

Origen de la distribución geográfica del haplogrupo R-M269

Origin of the geographical distribution of the R-M269 haplogroup

Yelko Bejarano, Jose A. Peña

Dpto. Genética, Antropología Física y Fisiología Animal, Fac. Ciencia y Tecnología, Universidad del País Vasco, UPV/EHU. ybejarano001@ikasle.ehu.eus, joseangel.pena@ehu.eus

Palabras clave: R-M269, Cromosoma Y, Haplogrupo, Clina

Keywords: R-M269, Y Chromosome, Haplogroup, Cline

Resumen

La distribución geográfica actual del haplogrupo del cromosoma Y más abundante en Europa, R1b-M269, se organiza de forma creciente de este a oeste, hecho que contrasta con su probable origen en la cultura Yamnaya de la estepa póntica, donde actualmente no presenta frecuencias. Se han analizado las frecuencias de R1b (y particularmente R1b-M269) y su distribución a lo largo del tiempo en el continente europeo, valorando los posibles eventos que han podido alterarla, incluyendo procesos migratorios, de deriva, expansiones culturales, invasiones, etc.

Se ha confirmado el origen del haplogrupo R1b-M269 en el oeste de la actual Rusia, en el contexto probablemente de la cultura Yamnaya, su expansión por Europa desde hace unos 4.500 años y su desaparición en el noreste del continente con el declive de dicha cultura. Por efecto probablemente de la deriva, combinada con la expansión sucesiva de las culturas Yamnaya, de la cerámica cordada y del vaso campaniforme, se llegaron a alcanzar frecuencias de 1 en el extremo occidental del continente. Posteriormente, la expansión del Imperio Romano, y las diferentes invasiones entre hace 2.000 y 1.000 años, diluirían las frecuencias para dar lugar a las observadas en la actualidad.

Abstract

The current geographical distribution of the most abundant Y-chromosome haplogroup in Europe, R1b-M269, is currently arranged in an increasing pattern from east to west, which contrasts with its probable origin in the Yamnaya culture in the Pontic Steppe, where it currently lacks frequencies. Frequencies of R1b (and particularly R1b-M269) have been analyzed, along with their distribution over time in the European continent, evaluating potential events that may have altered it, including migratory processes, drift, cultural expansions, invasions, etc.

The origin of haplogroup R1b-M269 has been confirmed in the western part of present-day Russia, likely within the context of the Yamnaya culture. It expanded across Europe around 4,500 years ago and disappeared in the northeastern part of the continent with the decline of that culture. Likely due to drift, combined with the successive

expansion of the Yamnaya, Corded Ware, and Bell Beaker cultures, frequencies as high as 1 were reached in the westernmost part of the continent. Subsequently, the expansion of the Roman Empire, and various invasions between 2,000 and 1,000 years ago, would dilute the frequencies, resulting in the current distribution.

Introducción

El cromosoma Y ha sido objeto de numerosos estudios sobre la evolución de las poblaciones humanas, particularmente acerca de sus procesos de microdiferenciación, que incluyen la interacción entre flujo génico y deriva. Por un lado, aporta una valiosa información sobre los linajes masculinos, complementando la información que para los linajes femeninos aporta el ADN mitocondrial. Por otra parte, dado que en su mayor parte este cromosoma no experimenta recombinación, pueden identificarse haplotipos y haplogrupos, que son herramientas muy adecuadas para la identificación de los procesos migratorios.

Esto ha permitido rastrear los sucesores de linajes paternos de poblaciones del pasado, como es el caso de los Yamnaya, componentes de una cultura que se remonta hasta hace 6.000 años (Morgunova y Khokhlova, 2013). Precisamente, se ha propuesto a esta civilización de la estepa pónica como el origen del haplogrupo más común en la actualidad en Europa occidental, el R1b-M269 (Balaesque *et al.*, 2010).

Sin embargo, esta cultura tuvo su origen en el noreste del continente y es en esta región donde menor frecuencia presenta el haplogrupo R1b-M269 en la actualidad. De hecho, se distribuye según una clina longitudinal que aumenta de este a oeste de forma excepcional, ya que la mayor parte de las clinas observadas en los europeos actuales son latitudinales (Valero Almingol y Peña, 2021).

Por ello, en este trabajo se pretenden establecer las causas que han podido dar lugar a la distribución geográfica actual de las frecuencias del haplogrupo R1b-M269.

La hipótesis que se propone es que el haplogrupo tuvo su origen en los Yamnaya y se expandió con ellos y su distribución actual se debe a diferentes procesos de microdiferenciación (flujo génico y deriva genética), que se pretenden identificar.

Materiales y métodos

Para la realización de este proyecto, se utilizó la base de datos AADR (Mallick y Reich, 2023), que incluye los genotipos de todas las muestras de ADN antiguo analizadas por los autores. Los datos se encuentran en formato Eigensoft, organizándose en 3 ficheros, uno con información sobre los individuos (.ind), otro con información sobre los SNPs (.snp) y un tercero con los genotipos (.geno). La base de datos, en formato binario, se descomprimió a formato texto mediante Eigensoft (Patterson *et al.*, 2013).

Se buscaron en esta base de datos los Y-SNPs que presentan una variación clinal idéntica en las poblaciones actuales del continente europeo (Valero Almingol y Peña, 2021), encontrándose 9 de ellos. Mediante la página de referencia del ISOGG (International Society of Genetic Genealogy) v.15.56 (<https://isogg.org/tree/index.html>) se ha asignado el haplogrupo R1b de cada individuo definido por estos SNPs, de acuerdo al criterio que se muestra en la Tabla 1.

SNP	Haplogrupo
rs67462087	R1b1a1b (M269)
rs66715481	R1b1a1 (P297)
rs7067226	R1b1a1b (M269)
rs9786169	R1b1a1 (P297)
rs9786276	R1b1a1b (M269)
rs877756	R1b1a1b (M269)
rs67558859	R1b1a1b (M269)
rs9786671	R1b1a1b (M269)
rs67494109	R1b1 (L278)

Tabla 1. SNPs utilizados y haplogrupo que identifican.
Table 1. SNPs used and haplogroup identified.

Se identificó el genotipo de cada individuo mediante el programa GWEdit (Peña, 2023) y se asignó el haplogrupo correspondiente. En general, se describirán las frecuencias de R1b, aunque puntualmente se mencionarán los resultados referentes al subtipo R1b-M269. Posteriormente, se agruparon los individuos en una serie de regiones geográficas. Dado que las muestras de ADN antiguo a menudo no dan resultados para la mayor parte de los SNPs analizados, el número de individuos con algún resultado no fue muy amplio, por lo que se definió un número limitado de regiones geográficas dentro del territorio de interés utilizando 2.853 de los 16.389 individuos totales que incluye la base de datos. Puesto que el objetivo ha sido analizar el continente europeo, se han descrito los resultados para este entorno geográfico, pero se han incluido algunas regiones asiáticas próximas en los mapas sintéticos, con el fin de mejorar su resolución. También se distribuyeron los individuos analizados por fecha de antigüedad en un número limitado de cohortes de 1.000 años, excepto el período desde hace 4.999 hasta hace 4.000 años, que fue dividido en 2, por su especial interés.

