

**Análisis Genético del Individuo de Chan do
Lindeiro: Caracterización de su Mitogenoma
y Situación de la Muestra en el Contexto
Paleogenético Europeo**
Genetic Analysis of the Individual of Chan do
Lindeiro: Characterization of her Mitogenome
and Situation of the Sample in the European
Paleogenic Context

GONZALEZ FORTES, G.^{1,2,3}, GRANDAL D'ANGLADE, A.³, VIDAL ROMANÍ, J.R.³
y HOFREITER, M.^{1,2}

(1) Department of Biology, The University of York, YO10 5DD, Reino Unido

(2) Institute for Biochemistry and Biology. University of Potsdam, 14476 Potsdam OT Golm, Alemania

(3) Instituto Universitario de Xeoloxía. Universidade da Coruña. ESCI, Campus de Elviña s/n. 15071
A Coruña, España

Resumen

En este trabajo se presentan los primeros resultados de un estudio de paleogenética realizado en los restos óseos de Elba, la mujer mesolítica de Chan do Lindeiro (Pedrafita do Cebreiro, Lugo). El estudio se realizó a nivel de ADN mitocondrial completo y de marcadores nucleares relacionados con rasgos fenotípicos, como pigmentación y tolerancia a la lactosa. El método se basa en el enriquecimiento mediante captura por hibridación de ADN mitocondrial y genómico, seguido de secuenciación de alto rendimiento. La cobertura obtenida para los marcadores genómicos resultó insuficiente, pero el genoma mitocondrial se recuperó en un 99%. La mujer de Chan do Lindeiro pertenece al haplogrupo U, característico de los cazadores recolectores europeos, y dentro de éste, en el subhaplotipo U5b1, cuyo origen se sitúa en la Península Ibérica hace unos 16-20000 años.

Palabras clave: ADN mitocondrial antiguo, secuencias NGS, Mesolítico, Chan do Lindeiro, Galicia

Abstract

In this work the first results of a paleogenetic study are presented in the bone remains of Elba, the mesolithic woman of Chan do Lindeiro (Pedrafita do Cebreiro, Lugo). The study was performed at the level of complete mitochondrial DNA and of nuclear markers related to phenotypic traits, such as pigmentation and lactose tolerance. The method is based on enrichment by hybridization capture of mitochondrial and genomic DNA, followed by high throughput sequencing. The coverage obtained for the genomic markers was insufficient, but the mitochondrial genome was recovered by 99%. The Chan do Lindeiro woman belongs to the haplogroup U, characteristic of the European hunter gatherers and within this, in the subhaplotype U5b1, whose origin is located in the Iberian Peninsula some 16-20000 years ago.

Key words: Ancient mDNA, NGS sequencing, Mesolithic, Chan do Lindeiro, Galicia

INTRODUCCIÓN

Desde el punto de vista genético, las poblaciones mesolíticas de la Europa Occidental se caracterizan por su homogeneidad. Los estudios genéticos de restos de cazadores-recolectores europeos coinciden en la prevalencia del haplogrupo mitocondrial U, desde el Pleistoceno Superior al Holoceno y hasta el comienzos del Neolítico, cuando nuevos haplogrupos asociados a los primeros agricultores empiezan a aparecer en Europa (PINHASI *et al.*, 2012; LAZARIDIS *et al.*, 2014, JONES *et al.*, 2015). Un estudio reciente que reúne mas de 60 muestras de cazadores-recolectores europeos de hasta 35000 años de antigüedad (Posth *et al.*, 2016) explica esta prevalencia como el resultado de una crisis demográfica en torno al último máximo glacial hace unos 14000 años, que habría reducido drásticamente la variabilidad genética de las poblaciones ancestrales en Europa.

