clasificación de las secuencias. En este sentido, analizar las secuencias por haplogrupos puede resultar un método interesante, pues a la vez que se reduce el ruido de fondo, se pueden obtener huellas de diferentes procesos demográficos. Es muy importante, para que la clasificación de las secuencias sea fiable, que la caracterización de los haplogrupos se realice con enzimas de restricción en el mayor número posible de muestras, obviamente de las que se cuente con la secuencia de la región de control, para poder establecer los límites de confianza de la correlación entre secuencia y haplogrupo, como ya se ha expuesto en secciones anteriores.

Hap.	N	<i>a</i>
H	251	0.53
U	94	0.21
T	49	0.10
J	53	0.42
K	29	0.2
I,W,X	37	0.2

Tabla D-4 Parámetro *a* de la distribución gamma estimado con un método de máxima verosimilitud con el programa PUZZLE, para cada haplogrupo. N, número total de secuencias utilizadas. Del haplogrupo V no se obtuvo un resultado fiable debido a la escasa variabilidad de sus secuencias.

Análisis filogenético

A pesar de que este trabajo es de carácter marcadamente metodológico, los análisis llevados a cabo han dado resultados que pueden interpretarse en el contexto de la historia de las poblaciones europeas. Tanto en el nivel particular de la población Catalana como en lo relativo a la estructura mitocondrial en Europa.

El acervo mitocondrial de las poblaciones europeas

En Europa, las poblaciones son bastante homogéneas en cuanto a su mtDNA, debido a su origen común. No obstante, es posible observar que existen determinadas poblaciones que probablemente debido a sus procesos demográficos resultan un poco más diferenciadas del resto.

En cuanto a la proporción de haplogrupos, la prueba exacta indica que las poblaciones de Barbagia, Galicia y País Vasco han sufrido esta diferenciación; resultado que fue corroborado en cierta medida mediante las reconstrucciones filogenéticas realizadas a partir de la frecuencia de los haplogrupos.

Por su parte, la reconstrucción filogenética basada en las secuencias, también mostró cierta estructura en las poblaciones analizadas, destacando la diferenciación de las poblaciones de Galicia, País Vasco, Cerdeña, Cataluña y notoriamente la de Galura. Estos resultados están en contradicción con el estudio de Simoni y col. (2000a) basado en métodos de autocorrelación, en el que no fue encontrada una estructuración significativa en el mtDNA de las poblaciones europeas. Sin embargo, el estudio que llevamos a cabo sobre las relaciones de los haplotipos de las distintas poblaciones, mostró que existen relativamente pocos haplotipos mitocondriales compartidos entre las poblaciones

(35/254), lo que indica que en algún momento ha debido de haber una diferenciación *in situ* de las poblaciones. Esta aparente contradicción podría ser explicada si la diferenciación de las poblaciones se produjo en tiempos relativamente recientes, ya que en el estudio de Simoni y col. (2000a) se utilizaron únicamente 20 sitios variables, elegidos por ser los más informativos y presumiblemente podrían ser los que contengan la información más antigua. En cualquier caso, este es un tema bastante polémico (Torróni et al. 2000; Simoni et al. 2000b) y nuestro análisis es limitado como para poder obtener conclusiones definitivas.

Los distintos índices de diversidad, estimados tanto a partir de los haplogrupos como de las secuencias de la región de control, mostraron un gradiente compatible con una expansión este - oeste descrita por otros autores. En un extremo se situaría la población de Turquía y en el otro la población de Galicia. El análisis de la distribución de las diferencias por parejas también es compatible con esta hipótesis. La única excepción, la población actual de Cataluña, parece ser debida a que ésta haya atravesado procesos demográficos diferentes a los sufridos por el resto de poblaciones europeas (ver abajo).

Las características de la población de Argelia, son compatibles con la hipótesis de que esta población representa el extremo occidental de una expansión equivalente a la ocurrida en el sur de Europa, pero que seguiría la ruta de la costa sur del mediterráneo.