Con el fin de representar de una forma visualmente asequible las frecuencias del haplogrupo R1b, se elaboraron una serie de mapas sintéticos por períodos mediante el programa Golden Software Surfer v16.

Finalmente, con el fin de poder interpretar conjuntamente todos los valores de las diferentes regiones en los diferentes períodos, se realizó un Análisis de Componentes Principales (ACP) mediante el programa PAST (Hammer *et al.*, 2001). Se utilizó la matriz de varianzas y covarianzas y con el fin de soslayar la existencia de valores perdidos, se realizó una imputación iterativa, donde los valores perdidos se sustituyen inicialmente por la media de su columna y se realiza un ACP previo para calcular los valores por regresión de los datos que faltan (Ilin y Raiko, 2010). Previamente se realizó un test de normalidad de Shapiro Wilk sobre los datos, también con PAST.

Resultados

Se han analizado las frecuencias de los haplotipos Y-SNPs en tres pasos. En primer lugar, se han considerado varias regiones geográficas, observando las variaciones temporales en las frecuencias de cada una de ellas. En segundo lugar, se han elaborado una serie de mapas sintéticos, en los que se han observado las variaciones geográficas en cada período. Y finalmente, se ha analizado la variación espacial y temporal conjuntamente mediante un Análisis de Componentes Principales.

Variación temporal de las frecuencias

Se han realizado una serie de gráficos lineales que representan el cambio de la frecuencia del haplogrupo R1b a lo largo del tiempo en cada zona geográfica previamente definida. Para ello se han considerado una serie de períodos de 1.000 años que incluyen la mayor parte de los individuos analizados, desde hace 7.999 años hasta la actualidad. Por su especial interés se dividió el período 4.999-4.000 en dos (4.999-4.500 y 4.499-4.000).

No obstante, es interesante mencionar por su posible trascendencia a varios individuos con una antigüedad superior a 8.000 años que han mostrado resultados positivos para algunos de los SNPs analizados. Por su antigüedad, todos ellos pertenecen al Mesolítico, por lo que son anteriores a la entrada en Europa de los primeros agricultores y lógicamente, a la expansión indoeuropea. Así, el resto de Villabruna (Italia), con 14.000 años de antigüedad porta el haplogrupo R1b, cualidad que también mencionan Fu *et al.* (2016). Por el genotipo que muestra, podría estar emparentado o incluso ser antecesor de R1b-M269. Con resultados muy similares han aparecido en Serbia 2 individuos de la cultura Iron Gates de 8.700 y 9.500 años y un individuo del yacimiento de Vlasac de 10.200 años, así como 3 individuos de la cultura Iron Gates de Rumanía (entre 8.600 y 9.400 años), 1 individuo de la misma cultura de Dereivka, en Ucrania (8.100 años) y un individuo en Zvejnieki (Letonia) con 9.200 años (Mathieson *et al.*, 2018). En todo caso, no se ha identificado ningún individuo con el haplogrupo R1b-M269. En Asia tan sólo se han encontrado 19 individuos con más de 8.000 años, ninguno de los cuales ha mostrado resultados positivos para los SNPs analizados.

En los Países Bálticos se observan unas frecuencias para R1b de 0'33 en periodos muy tempranos, que descienden posteriormente hasta desaparecer. Esta zona geográfica es una de las pocas regiones que presentan frecuencia, aunque baja, en el periodo de hace 7.000-7.999 años, que mantendrá también en el milenio posterior. Si bien es cierto que hay varios periodos sin datos hasta

hace 2.999-1.000 años, en este período presenta un valor de 0 (Figura 1). En todo caso, los haplogrupos encontrados son R1b, pero no R1b-M269.

En el oeste de Rusia comienzan a aparecer valores para R1b a partir de hace 6.999-6.000 años con una frecuencia relativamente baja, que sube drásticamente hace 4.999-4.500 años, con un valor máximo ligeramente superior a 0,7, para luego descender hasta frecuencias prácticamente nulas. A partir de hace 1.999 años se estabiliza en valores bajos de alrededor de 0,25, hasta la actualidad (Figura 1). R1b-M269 aparece en el último siglo del período 5.999-5.000, siendo a partir de ese momento mayoritario entre los R1b.

La región integrada por Ucrania, Rumanía y Moldavia forma parte del pequeño conjunto de zonas geográficas que presentan frecuencias de R1b en el periodo más antiguo (7.999-7.000). Los escasos valores obtenidos toman valores por debajo de 0,2 a lo largo del tiempo excepto hace 2.999-2.000 años, cuando toma un valor máximo de 0,5 (Figura 1). Durante el período 4.999-4.000 el número de muestras disponibles es escasa para ser representativa, por lo que no aparece en la gráfica, pero es precisamente hace 4700 años cuando aparecen los primeros individuos con el haplogrupo R1b-M269.

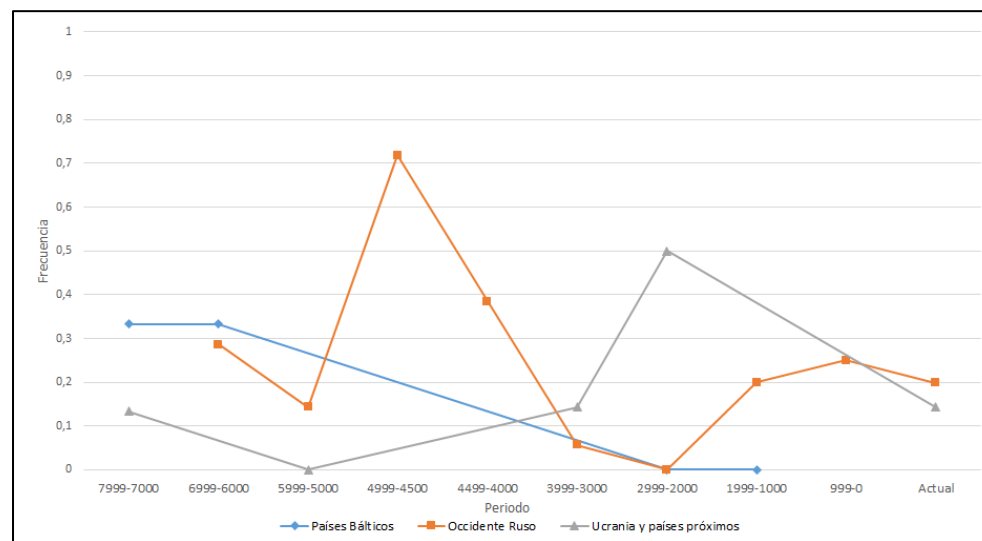


Figura 1. Variación temporal de la frecuencia de R1b en Europa Oriental.
Figure 1. Temporal variation of R1b frequency in Eastern Europe.