Particularmente en la Península Ibérica se han estudiado un total de 25 muestras humanas prehistóricas a nivel de ADN mitocondrial completo (SÁNCHEZ-QUINTO *et al.*, 2012; GUNTER *et al.*, 2015; OLALDE *et al.*, 2015; POSH *et al.*, 2016). Solo una de estas muestras corresponde al periodo mesolítico en España (La Braña 1, SÁNCHEZ-QUINTO *et al.*, 2012; OLALDE *et al.*, 2014), mientras que las otras 24 presentan dataciones correspondientes a diferentes épocas del Neolítico. El individuo de La Braña fue asignado al haplotipo U5b2c1 (SÁNCHEZ-QUINTO *et al.*, 2012; OLALDE *et al.*, 2015) y un 16.6% de las muestras neolíticas se asignaron a diferentes subhaplotipos del U5b, lo que constata la supervivencia de este grupo en la Península durante la época Neolítica, si bien se observa una disminución de su fre-

cuencia en consonancia con lo que ocurre en el resto de Europa con la llegada de la agricultura.

A nivel de ADN nuclear, el reciente desarrollo de nuevas tecnologías de captura y secuenciación genómica ha permitido la recuperación de un elevado número de marcadores nucleares en muestras prehistóricas y en algunos casos, de genomas nucleares completos (LAZARIDIS *et al.*, 2014; 2016; GAMBA *et al.*, 2015, JONES *et al.*, 2015, HAAK *et al.*, 2015). En base al estudio de estos genomas antiguos y su comparación con poblaciones actuales se ha llegado a la identificación de al menos tres componentes ancestrales en las poblaciones europeas: una componente que deriva de poblaciones paleolíticas del Norte Euroasiático; una que procede de las poblaciones de cazadores-recolectores de la Europa Occidental; y por último, una componente derivada de los primeros granjeros europeos (LAZARIDIS *et al.*, 2014). Todos los individuos mesolíticos analizados hasta la fecha en la Europa Occidental (LAZARIDIS *et al.*, 2014; SCKOGLUND *et al.*, 2014; JONES *et al.*, 2015; HAAK *et al.*, 2015), se sitúan fuera de la variabilidad genética de las poblaciones europeas actuales, pero se agrupan entre ellos formando el clúster denominado WHG (west hunter-gatherers, cazadores-recolectores occidentales), lo que confirma las observaciones basadas en ADNmt en cuanto a la poca variabilidad genética de estas poblaciones mesolíticas.

Además de estudios poblacionales, en algunos casos, la disponibilidad de marcadores nucleares ha permitido estudiar caracteres fenotípicos en muestras prehistóricas. Concretamente, en la Península Ibérica se han podido identificar en los restos de La Braña marcadores nucleares relacionados

con pigmentación de piel, cabello y ojos, lo que llevo a la identificación de ojos de coloración clara por primera vez en un individuo preneolítico. Así, el individuo de La Braña se describió como un hombre de aspecto robusto, tez oscura y ojos azules y esta imagen se ha venido utilizando como imagen característica de los cazadores-recolectores de la Europa Occidental (Fig. 1). Sin embargo, en los últimos dos años, el estudio de nuevas muestras ha llevado a la identificación de distintos fenotipos para pigmentación en el mesolítico europeo (JONES *et al.*, 2015), poniendo en entredicho la prevalencia de los ojos claros en estas poblaciones. El caso de la Braña nos recuerda lo poco que sabemos aun de las poblaciones humanas antiguas, especialmente en la Península Ibérica, y la necesidad de aportar nuevos datos a fin de evitar generalizaciones desacertadas.

En este estudio llevamos a cabo el estudio genético de la muestra humana de Chan do Lindeiro a nivel de ADN mitocondrial completo y de marcadores nucleares relacionados con rasgos fenotípicos, como pigmentación y tolerancia a la lactosa. El ADN mitocondrial ha sido el marcador de elección en la mayor parte de los estudios de ADN antiguo (ADNa) realizados hasta la fecha. Es el ADN de las mitocondrias, que pueden estar presentes en las células vivas en número de cientos, por lo que cada célula contiene centenares de copias de los genes mitocondriales, en contraste con el ADN nuclear, del que solo porta dos copias por gen, la materna y la paterna. Así, tras la muerte

del individuo y con el paso del tiempo, es mas probable que se conserven fragmentos de ADN mitocondrial que de nuclear. En el caso de Elba, una muestra de casi 10000 años de antigüedad, se recurrió a tecnología genómica puntera para la recuperación de ADNa, concretamente se llevo a cabo un proceso de enriquecimiento basado en la captura por hibridación de ADN mitocondrial y genómico, seguido de secuenciación de alto rendimiento en una plataforma Illumina HiSeq2000. Gracias a esta tecnología se pudo verificar la autenticidad del ADN de la muestra y recuperar su genoma mitocondrial con una cobertura media en torno a 50X. En el caso de los marcadores nucleares, la cobertura fue del 0.1x, lo que no es suficiente para su genotipado, por lo que su estudio se esta completando en la actualidad con otro proyecto de mayor envergadura que incluye la secuenciación genómica completa del individuo de Chan do Lindeiro (FORTES *et al.*, en preparación).