La información contenida en las secuencias de los distintos haplogrupos

Expansión poblacional Mesolítica

De acuerdo con algunos investigadores, la forma de la filogenia de algunos haplogrupos, así como su distribución, indican que después del último pleniglacial hubo una expansión poblacional hacia el norte de Europa desde las zonas de refugio situadas en

el sur del continente. Particularmente, las filogenias en forma de estrella de los haplogrupos H, V y K podrían estar relacionadas con esta expansión (Richards et al. 1998). No obstante, esta expansión debería de haber afectado también al resto de haplogrupos presentes en la época, produciendo ramificaciones en forma de estrella en sus filogenias, ya que cuando una población se expande, también se expanden todos sus alelos (Excoffier y Schneider, 1999).

En nuestro análisis, los haplogrupos H, V y K muestran una filogenia en forma de estrella y el haplogrupo U también muestra signos de contener en el centro una estructura de este tipo, pero es probable que las secuencias que deberían conformarla hayan sido incluidas en la filogenia del haplogrupo H. Sin embargo, el árbol del haplogrupo T, no presenta ningún indicio de contener estructuras en forma de estrella. Como ya lo habíamos expresado, este hecho nos parece sorprendente, ya que de acuerdo con Richards y col. (1998) este haplogrupo estaba presente en Europa antes de la supuesta expansión post-glacial y por lo tanto debería de haber sido afectado por ella, generando en su filogenia, al igual que para la mayoría de haplogrupos, ramificaciones en forma de estrella, a pesar de que pudiese contener estructuras más profundas, como se observa en los haplogrupos H ó U.

Conclusiones similares pueden obtenerse a partir del árbol de los haplogrupos I, X y W, que tampoco muestran ramificaciones en forma de estrella. Además, el haplogrupo J, que de acuerdo con Richards y col. (1996) entró en Europa durante el neolítico y que por lo tanto no participó de la expansión post-glacial, presenta una estructura peculiar, indicando cierta antigüedad de algunos tipos mitocondriales que estarían en la base de los distintos subgrupos pero también presenta una forma de estrella en el centro de la filogenia, indicando una expansión reciente.

Estas observaciones nos llevan a considerar hipótesis alternativas que expliquen las filogenias en forma de estrella que presentan algunos haplogrupos. En este sentido debemos preguntarnos si un aumento en la frecuencia de un haplogrupo podría tener los mismos efectos que una expansión poblacional en cuanto a la variabilidad producida. Es

bastante probable que esto sea así, ya que un aumento en la frecuencia implica el aumento del número de individuos pertenecientes al mismo haplogrupo y con una tasa rápida de mutación, muchos de estos individuos presentarán nuevas mutaciones. Si esta hipótesis es cierta, es probable que nos encontramos ante una expansión del haplogrupo H y sería muy interesante intentar modelar el proceso y predecir el tiempo aproximado que tardará este haplogrupo en substituir a los haplogrupos restantes. Esta hipótesis puede ser corroborada estudiando la frecuencia del haplogrupo H en poblaciones antiguas, preferiblemente preneolíticas. De hecho, Izagirre y De la Rúa (1999) encontraron en poblaciones prehistóricas del País Vasco una frecuencia relativamente baja del haplogrupo H, en comparación con la frecuencia determinada para poblaciones Vascas actuales.

Por otra parte, bajo esta hipótesis, los haplogrupos K y V, también habrían aumentado su frecuencia, en tanto que los haplogrupos T, I, W y X, serían los haplogrupos cuya frecuencia habría disminuido. Por su parte, los haplogrupos U y J podrían haber mantenido una frecuencia constante desde su llegada a Europa.