En Europa Central el valor de las frecuencias de R1b, que es muy bajo en edades tempranas, experimenta un gran aumento hace alrededor de 4.499-4.000 años para dar lugar a valores máximos de entre 0,7 y 0,85, en el caso de las dos zonas más al norte y de 0,55 en la región más hacia la zona balcánica, al sur. Estos valores máximos van descendiendo progresivamente hasta la actualidad, adquiriendo valores cercanos a 0,2 y 0,4 para la zona oriental y occidental respectivamente. Este descenso se acentúa hacia el sur, en la zona de Europa Central-Balcánica, donde el valor actual obtenido se encuentra ligeramente por debajo de 0,1 (Fig. 2). Los registros más antiguos de R1b-M269 en esta región son de hace unos 4800 años en la República Checa, 4700 años Alemania y Croacia y 4600 en Suiza. Por otra parte, como en el resto de regiones europeas, el aumento de las frecuencias del haplogrupo R1b se debe en exclusiva a una mayor presencia de R1b-M269.

Las Islas Británicas y Francia presentan un comportamiento muy similar para R1b. Las frecuencias son prácticamente nulas hasta que empiezan a aumentar a partir de hace 4.999-4.500 años, presentan un aumento drástico en la siguiente mitad del milenio alcanzando cifras superiores a 0,8 y llegando incluso a 0,9 y que luego se va a mantener con un ligero descenso. Excepcionalmente hay una considerable bajada de las frecuencias en ambas zonas geográficas hace 1.000-1.999 años. En Inglaterra el descenso de este período se debe a la presencia de yacimientos sajones, con frecuencias menores de R1b. En Francia la muestra es pequeña y concentrada en el yacimiento de Sarreburgo, con restos de hace entre 1.742 y 1.587 años, es decir, posteriores a la caída del Imperio Romano y en pleno período de las invasiones de francos y alamanes. Posteriormente se recuperan unos valores ligeramente más bajos que en milenios anteriores, alcanzándose unos valores de alrededor de 0,8 en la actualidad (Fig. 3).

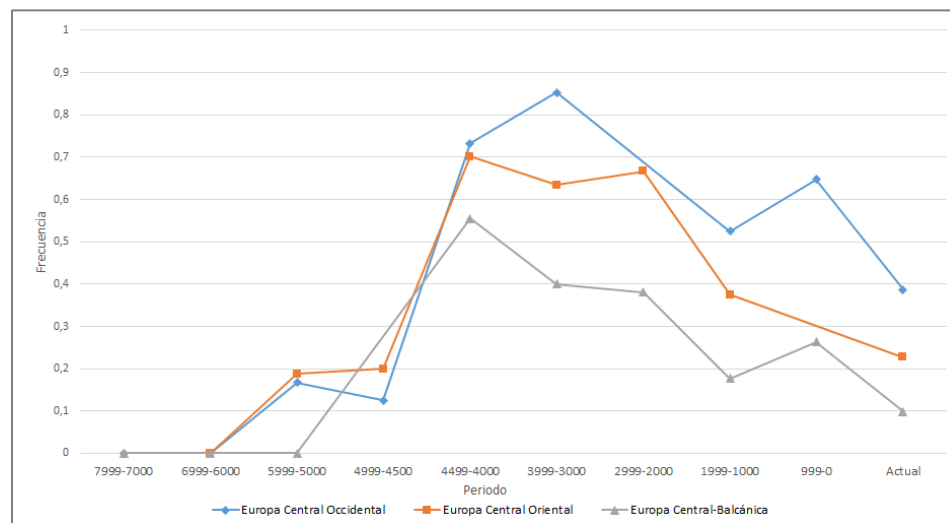


Figura 2. Variación temporal de la frecuencia de R1b en Europa Central.
Figure 2. Temporal variation of R1b frequency in Central Europe.

Sorprendentemente hay un individuo R1b-M269 de 5.700 años en Inglaterra (Ingleborough Hill, Fox Holes Cave) y otro de 4900 años (Blaydon, Tyne and Wear, Summerhill) si bien son 2 casos aislados, ya que el resto comienza a aparecer a partir de hace 4.300 años asociados a la cultura del vaso campaniforme (Olalde *et al.*, 2018). En Francia aparecen también hace unos 4.300 años (La Clape, Grotte Basse de la Vigne Perdue).

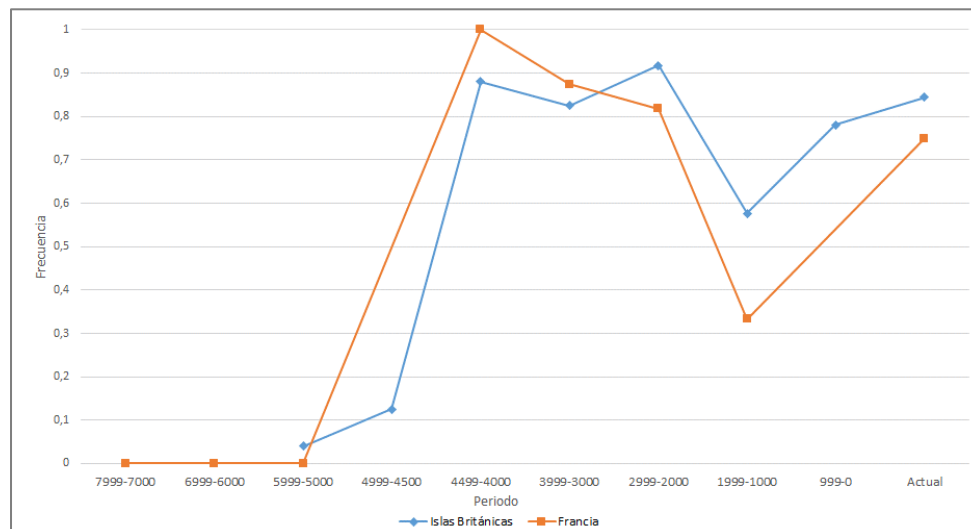


Figura 3. Variación temporal de la frecuencia de R1b en las Islas Británicas y Francia.
Figure 3. Temporal variation of R1b frequency in the British Isles and France.