Elba es el individuo mas antiguo estudiado a nivel genético en el mesolítico de la Península Iberia y es también la muestra antigua mas occidental estudiada en Europa. La caracterización de su genoma contribuye a profundizar en el conocimiento de las poblaciones antiguas, cuáles eran los vínculos genéticos entre ellas, como se desarrollaron demográficamente a lo largo de la prehistoria y en que medida su legado genético ha llegado a nuestros días, contribuyendo o no a la composición genómica de las poblaciones europeas actuales.

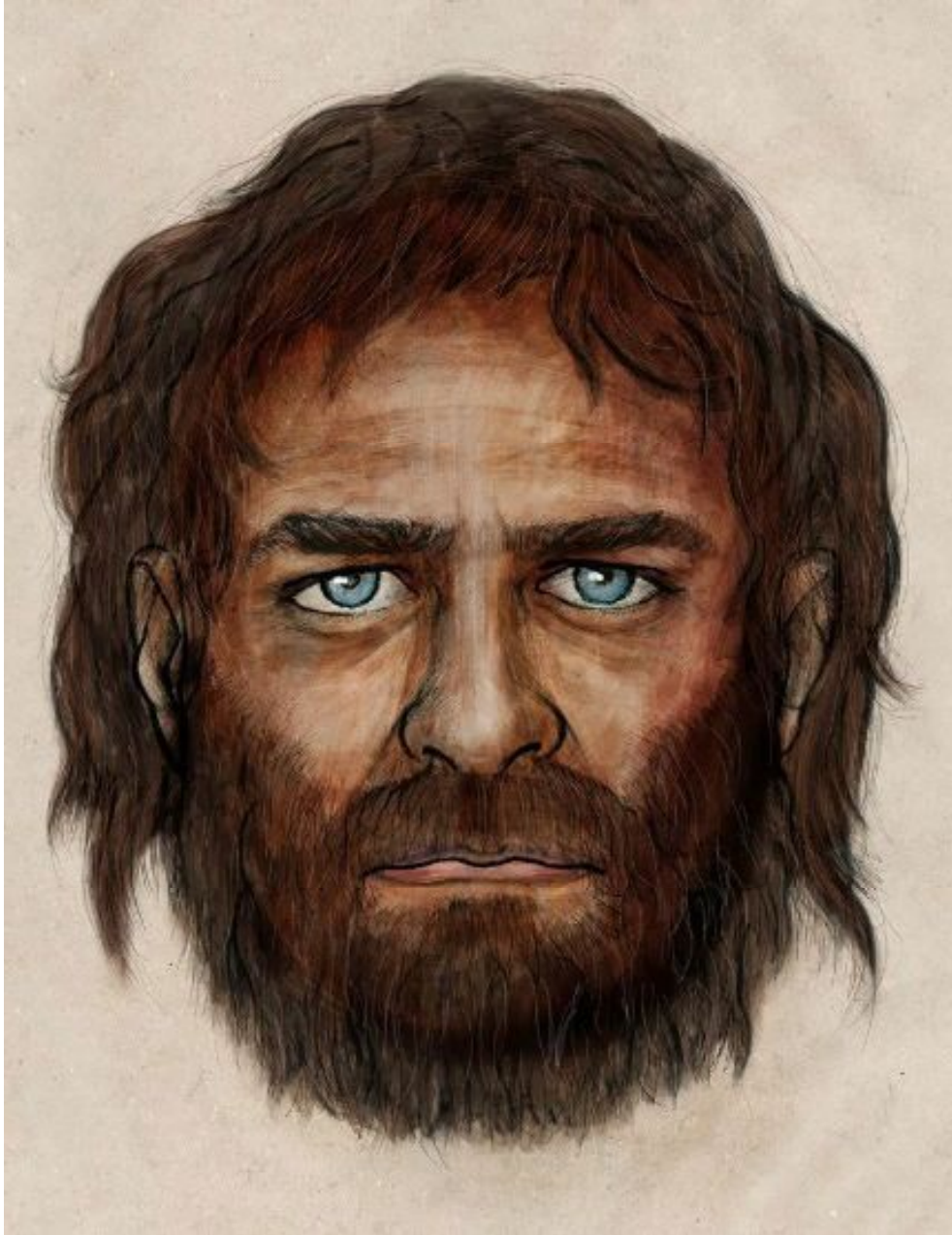


Fig. 1. Reconstrucción del aspecto típico de un cazador-recolector del mesolítico Ibérico basada en las predicciones fenotípicas de OLALDE et al., 2014.

MATERIAL Y MÉTODOS

El procesado de las muestras de Chan do Lindeiro, incluyendo la descontaminación de las mismas, la extracción de ADN y la construcción de librerías genómicas para secuenciación se llevo a cabo en la Universidad de York (Reino Unido), en un laboratorio de ADN especialmente diseñado para la manipulación de restos arqueológicos. Dicho laboratorio dispone de estancias exclusivamente dedicadas al procesado de muestras antiguas, con espacios presurizadas para evitar la entrada de aire no filtrado del exterior, así como lámparas de luz ultravioleta para la esterilización de todas las superficies y materiales que han de estar en contacto con las muestras antiguas.

Descontaminación de la muestra y extracción del ADN

La extracción de ADN se llevo a cabo a partir de material óseo y dentario de la muestra, concretamente se utilizaron un fragmento del hueso femoral y la raíz de un molar. Previamente a la extracción, los restos óseos y dentario fueron descontaminados a fin de minimizar la presencia de ADN exógeno. En este proceso, los fragmentos fósiles se expusieron a luz ultravioleta durante 10 minutos cada lado. A continuación se procedió a la eliminación de la capa mas externa del hueso, utilizando una sierra eléctrica y material de pulido desechable. Finalmente, el hueso pulido se expuso nuevamente a luz u.v. durante 10 minutos por cada una de sus caras. El hueso descontaminado fue pulverizado usando morteros de cerámica previamente esterilizados con lejía durante 24h.