Antigüedad de los haplogrupos en Europa

En cuanto a la fecha en la que entraron los distintos haplogrupos, nuestros resultados coinciden parcialmente con lo observado por Richards y col. (1998; ver Marco de Referencia). La filogenia que obtuvimos a partir de las secuencias del haplogrupo U es bastante peculiar, pues es la única que muestra una diferenciación más o menos clara entre algunas poblaciones. Por ejemplo, el subgrupo U6 lo conforman mayoritariamente haplotipos de la población de Argelia; el clúster del subgrupo U1 está formado exclusivamente por secuencias de Cerdeña, en tanto que en el grupo de las secuencias U3, únicamente hay secuencias de Turquía y Argelia. En ningún otro de los árboles obtenidos observamos grupos formados por secuencias exclusiva o mayoritariamente de una sola población ni tampoco procedentes de dos poblaciones como en el caso del subgrupo U3. Estas características indican que este haplogrupo ha estado presente en Europa desde

tiempos muy antiguos. El árbol del haplogrupo T también muestra una estructura compatible con una antigüedad considerable, aunque menor que la del haplogrupo U, ya que no existen grupos de secuencias de poblaciones concretas. Una conclusión similar puede extraerse del árbol de los haplogrupos I, W y X.

El haplogrupo J también muestra una estructura compatible con una importante antigüedad, ya que presenta grupos bien diferenciados con ramas profundas. No obstante, según el estudio de Richards y col. (1996) este haplogrupo habría sido introducido en Europa en tiempos más recientes, durante el Neolítico. En su análisis estos investigadores también notaron la antigüedad de la filogenia, sin embargo, postularon la existencia de distintos ancestros cuya variabilidad ya era importante al entrar en Europa, lo que resolvería la contradicción. En nuestra opinión, este resultado debe ser corroborado, idealmente con estudios de DNA antiguo.

Las filogenias construidas con las secuencias de los haplogrupos K y V, son compatibles con la idea de que surgieron localmente en Europa en tiempos relativamente recientes. Pero de más difícil interpretación es la filogenia del haplogrupo H. A primera vista esta filogenia tiene una forma marcadamente estrellada, no obstante, es posible observar una estructura con grupos bien definidos formados por ramas profundas evidenciando una antigüedad considerable. En nuestra opinión este haplogrupo podría ser incluso más antiguo que el haplogrupo T, pero ya sea por la expansión poblacional post-glacial, o por el simple aumento de su frecuencia, presenta una gran cantidad de nuevos haplotipos que conforman una filogenia en forma de estrella superpuesta a la filogenia de mayor antigüedad. Estas características también podrían ser explicadas por al menos dos olas migratorias desde el Este Medio, ocurridas de tal forma que el haplogrupo H fuese el único en ser reintroducido, ya que el resto de haplogrupos no muestran características parecidas. Esta idea no contradice la hipótesis de que las características peculiares de este haplogrupo se deban a que en algún momento de la historia comenzó a expandirse cuando su frecuencia superó cierto umbral, ya que una segunda introducción pudo elevar la frecuencia del haplogrupo. De otra forma, en el caso de que haya sido introducido únicamente por un evento migratorio, este umbral ha debido ser superado en la población

que constituyó la base común del haplogrupo, la población que lo introdujo en todas las poblaciones europeas.

Así, el haplogrupo U habría sido el primero en ser introducido en Europa, el haplogrupo H también pudo haber sido introducido al mismo tiempo o pudo haber entrado posteriormente junto con los haplogrupos T, I, W y X o incluso pudo ser introducido en dos eventos migratorios, uno relacionado con el haplogrupo U y el otro con el resto de haplogrupos mencionados. La entrada del haplogrupo J no está muy clara, pero por la filogenia que presenta pudo haber sido introducido antes del neolítico junto con la mayoría de haplogrupos. Finalmente, los haplogrupos K y V habrían surgido localmente en Europa, a partir de los haplogrupos U y H respectivamente.