En el sur de Europa, los primeros registros corresponden a los Balcanes, correspondiendo con los últimos representantes del yacimiento Iron Gates. En todo caso, todas las regiones alcanzan las frecuencias más altas de R1b a partir de hace 4.499-4.000 años, las cuales se mantendrán relativamente altas hasta hace 1.999-1.000 años, cuando experimentan un considerable descenso. Después volverán a subir, excepto en la Península Ibérica, donde toma un valor de 0 en el periodo de hace 999-0 años. Esta frecuencia se debe a la composición de la muestra, ya que se corresponde con individuos encontrados en yacimientos de Granada y Valencia, de origen islámico, lo que justifica la ausencia del haplotipo.

En general, se observa que los valores en el sur de Europa son menores hacia el este, aunque sigan un patrón temporal similar (Fig. 4).

En España R1b-M269 aparece hace unos 4.300 años en un yacimiento asociado a los hombres de las estepas (Monasterio de Rodilla, Burgos) pasando a ser rápidamente el único haplogrupo del cromosoma Y (Olalde *et al.*, 2019) y en Macedonia hace unos 3.200 años (Dimov Grob).

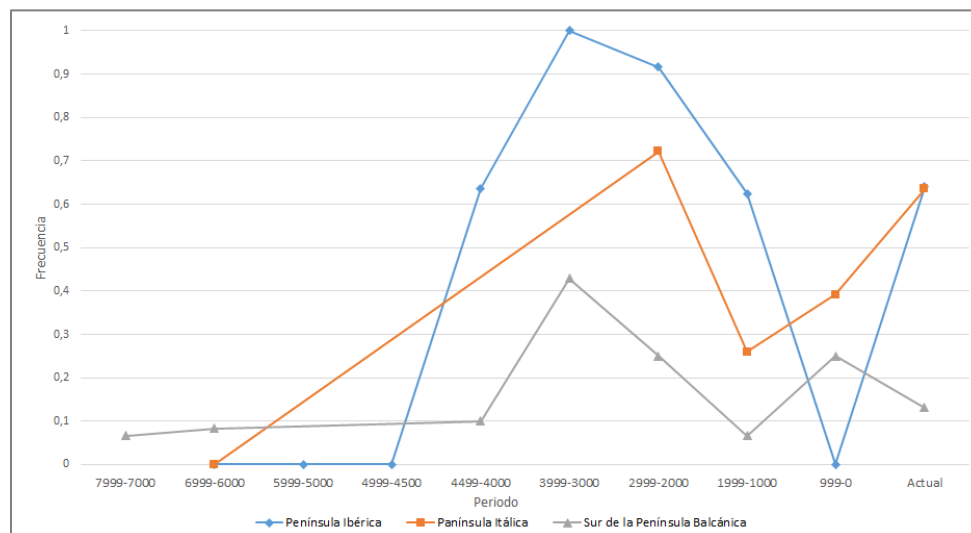


Figura 4. Variación temporal de la frecuencia de R1b en el Sur de Europa.
Figure 4. Temporal variation of R1b frequency in Southern Europe.

En Oriente Próximo las frecuencias de R1b tienden a ser muy bajas, hasta que experimentan un ligero aumento hace 1.999-1.000, que propiciará en el periodo posterior a este un máximo de 0'5 en el caso de Iraq y el Levante Mediterráneo y de 0'2 en el de Turquía, para volver a descender rápidamente en ambas regiones tomando valores cercanos a 0'1 en la actualidad (Fig. 5).

R1b-M269 se detecta por primera vez hace unos 3500 años en Turquía (Hatay).

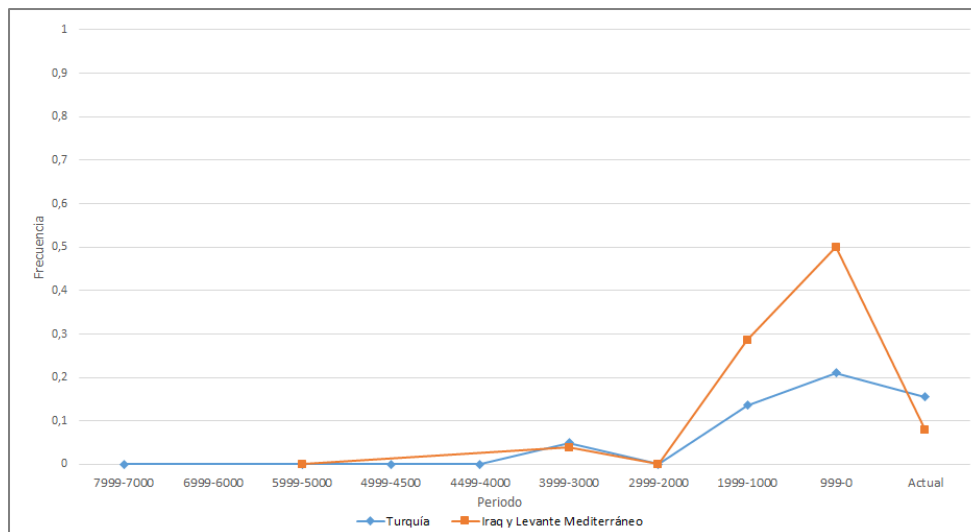


Figura 5. Variación temporal de la frecuencia de R1b en Oriente Próximo.
Figure 5. Temporal variation of R1b frequency in the Near East.

En Asia se ha observado un individuo probablemente R1b-M269 en Shatar Chuluu (Mongolia) asociado a la cultura Afanasievo, con unos 5.000 años de antigüedad y varios más con unos 4.500 años. La cultura Afanasievo se considera descendiente de la cultura pre Yamnaya Repin, que migró desde las estepas euroasiáticas a la región de los montes Altai (Anthony, 2010).

Entre otros individuos R1b-M269, cabe destacar un individuo de la cultura Yamnaya de unos 4.900 años de antigüedad en Karagash (Kazajistán) y otro de 4.200 años en Xinjiang, actual territorio uigur.

Mapas sintéticos

A continuación, se han confeccionado los mapas sintéticos de las frecuencias obtenidas en las diferentes zonas geográficas para cada periodo analizado, ordenados desde el más antiguo hasta la actualidad.

En el periodo 7999-7000 se observan frecuencias cercanas a valores entre 0'1 y 0'4 en regiones del este de Europa, siendo más altas hacia el norte. En la mitad occidental del continente, cuando existen datos, se corresponden con valores nulos, al igual que ocurre con Oriente Próximo (Fig. 6A).