Doscientos cincuenta miligramos de polvo del hueso femoral y 50 mg de polvo de la

raíz del molar fueron utilizados para llevar a cabo dos extracciones independientes. En cada caso el polvo de hueso se incubo con EDTA y proteinasa K a 37 grados y en rotación suave durante 12-16h. A continuación se siguió el protocolo de ROHLAND *et al.* (2010), especialmente diseñado para la extracción de ADN a partir de muestras antiguas. Finalmente se obtuvo un volumen de 50 µl de extracto a partir de cada una de las dos muestras (femoral y raíz).

Construcción de librerías NGS y captura por hibridación

De cada uno de los extractos se utilizaron 20 µl para la construcción de librerías NGS (Next Generation Sequencing) siguiendo el protocolo de MEYER Y KIRCHER 2010, con las modificaciones propuestas por FORTES Y PAIJMANS 2015. En la construcción de las librerías a partir de la muestra de fémur se utilizaron adaptadores marcados con una secuencia index diferente a la utilizada en las librerías construidas a partir de la raíz del molar, a fin de diferenciar el ADN extraído a partir de una y otra muestra. Las librerías construidas se amplificaron en cuatro PCRs paralelas de 12 ciclos cada una, y los productos de estas PCRs se juntaron y purificaron en una columna MinElute. A fin de aumentar el numero de copias de los fragmentos de ADN para la captura, el volumen total del purificado (20 µl) se utilizo como molde para otras cuatro reacciones paralelas de 12 ciclos, y los productos de estos se volvieron a juntar para su purificación en una sola columna MinElute. Este ultimo purificado se cuantifico en un aparato Nanodrop y se sometió al protocolo de enriquecimiento mediante captura por hibridación.