No obstante, todas estas observaciones deben ser consideradas como provisionales, ya que los cambios en las frecuencias de los haplogrupos podrían “borrar” algunos aspectos importantes de las filogenias. El efecto del cambio en las frecuencias podría sesgar también el cálculo de la edad del antecesor común de cada haplogrupo. Resulta entonces fundamental, determinar si ha habido cambios notables en la frecuencia de los haplogrupos, lo que puede realizarse de manera fiable a través de estudios de DNA antiguo en poblaciones pre-neolíticas.

Diferenciación de la población Catalana

El análisis de la distribución de las diferencias por parejas de individuos, mostró ciertas peculiaridades en la historia demográfica de Cataluña, que la hacen diferente del resto de poblaciones europeas. El análisis de máxima verosimilitud permitió detallar más las causas de estas diferencias, mostrando que pueden estar relacionadas con un mayor incremento de la población durante la expansión que originó la curva. Aunque esta expansión pudo producirse en el Paleolítico, los procesos demográficos ocurridos durante y después de la última glaciación en el Mesolítico, podrían haber influido en la estructura

del mtDNA de la población Catalana. Sobretudo considerando que los modelos utilizados no contemplan el efecto de la migración.

También se debe considerar que nuestra muestra de la población Catalana no está bien definida genealógicamente, y este hecho podría estar sesgando el resultado y ser responsable de las diferencias observadas en esta población. Sin embargo, otras muestras que tampoco tienen una filiación filogenética bien definida, como la británica de Piercy y col. (1993) o la de Galicia de Salas y col. (1998), no presentan características tan peculiares. Además, otros estudios también han encontrado cierta diferenciación en la población Catalana (Côrte-Real et al. 1996, Simoni et al. 2000a). Todo esto justifica la necesidad de llevar a cabo un estudio más detallado del mtDNA de esta población, en el que sería deseable un muestreo en el que sean registrados los lugares de nacimiento de las abuelas y en el que las muestras sean colectadas en pueblos pequeños del interior de Cataluña. También es importante ampliar el número de secuencias obtenidas, quizá hasta un mínimo de 45 individuos.

Diferencias entre las poblaciones antigua y actual de Cataluña

En general no existe diferencia entre las poblaciones antigua y actual de Cataluña en cuanto a su distribución de haplogrupos según la prueba exacta de diferenciación poblacional, a pesar del bajo número de muestras de la población antigua y su relativa escasa representación del haplogrupo H. Asimismo, la distancia genética de Reynolds estimada a partir de la frecuencia de los haplogrupos indica que las poblaciones actual y antigua muestran un patrón de distancias genéticas similar, ya que las dos presentan las menores distancias con las mismas poblaciones. No obstante también es posible observar ciertas diferencias, ya que la población antigua presenta una menor distancia con la población actual, en tanto que la población actual presenta distancias menores con otras poblaciones antes que con la población antigua.

Por otra parte, entre las poblaciones antigua y actual no existe diferencia significativa en cuanto a la proporción de haplotipos diferentes ($\%k$) que presentan ni en cuanto a la proporción de sitios variables ($\%v$).

No obstante, en la distribución de las diferencias por parejas, se observó una diferencia importante en la situación de la curva, que para la población antigua se sitúa más a la izquierda indicando un tiempo de expansión más reciente. Sin embargo, con el fragmento estudiado (153pb) se pierde información relevante para el análisis de la distribución, como lo demuestra el hecho de que para este segmento el ajuste a la distribución de Poisson sea mejor que el ajuste al modelo de expansión súbita, contrariamente a lo observado para los segmentos grandes. Además, cuando se realizó el cálculo del tiempo de expansión utilizando una tasa de mutación en función del valor de \hat{U} se encontraron discrepancias con lo observado gráficamente. De acuerdo con la posición de las curvas, la población de la Plaça Vella se habría expandido después de que lo hicieran las poblaciones de Galicia y País Vasco, pero la estimación con la tasa de mutación mostró que estas últimas se expandieron con posterioridad. Por lo tanto, la diferencia en la distribución de las diferencias por parejas entre las poblaciones antigua y actual no puede ser evaluada.