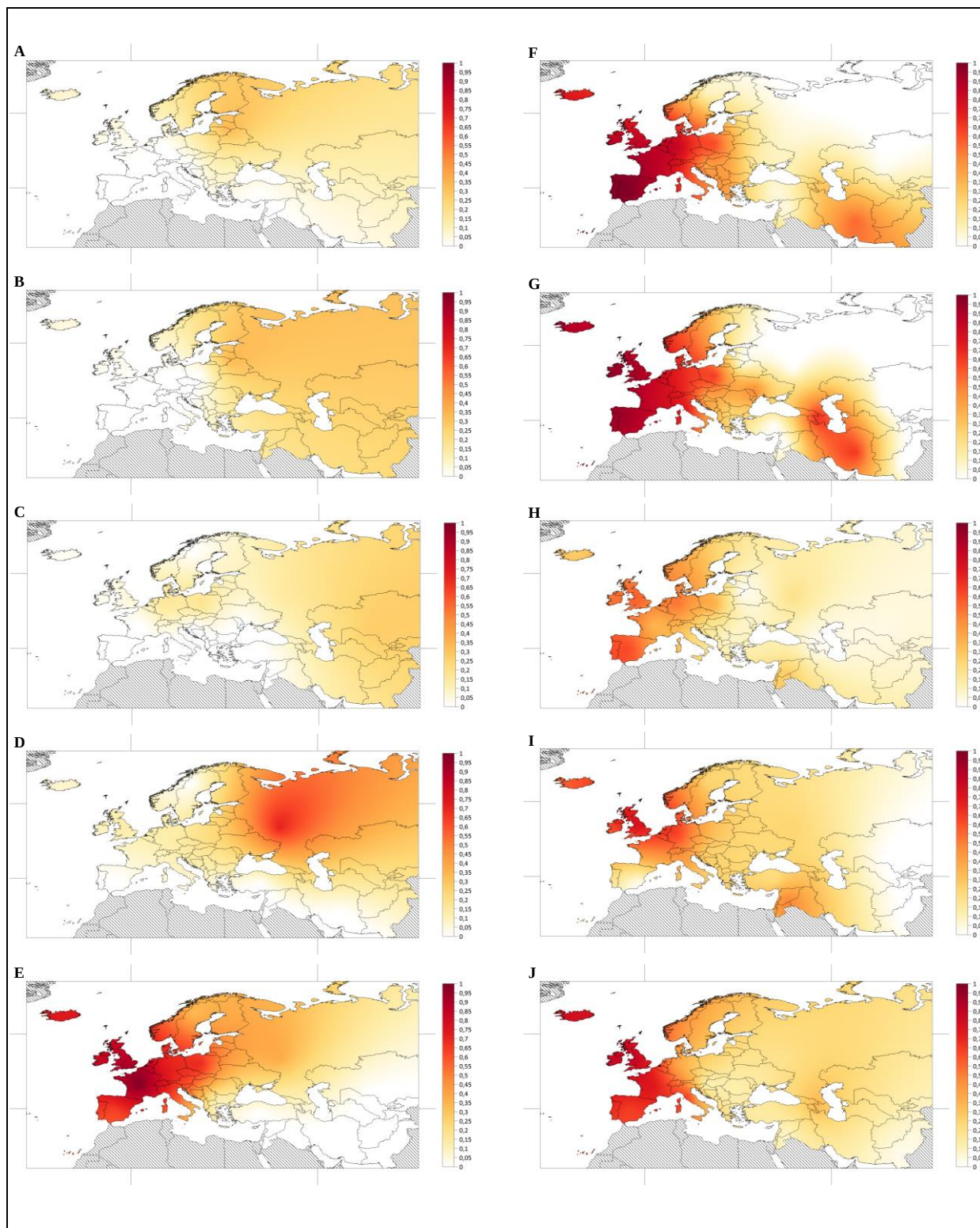


Figura 6. Mapas sintéticos de las frecuencias obtenidas para R1b en los periodos: 7999-7000 (A), 6999-6000 (B), 5999-5000 (C), 4999-4500 (D), 4499-4000 (E), 3999-3000 (F), 2999-2000 (G), 1999-1000 (H), 999-0 (I) y la actualidad (J).

Figure 6. Synthetic maps of the frequencies obtained for R1b in the periods: 7999-7000 (A), 6999-6000 (B), 5999-5000 (C), 4999-4500 (D), 4499-4000 (E), 3999-3000 (F), 2999-2000 (G), 1999-1000 (H), 999-0 (I) and the present (J).

Entre 6999 y 6000 las frecuencias halladas para Europa oriental durante este periodo se encuentran de nuevo en el intervalo de entre 0'1 y 0'4, mostrándose más altos en el norte. En cuanto al oeste de Europa, los valores observados son en todos los casos de 0. Para la región de Oriente Próximo, no se han obtenido frecuencias (Fig. 6B).

Entre 5999 y 5000 se pueden observar frecuencias ligeramente por debajo de 0'2 sobre todo en Europa Central y en el oeste de Rusia. El resto de datos obtenidos a lo largo de todo el territorio europeo y de Oriente Próximo es nulo o, como ocurre en el caso de las Islas Británicas, muy cercano a 0 (Fig. 6C).

Entre 4999 y 4500 el valor más alto observado, de alrededor de 0'7 se encuentra en el Occidente Ruso, que va diluyéndose hacia el oeste de manera progresiva. En el caso de Europa Central e Islas Británicas las frecuencias tomarían valores cercanos a 0'15, mientras que en la Península Ibérica la frecuencia obtenida ha sido de 0. Además de esto, también se han observado en Oriente Próximo y en los Países Nórdicos frecuencias nulas (Fig. 6D).

En el periodo 4499-4000 las frecuencias más altas se concentran en el oeste de Europa, donde llegan a valores superiores a 0'9, como en el caso de las Islas Británicas y Francia. Estos valores van descendiendo hacia el este, donde se encuentran valores entre 0'55 y 0'75 en Europa Central y todavía más bajos en el occidente de Rusia, con una frecuencia alrededor de 0'4. Hacia el sur, en la Península Balcánica y Oriente Próximo, todavía se hace más notable este descenso de las frecuencias con valores no superiores a 0'1 (Fig. 6E).

Durante el intervalo 3999-3000, el máximo de las frecuencias se encuentra en la Península Ibérica, con un valor de 1. Este valor va decreciendo gradualmente hacia las zonas geográficas que se encuentran más hacia el este, hasta llegar a valores por debajo de 0'1 en el occidente de Rusia y Oriente Próximo. También puede observarse otro foco de frecuencias relativamente altas en el sudoeste de Asia, con valores cercanos a 0'55 (Fig. 6F).

Entre 2999 y 2000 en Europa hay una tendencia descendente de las frecuencias hacia el este, teniendo el valor máximo en la Península Ibérica por encima de 0'9. Se puede diferenciar a su vez un foco de frecuencias de alrededor de 0'65 en el sudoeste de Asia de forma aislada (Fig. 6G).