Para la captura de ADN mitocondrial y de los SNPs nucleares se siguió el protocolo de HODGES *et al.* (2009), con las optimiza-

ciones descritas por FORTES y PAIJMANS (2015) para muestras antiguas. La captura se llevo a cabo en un microarray 244k DNA SureSelect™ (Agilent, Boblingen, Germany), cubierto por sondas específicas para la hibridación de ADN mitocondrial humano y de las secuencias de ADN que contienen los SNPs de pigmentación y tolerancia a la lactosa. Para el diseño de las sondas mitocondriales se utilizo como referencia la secuencia revisada de Cambridge (NC_012920), mientras que para los SNPs las sondas se diseñaron a partir de secuencias de 100 pares de bases en torno a los mismos, tomando como genoma de referencia el hg19 (también conocido como Built37). En ambos casos, tanto para el mitocondrial como para el nuclear, se construyeron sondas de 60 nucleótidos de longitud, con un solapamiento de 2 bases hasta cubrir la longitud total de las secuencias de referencia.

El array y las librerías se mantuvieron en contacto y en rotación suave a 65 grados durante 72h. Concluido el tiempo de hibridación, el eluido fue amplificado y usado de nuevo para una segunda ronda de hibridación de otras 72h, siguiendo las indicaciones de FORTES y PAIJMANS (2015).

Secuenciación del ADN y procesado de los datos NGS

El pool de librerías femorales y las del molar se secuenciaron en un lane de la plataforma HiSeq2000 de Illumina, en modo 100SE en el Danish National Sequencing Centre de la Universidad de Copenhague. Los archivos BCL resultantes se convirtieron al formato fastq usando el pipeline de Illumina para la identificación de bases (Illumina Pipeline v1.4). A continuación se uso el programa Cutadapt v1.3 (MARTIN,

2011) para eliminar restos de la secuencia del adaptador P7 en los extremos 3' y la secuencia index del extremo 5', así como todas aquellas secuencias en las que los índices de los extremos 5' y 3' no fueran coincidentes, ya que esto es indicativo de la formación de quimeras durante la amplificación PCR de los productos de la captura. Durante este procedimiento, fueron descartadas también todas las secuencias de longitud inferior a 25 nucleótidos a fin de evitar alineamientos espurios en etapas posteriores. Las secuencias libres de adaptadores que superaron el filtro de longitud, se alinearon al genoma de referencia -rCRS para el mitocondrial y los correspondientes cromosomas del hg19 para la localización de los SNPs-. Para el alineamiento se utilizo el programa bwa-0.5.9 (LI y DURBIN, 2009) y a continuación samtools 0.1.19-44428 para el indexado, eliminación de secuencias clónicas y filtrado de calidad (-q30) de las secuencias mapeadas. Para evitar errores de genotipado asociados a la degradación del ADN en muestras antiguas, se descartaron 5 nucleótidos en cada uno de los extremos 5' y 3' de las secuencias, ya que estos son mas sensibles a la desaminación de las citosinas a timinas (BRIGGS *et al.*, 2007). Todos los sitios polimórficos identificados se confirmaron a través de la visualización directa de los alineamientos en Tablet 1.13.05.02 (MILNE *et al.*, 2013) en el caso de las secuencias mitocondriales y en IGV, en el caso de las secuencias SNP. La cobertura del genoma mitocondrial se determino usando el programa GATK (MACKENNA *et al.*, 2010). A fin de testar la autenticidad de las secuencias de ADN se utilizó el software MapDamage (GINOLHAC *et al.*, 2011), que cuantifica y representa gráficamente la presencia de desaminación de las citosinas en los extremos de las secuen-

cias. Este daño molecular se ha encontrado de forma consistente en muestras antiguas y es una característica que actualmente se utiliza para confirmar la autenticidad de las mismas (KRAUSE *et al.*, 2010; ORLANDO *et al.*, 2011).

Caracterización del haplotipo mitocondrial

Los polimorfismos con una cobertura de 3 o más secuencias fueron utilizados para identificar el haplotipo mitocondrial de los restos de Elba. Para la asignación del haplotipo se usó Haplogrep (<http://haplogrep.uibk.ac.at/index.html>), una herramienta online que permite la comparación automática del conjunto de posiciones polimórficas con la completa y actualizada filogenia mitocondrial de Phylotree (<http://www.phylotree.org/>).