Finalmente, el árbol filogenético construido a partir de las secuencias de la región de control, mostró la relación cercana de las poblaciones antigua y actual de Cataluña. Este resultado contrasta con lo obtenido en la reconstrucción filogenética a partir de los haplogrupos, sin embargo, el análisis a través de las secuencias es mucho más completo pues considera el verdadero contenido genético de las poblaciones. Por ejemplo, la frecuencia de un determinado haplogrupo podría ser igual para dos poblaciones y sin embargo, los haplotipos de dicho haplogrupo pertenecientes a cada población podrían ser filogenéticamente distantes de los haplotipos de la otra población; por el contrario, la frecuencia del haplogrupo podría ser muy distinta y sin embargo, los tipos mitocondriales estar cercanamente relacionados. Es por esto que el análisis a partir de las secuencias se debe considerar más completo y más adecuado y su resultado debe primar sobre el resultado obtenido mediante la proporción de haplogrupos.

De hecho, la proporción de haplotipos compartidos entre las muestras antigua y actual (9/23, 39.1%) es relativamente alto y significativamente diferente (prueba exacta de Fisher bilateral, $P=0.0042$) de la proporción de haplotipos compartidos observada para todo el conjunto de poblaciones (35/254, 13.8%).

Por otra parte, es muy poco probable que las pequeñas diferencias entre las poblaciones antigua y actual de Cataluña pueden deberse a divergencia evolutiva, ya que la distancia genética entre las dos poblaciones, estimada a partir de las secuencias, es cero; mientras que la diferencia observada en la distribución de haplogrupos puede deberse a factores como el reducido tamaño de muestra o la posibilidad de consanguinidad en la población antigua. En cierta medida, esto nos indica que la tasa de mutación del mtDNA no es lo suficientemente rápida como para generar cambios importantes en 500 años.

En todo caso, es fundamental caracterizar la población actual de forma más precisa, ya que de haber encontrado diferencias importantes, éstas podrían atribuirse a migraciones recientes. De igual forma, también es interesante el estudio de poblaciones Catalanas más antiguas, así como la ampliación de la muestra de la población de la Plaça Vella que podría convertirse en una población de referencia para el estudio del mtDNA de la población Catalana.

Relación de la población Catalana con otras poblaciones

Los análisis a partir de la distribución de haplogrupos, mostraron que la población Catalana difería de las poblaciones ibéricas de Galicia y el País Vasco pero no de las poblaciones de las islas mediterráneas de Córcega y Cerdeña. Asimismo, la reconstrucción filogenética a partir de las secuencias mostró que la población catalana se sitúa en estrecha relación con la población Galuresa y se encuentra en un punto más cercano a las poblaciones de Cerdeña y de la Toscana que con las poblaciones ibéricas. Más aún, el

análisis interpoblacional de las relaciones haplotípicas, mostró la estrecha relación que existe entre las secuencias de los haplogrupos U, J y T de la población Catalana con las secuencias correspondientes de otras poblaciones mediterráneas. Estas características parecen ser la consecuencia tanto de la influencia de poblaciones mediterráneas en la población Catalana, como de la influencia ejercida por la población Catalana sobre las islas mediterráneas en tiempos históricos más recientes. Por ejemplo, la secuencia CAT34-U5a se sitúa en la misma rama que la secuencia G16-U5a de la población de Galura y la secuencia CAT23-J2a, es idéntica a una secuencia de la población Sarda y se encuentra situada en una rama que conduce a dos secuencias de la población Galuresa. Estos linajes mitocondriales podrían haber sido introducidos a la isla de la Cerdeña por la población Catalana cuando ésta expandió su dominio en la zona mediterránea. Por otra parte, la secuencia CAT24-T2 se sitúa en un grupo formado por dos secuencias de la población Toscana, pero también es compartido por la población Turca, por lo que este linaje podría representar un aporte desde el Medio Este que hubiese seguido la ruta mediterránea.