Entre 1999 y 1000 las mayores frecuencias que se han obtenido en este periodo no superan en ningún caso el valor de 0'65, siendo el más alto en la Península Ibérica, seguido por las Islas Británicas y Europa Central Occidental. Se puede distinguir de nuevo una tendencia descendente del valor de las frecuencias hacia la parte oriental. También se obtiene por primera vez un valor para Islandia e Islas Feroe, cercano a 0'3 (Fig. 6H).

Entre el año 999 y la actualidad, la frecuencia más alta obtenida ronda los 0'8 en las Islas Británicas seguida por Europa Central Occidental y las Islas Feroe e Islandia, con un valor cercano a 0'6. Hacia el este de Europa las frecuencias son, en su mayor parte, de alrededor de 0'25 (Fig. 6I). Cabe destacar que el valor nulo que se encuentra en la Península Ibérica se corresponde con individuos encontrados en yacimientos de Granada y Valencia, de origen islámico, lo que podría explicar la ausencia del haplotipo, como ya se ha apuntado anteriormente.

En la actualidad, en las Islas Británicas se encuentra la frecuencia más alta, cercana a 0'85, seguida por Francia y la Península Ibérica con valores entre 0'65 y 0'75 aproximadamente. Estas frecuencias descienden progresivamente hacia las zonas más orientales del territorio analizado (Fig. 6J). Prácticamente la totalidad de los haplogrupos detectados a partir de estos SNPs en las poblaciones actuales han sido R1b-M269.

Análisis de Componentes Principales

Con el fin de valorar de forma conjunta la variación temporal de las frecuencias entre todos los territorios, se ha realizado un Análisis de Componentes Principales sobre la matriz de varianzas y covarianzas. Previamente, para valorar la adecuación del método, se realizó un test de normalidad. El resultado de la W de Shapiro Wilk, con la corrección de Bonferroni no dio significación en ningún caso. Con el fin de soslayar la existencia de valores perdidos, se ha realizado una imputación iterativa: los valores perdidos se sustituyen inicialmente por la media de su columna y se realiza un ACP previo para calcular los valores por regresión de los datos que faltan (Ilin & Reaiko, 2010).

Se puede observar una agrupación desde hace 7999 años hasta hace 4500, que son los que conforman los periodos analizados más tempranos, en el lado negativo del eje del Componente 1 (Fig. 7). A partir de esa fecha, se da un cambio drástico, en el periodo de hace 4499-4000 años (codificado como 5), el cual toma el mayor valor en el extremo positivo del Componente 1. Los periodos que suceden a este también se encuentran en el lado positivo del eje. Por esto se diferenciarían dos claros grupos, los que se encontrarían antes de hace 4500 años y los posteriores a esta fecha.

El Componente 2, por su parte, diferencia tendencias más recientes separando por un lado el período 999-0 (9) de los períodos 3999-3000 y 2999-2000 (6 y 7).

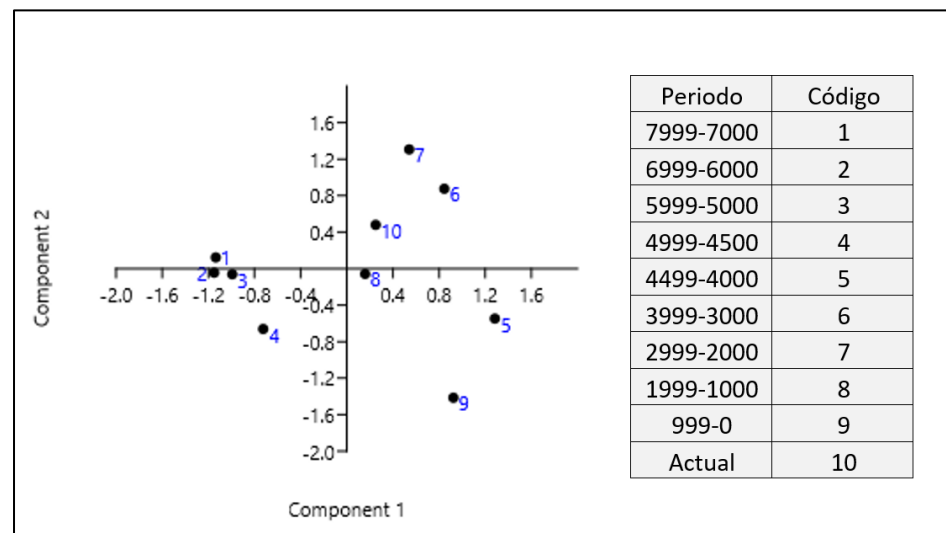


Figura 7. Análisis de Componentes Principales de las frecuencias de R1b en diferentes regiones y varios períodos de tiempo. Varianza explicada por el Componente 1: 55.01% y por el Componente 2: 38.95%.

Figure 7. Principal Component Analysis of R1b frequencies in different regions and various time periods. Variance explained by Component 1: 55.01% and by Component 2: 38.95%.

Discusión

El haplogrupo R1b se encuentra presente en Europa al menos desde el Mesolítico, como se deduce de los genotipos de varios individuos de este período incluidos en este trabajo y como han descrito previamente otros autores (Fu *et al.*, 2016; Mathieson *et al.*, 2018). Su presencia dataría de al menos 14000 años (en Villabruna, Italia) y se extendería en este período por una buena parte del continente, desde Italia y Serbia en el sur, incluyendo a Rumanía y Ucrania, hasta Letonia en el norte.

R1b-M269, por el contrario, parece posterior, ya que el individuo más antiguo registrado en este trabajo data de hace 5100 años, correspondiendo al yacimiento de Samara (Rusia), de cultura Yamnaya (Mathieson *et al.*, 2015)

Si se analizan los datos disponibles para el período más antiguo que hemos considerado, desde hace 7999 años hasta hace 7000, se puede observar una distribución peculiar de las frecuencias, concentradas en los yacimientos de los Países Bálticos y Ucrania principalmente (Fig. 6A). Posteriormente, durante el siguiente milenio (6999-6000) se aprecia una expansión hacia el sur y el este, encontrándose frecuencias apreciables también en Asia central y occidental (Fig. 6B). También se observan frecuencias apreciables en el oeste de Rusia, de donde no hay datos para el milenio anterior. Durante estos dos milenios no se han encontrado evidencias de la presencia de los alelos R1B de ninguno de los SNPs analizados en Europa occidental. Aparentemente, por tanto, el foco original sería la región nororiental de Europa, para posteriormente expandirse hacia el sur y el oeste. Esto concordaría a grandes rasgos con la hipótesis que propone Merpert (1974), según la cual el oeste de Rusia y Ucrania serían el territorio original de la cultura Yamnaya. Y también también a grandes rasgos con la hipótesis de Gimbutas (1993), quien propone que los protoindoeuropeos o kurganes se habrían expandido desde el sur de Rusia hacia el Dnieper y el Mar Negro hacia la mitad del 5º milenio antes de Cristo. Del mismo modo, podría asumirse a la cultura Jvalynsk (entre hace 6700 y 5800 años) y la Sredny Stog (entre hace 6500 y 5500 años) como los orígenes posibles de la cultura Yamnaya, ya que estas culturas pertenecieron a las regiones de Rusia Occidental, en zonas cercanas al río Volga la primera y al norte del mar Negro, en Ucrania y países próximos la segunda, como afirma Mallory (2003).