Estima de los niveles de contaminación

La presencia de contaminantes en las secuencias de ADNmt se evaluó como la proporción de bases secundarias (distintas a la base mayoritaria) en posiciones diagnósticas para la definición del haplogrupo. Se consideraron posiciones diagnósticas aquellas identificadas por Haplogrep para asignar un determinado haplotipo. El porcentaje de contaminación en la muestra se estimó como la proporción de bases secundarias del total de secuencias cubriendo las posiciones diagnósticas.

Análisis filogenético

La secuencia consenso del mitocondrial completo de Chan do Lindeiro se alineó a otras 60 secuencias de mitogenomas anti-

guos usando el software MUSCLE. Para el alineamiento se incluyeron las secuencias mitocondriales completas de La Braña (OLALDE *et al.*, 2014), de 35 individuos del Paleolítico Superior y Holoceno europeo publicadas recientemente por POSTH *et al.* (2016) y 24 mitogenomas de distintas épocas del neolítico en España (GUNTHER *et al.*, 2015; HAAK *et al.*, 2015; OLALDE *et al.*, 2015). El alineamiento se utilizó para la construcción de un árbol filogenético de máxima verosimilitud (Maximum Likelihood Tree) en RaxML (STAMAKIS, 2014). Como outgroup se utilizó el mitocondrial completo de un individuo Neanderthal de Vindija (FM865410.1). El árbol se construyó con un 95% de delección parcial y bajo el método GTR+G. El soporte estadístico del árbol generado se testó con 500 replicas bootstrap.

La filogenia generada se visualizó en FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/>) y finalmente se redefinió en Inkscape (<https://inkscape.org/en/>).

Caracterización fenotípica

La Tabla 1 recoge los SNP incluidos en el microarray para la captura de marcadores nucleares asociados con rasgos fenotípicos. Esta tabla incluye los ocho marcadores del sistema 8-plex (SPICHENOK *et al.*, 2011; PNEUMAN *et al.*, 2012) y los 24 del Hirisplex (WALSH *et al.*, 2013) para la predicción del color de ojos, cabello y piel. Además de estos, también se incluyeron en el array sondas que cubren loci relacionados con la tolerancia a la lactosa: rs4988235 y rs4988235.

Tabla 1. Panel de SNPs utilizados en la captura para la predicción de rasgos fenotípicos.