Por lo que se refiere a su desarrollo temporal, Morgunova y Khokhlova (2013) analizan los yacimientos situados en la cuenca del Volga y proponen una etapa previa (entre hace 6000 y 5300 años), una etapa clásica o avanzada (entre hace 5300 y 4600 años) dividida en dos fases, A (entre hace 5300 y 4900 años) y B (entre hace 4900 y 4600 años) y una etapa tardía (entre hace 4600 y 4300 años).

A grandes rasgos, por tanto, la aparición de R1b-M269 en Europa parece ir asociada al origen y expansión de los Yamnaya a partir de la fase A.

En el milenio 5999-5000 las frecuencias se encuentran distribuidas de forma más irregular y con menor presencia en los territorios anteriormente mencionados; es llamativa particularmente la existencia de una frecuencia de 0'27 en Asia Central (*Fig. 6C*). Probablemente, una parte de la distribución observada en este período sea debida a los yacimientos representados en la base de datos. Así, en Ucrania sólo se encuentran en este período individuos de la cueva Verteba, asociados a la primera cultura neolítica de la región y sin parentesco con los Yamnaya (Gelabert *et al.*, 2022). En todo caso, puede apreciarse una primera expansión de haplogrupos R1b hacia el este y el oeste presumiblemente asociados a los hombres de las estepas, quizá motivada por la llegada de otros grupos a esta región, portadores de culturas neolíticas.

Para la primera mitad del milenio posterior, desde hace 4999 hasta hace 4500 años, se obtiene de nuevo un patrón muy similar a los observados anteriormente como se ha podido confirmar mediante el Análisis de Componentes Principales (*Fig. 7*), presentando frecuencias en el este de Europa, que en este caso son mucho más elevadas (*Fig. 6D*). Las altas frecuencias observadas en esta mitad del milenio corresponden a individuos asociados a la cultura Yamnaya o emparentados con culturas pre-yamnaya, como los encontrados en los yacimientos de Samara y Afanasievo; este periodo coincide en buena medida con la fase B de la etapa avanzada de los Yamnaya (4900–4600) (Morgunova & Khokhlova, 2013). Dicho aumento sustancial de las frecuencias en comparación a periodos anteriores puede ser la consecuencia de una mejoría en la productividad del ganado provocada por un aumento en la humedad en la zona que beneficiaría a los pastos que dicho ganado consumiría (Kristiansen *et al.*, 2017).

En el período 4499-4000, el patrón de períodos anteriores experimenta un cambio radical como refleja el ACP (*Fig. 7*). En este caso, los valores más altos se encuentran hacia el oeste de Europa (*Fig. 6E*), dejando valores más bajos que en periodos anteriores en Rusia Occidental (*Fig. 1*), lo que podría corresponderse con la última época de la cultura Yamnaya, que se encontraría en su fase tardía (4600–4300). En Europa central y occidental se observan aumentos en las frecuencias de R1b por todas las regiones; este cambio drástico de las frecuencias podría explicarse por una migración masiva y muy rápida ocurrida hace 4500 años aproximadamente por parte de la civilización Yamnaya en dirección oeste (Haak *et al.*, 2015; Allentoft *et al.*, 2015), posiblemente facilitada por la domesticación del caballo hace unos 5000 años, lo que conllevaría un medio de transporte más eficaz (Trautmann *et al.*, 2023). Efectivamente, se observan valores de alrededor de 0'7 en Centroeuropa, cerca de 0'8 en Francia, de 1 en las Islas Británicas, más de 0'6 en la Península Ibérica y valores sustanciales, aunque más bajos, en otras regiones del norte y del sur de Europa. La frecuencia de 1 en las Islas Británicas, así como en el siguiente período en la Península Ibérica probablemente se han alcanzado por un cierto efecto de la deriva en las poblaciones descendientes de los Yamnaya que llegaron a estos lugares, pero también por un proceso de dominación sobre las poblaciones locales, de modo que se imponen esencialmente los linajes masculinos. Esto puede observarse en la Península Ibérica, donde se sustituyen completamente los linajes masculinos previos, pero tan sólo cambia el 40% del patrimonio genético total (Olalde *et al.*, 2019).

El flujo de genes hacia el oeste habría ido acompañado de una asimilación o bien de una fusión con otras culturas. Así, se observan rituales muy similares entre la cultura Yamnaya y la cultura de la cerámica cordada (5000–4350), por lo que se ha propuesto una relación cultural y genética entre ambas (Kristiansen *et al.*, 2017; Allentoft *et al.*, 2015; Lazaridis *et al.*, 2014). Además, se conoce una relación entre la dispersión de la cultura de la cerámica cordada y tendencias migratorias desde el nordeste hacia el sudoeste (Price *et al.*, 1998).

En cuanto a las regiones que se encuentran más hacia el oeste, durante este periodo predomina la cultura del vaso campaniforme, que en su versión septentrional guarda un alto grado de parentesco genético con los individuos de las estepas, aunque un poco menor que con los de la cerámica cordada (Allentoft *et al.*, 2015; Olalde *et al.*, 2018).

Un nuevo cambio sustancial en las frecuencias ocurre en el periodo que transcurre desde hace 1999 años hasta justo antes de la actualidad (*Fig. 6H y 6I*), ya que en Europa central y occidental descienden considerablemente. Durante este período ocurre en primer lugar la máxima expansión del Imperio Romano y después la desintegración del Imperio Romano de Occidente. Debido a ello ocurre primero un cierto flujo génico desde la mitad oriental del Mediterráneo, donde las frecuencias de R1b son más bajas, por influencia de la romanización y después un nuevo flujo génico desde el centro y este de Europa por la llegada de los bárbaros y la expansión del Imperio Huno (Lindner, 1995) además

del asentamiento del kanato ávaro a lo largo de la llanura panónica (Pohl *et al.*, 2018), que probablemente también contribuye a diluir las frecuencias de este haplogrupo. Así, en las Islas Británicas todos los individuos que no portan haplogrupos R1b son sajones y vikingos, excepto un individuo procedente de Próximo Oriente, en tanto que en España y Portugal son romanos procedentes de África.