Gen	SNP	Posición cromosómica en hg19	Sistema predictivo	Referencia
MC1R	N29insA		Hirisplex	Walsh <i>et al.</i> , 2013
MC1R	rs11547464	Chr16: 89,986,091	Hirisplex	Walsh <i>et al.</i> , 2013
MC1R	rs885479	Chr16: 89,986,154	8-Plex, Hirisplex	Walsh <i>et al.</i> , 2013, Pneuman <i>et al.</i> , 2012
MC1R	rs1805008	Chr16: 89,986,144	Hirisplex	Walsh <i>et al.</i> , 2013
MC1R	rs1805005	Chr16: 89,985,844	Hirisplex	Walsh <i>et al.</i> , 2013
MC1R	rs1805006	Chr16: 89,985,918	Hirisplex	Walsh <i>et al.</i> , 2013
MC1R	rs1805007	Chr16: 89,986,117	Hirisplex	Walsh <i>et al.</i> , 2013
MC1R	rs1805009	Chr16: 89,986,546	Hirisplex	Walsh <i>et al.</i> , 2013
MC1R	Y152OCH		Hirisplex	Walsh <i>et al.</i> , 2013
MC1R	rs2228479	Chr16: 89,985,940	Hirisplex	Walsh <i>et al.</i> , 2013
MC1R	rs1110400	Chr16: 8,998,6130	Hirisplex	Walsh <i>et al.</i> , 2013
SLC45A2	rs28777	Chr5: 33,958,959	Hirisplex	Walsh <i>et al.</i> , 2013
SLC45A2	rs16891982	Chr5: 33,951,693	8-Plex, Hirisplex	Walsh <i>et al.</i> , 2013, Pneuman <i>et al.</i> , 2012
KITLG	rs12821256	Chr12: 89,328,335	Hirisplex	Walsh <i>et al.</i> , 2013
EXOC2	rs4959270	Chr6: 457,748	Hirisplex	Walsh <i>et al.</i> , 2013
IRF4	rs12203592	Chr6: 396,321	8-Plex, Hirisplex	Walsh <i>et al.</i> , 2013, Pneuman <i>et al.</i> , 2012
TYR	rs1042602	Chr11: 88,911,696	Hirisplex	Walsh <i>et al.</i> , 2013
OCA2	rs1800407	Chr15: 28,230,318	Hirisplex	Walsh <i>et al.</i> , 2013
SLC24A4	rs2402130	Chr14: 92,801,203	Hirisplex	Walsh <i>et al.</i> , 2013
HERC2	rs12913832	Chr15: 28,365,618	8-Plex, Hirisplex	Walsh <i>et al.</i> , 2013, Pneuman <i>et al.</i> , 2012
ASIP/PIGU	rs2378249	Chr20: 33,218,090	Hirisplex	Walsh <i>et al.</i> , 2013
SLC24A4	rs12896399	Chr14: 92,773,663	8-Plex, Hirisplex	Walsh <i>et al.</i> , 2013, Pneuman <i>et al.</i> , 2012
TYR	rs1393350	Chr11: 89,011,046	Hirisplex	Walsh <i>et al.</i> , 2013
TYRP1	rs683	Chr9: 12,709,305	Hirisplex	Walsh <i>et al.</i> , 2013
OCA2	rs1545397	Chr15: 28,187,772	8-Plex	Pneuman <i>et al.</i> , 2012
SLC45A5	rs1426654	Chr15: 48,426,484	8-Plex	Pneuman <i>et al.</i> , 2012
ASIP	rs6119471	Chr20: 32,785,212	8-Plex	Pneuman <i>et al.</i> , 2012
MCM6	rs4988235	Chr2: 136,608,646	Tolerancia a la lactosa	Kettunen <i>et al.</i> , 2010
MCM6	rs182549	Chr2: 136,616,754	Tolerancia a la lactosa	Ye and Gu 2011

RESULTADOS

Cobertura genética e identificación del haplotipo mitocondrial

La Tabla 2 recoge los resultados de secuenciación NGS de los productos de captura mitocondrial en la muestra de Chan do Lindeiro. Un 19% de las secuencias obtenidas mapearon al genoma mitocondrial de referencia (rCRS), si bien tras la eliminación de las secuencias clonales y de aquellas que no superaron los criterios de calidad, el porcentaje

de ADN endógeno y útil se redujo al 0.41%, un porcentaje similar de otras muestras antiguas sometidas a captura mitocondrial y que se explica por el alto porcentaje de secuencias clonales que resulta esta técnica (SÁNCHEZ-QUINTO et al., 2012). Finalmente, el 99% de la secuencia del genoma mitocondrial fue cubierta con una densidad media de 58.47 lecturas por cada posición. Los sitios variables con respecto a la secuencia de referencia y con una cobertura de al menos tres lecturas, se utilizaron para la determinación del haplotipo mitocondrial en Haplogrep.

Tabla2. Descripción de los resultados NGS obtenidos a partir de la librería de Chan do Lindeiro.

Muestra	Numero de secuencias						
	Totales	Sin adaptadores	Mapeadas	Mapeadas que superan los filtros de calidad	Mapeadas no clonales	Cobertura media por sitio	% del genoma cubierto
Chan	2006422	1934347	394300	375806	8341	58.47	99.1

Las posiciones diagnóstico recogidas en la Tabla 3 fueron utilizadas para identificar el haplotipo mitocondrial en Haplogrep. En base a la comparación de estas posiciones con Phylotree se identificó el haplotipo de Elba como U5b1 con una calidad del 94.73%. Un total de 20 posiciones definen el haplotipo U5b, dos de las cuales, 4769 y 8860, faltan en

Chan do Lindeiro, ya que se trata de posiciones no cubiertas por los fragmentos resultantes de la captura. Al mismo tiempo, la variante 6713T de Chan do Lindeiro fue identificada como mutación privada por Haplogrep, lo cual significa que en la base de datos de PhyloTree ninguna muestra de perteneciente al haplotipo U5b porta dicha variante.