En los últimos 2000 años hay una bajada de frecuencias generalizadas en la Europa más occidental a la que sucede un aumento aparente en las poblaciones actuales. La razón puede estar en una cierta sobrerrepresentación de yacimientos correspondientes a las invasiones asociadas a los romanos, los bárbaros y en el caso de la Península Ibérica los árabes, cuya influencia real fue finalmente menor. Es el caso, por ejemplo, de los visigodos y los árabes en la Península Ibérica, que dejaron muy poca huella en su patrimonio genético (Olalde *et al.*, 2018).

Conclusiones

El haplogrupo R1b se encuentra presente en Europa al menos desde el Mesolítico, con unos 14000 años de antigüedad mínima.

R1b-M269, derivado de R1b, aparece asociado a la cultura Yamnaya, con una antigüedad mínima de 5100 años.

La expansión de R1b-M269 en Europa occidental parece ir asociada a las migraciones asociadas a la expansión de la cultura Yamnaya y otras culturas descendientes de la misma, a partir de hace 4500 años.

La disminución de la frecuencia de R1b-M269 en el oeste de Rusia, donde probablemente se originó, coincide en el tiempo con el declive de la cultura Yamnaya.

Como consecuencia de la expansión Yamnaya y probablemente asociadas a procesos de deriva genética y dominación cultural, se observan unas sustituciones completas o casi completas de linajes previos del cromosoma Y en algunas regiones, como las Islas Británicas, Francia y la Península Ibérica.

Entre hace 2000 y hace 1000 años las frecuencias de R1b-M269 descendieron en el extremo occidental de Europa, probablemente por las migraciones asociadas a la expansión del Imperio Romano primero y las invasiones bárbaras después.

La sucesión de eventos descrita explica verosímelmente la existencia de una clina en Europa que aumenta de este a oeste para el haplogrupo del cromosoma Y R1b-M269, cuyo origen se encuentra sin embargo se encuentra en el extremo nororiental del continente.

Referencias

- Allentoft, M. E., Sikora, M., Sjögren, K. G., *et al* (2015). Population genomics of Bronze Age Eurasia. *Nature*, 522(7555), 167–172. <https://doi.org/10.1038/nature14507>
- Anthony, D. W. (2010). *The horse, the wheel, and language: How bronze-age riders from the Eurasian steppes shaped the modern world*. Princeton University Press.
- Balaresque, P., Bowden, G. R., Adams, S. M., *et al.* 2010. A predominantly neolithic origin for European paternal lineages. *PLoS biology*, 8(1), e1000285. <https://doi.org/10.1371/journal.pbio.1000285>
- Fu, Q., Posth, C., Hajdinjak, M., *et al.*, 2016. The genetic history of Ice Age Europe. *Nature*, 534(7606), 200–205. <https://doi.org/10.1038/nature17993>
- Gelabert, P., Schmidt, R. W., Fernandes, D. M., *et al.* (2022). Genomes from Verteba cave suggest diversity within the Trypillians in Ukraine. *Scientific reports*, 12(1), 7242. <https://doi.org/10.1038/s41598-022-11117-8>
- Gimbutas, M. (1993). The Indo-Europeanization of Europe: the intrusion of steppe PASToralists from south Russia and the transformation of Old Europe, *WORD*, 44:2, 205-222. <https://doi.org/10.1080/00437956.1993.11435900>
- Haak, W., Lazaridis, I., Patterson, N., *et al.* (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522(7555), 207–211. <https://doi.org/10.1038/nature14317>

- Hammer, Ø., Harper, D., and Ryan, P., 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, 4(1):1-9.
- Ilin, A., T. Raiko., 2010. Practical approaches to Principal Component Analysis in the presence of missing values. *Journal of Machine Learning Research* 11:1957-2000.
- Kristiansen, K., Allentoft, M., Frei, K., *et al.* (2017). Re-theorising mobility and the formation of culture and language among the Corded Ware Culture in Europe. *Antiquity*, 91(356), 334-347. <https://doi.org/10.15184/aqy.2017.17>
- Lazaridis, I., Patterson, N., Mittnik, A., *et al.* (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 513(7518), 409–413. <https://doi.org/10.1038/nature13673>
- Lindner, R. P. (1981). Nomadism, Horses and Huns. *Past & Present*, 92, 3–19.
- Mallick, S., Reich, D., 2023. The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes (V8). Harvard Dataverse. <https://doi.org/10.7910/DVN/FFIDCW>
- Mallory, J.P. (2003). *In search of the indo-europeans: Language, archaeology, and myth*. Thames & Hudson.
- Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., *et al.*, 2018. The genomic history of southeastern Europe. *Nature*, 555(7695), 197–203. <https://doi.org/10.1038/nature25778>
- Mathieson, I., Lazaridis, I., Rohland, N., *et al.* (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528(7583), 499–503. <https://doi.org/10.1038/nature16152>
- Merpert, N. I. (1974). *Drevneishie skotovody Volzhsko-Ural'skogo mezhdurech'ia*. Nauka.
- Morgunova, N., Khokhlova, O., 2013. Chronology and Periodization of the Pit-Grave Culture in the Region Between the Volga and Ural Rivers Based on Radiocarbon Dating and Paleopedological Research. *Radiocarbon*, 55(3), 1286-1296. <https://doi.org/10.1017/S0033822200048190>
- Olalde, I., Brace, S., Allentoft, M. E., *et al.* (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*, 555(7695), 190–196. <https://doi.org/10.1038/nature25738>
- Olalde, I., Mallick, S., Patterson, N., *et al.* (2019). The genomic history of the Iberian Peninsula over the PAST 8000 years. *Science* (New York, N.Y.), 363(6432), 1230–1234. <https://doi.org/10.1126/science.aav4040>
- Patterson, N., Price, A. L., Reich, D., *et al.*, 2013. EIGENSOFT v5.01. Harvard University.
- Peña, J.A., 2023. Comunicación personal.
- Pohl, W., Sayers, W. (2018). *The Avars: A Steppe Empire in Europe, 567-822*. Cornell University Press.
- Price, T., Grupe, G., Schröter, P. (1998). Migration in the Bell Beaker period of central Europe. *Antiquity*, 72(276), 405-411. <https://doi.org/10.1017/S0003598X00086683>
- Trautmann, M., Frînculeasa, A., Preda-Bălănică, B., *et al.* (2023). First bioanthropological evidence for Yamnaya horsemanship. *Science advances*, 9(9), eade2451. <https://doi.org/10.1126/sciadv.ade2451>
- Valero Almingol, A., Peña, J.A., 2021. Origen de la variación clinal en el genoma humano europeo. *Antropo*, 46, 41-56.