Tabla 3. Haplotipo mitocondrial y posiciones diagnóstico en Chan do Lindeiro.

Muestra	Haplogrupo	Posiciones diagnóstico	Otras variantes
Chan	U5b	73G, 150T, 263G, 750G, 1438G, 2706G, 3197C, 7028T, 7768G, 9477A, 11467G, 11719A, 12308G, 12372A, 13617C, 14182C, 14766T, 15326G, 16192T, 16270T	6713T

AUTENTICIDAD DE LOS RESULTADOS

La condición haploide del genoma mitocondrial ofrece la posibilidad de identificar con relativa facilidad la presencia de contaminantes entre las secuencias generadas. Concretamente, esta estima se basa en la cuantificación del número de bases secundarias, es decir, bases que difieren de la variante mayoritaria en una determinada posición. En el caso de Chan do Lindeiro, el porcentaje de contaminantes es del 2.14%, sin embargo, algunas de estas variantes pueden explicarse como el resultado de daños postmortem esperables en el ADN, concretamente las desaminaciones que llevan al cambio de citosinas por timinas y de adeninas por

guaninas. Si excluimos estas posibles transiciones, el porcentaje de bases secundarias o contaminantes disminuye al 1.7%.

El proceso de desaminación ocurre fundamentalmente en los extremos de los fragmentos de ADN, que al romperse generan bordes de cadena simple más sensibles a daños moleculares que la doble cadena. La Figura 2 generada por MapDamage (GINOLHAC et al., 2011) representa gráficamente la proporción de este tipo de daños que en los extremos 5' y 3' de las secuencias de Chan do Lindeiro estaría en torno al 10%, lo que confirma la antigüedad de las moléculas de ADN y el origen endógeno de las secuencias generadas.

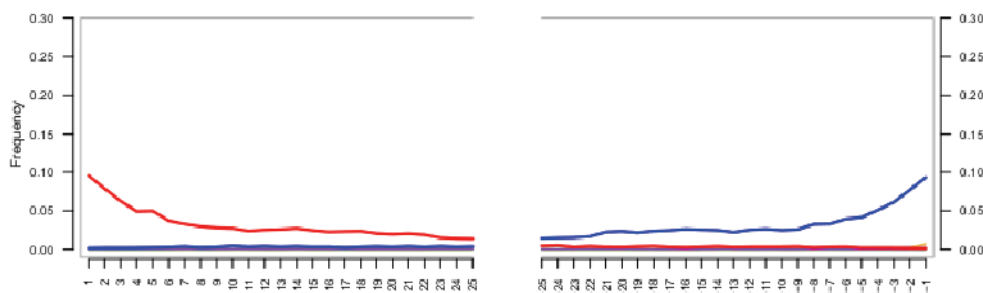


Fig 2. Representación gráfica obtenida con MapDamage (GINOLHAC et al., 2011) de los porcentaje de desaminación en los extremos de las secuencias de Chan do Lindeiro. A. Desaminación de citosinas a timinas en los extremos 5'. Desaminación de adeninas a guaninas en los extremos 3'.

Análisis filogenético del linaje mitocondrial de Elba

La Figura 3 representa gráficamente la filogenia obtenida a partir del alineamiento de la secuencia mitocondrial completa de Chan do Lindeiro junto con los mitogenomas de otros 39 individuos prehistóricos europeos. En el árbol se distinguen tres clados:

1. Un clado formado por tres individuos del paleolítico medio que fueron todos ellos

asignados al haplotipo mitocondrial M por Haplogrep. Este haplotipo M se encuentra actualmente en Asia, Australia y Sudamérica, pero no en Europa, donde solo se ha identificado en estas tres muestras antiguas (POSTH et al., 2016).

2. Un clado que agrupa la mayor parte de las muestras estudiadas, desde el Paleolítico al Neolítico y presentan todas ellas haplotipos todavía existentes en la Europa actual.

3. El individuo CS7675 es una muestra neolítica procedente de Cova de la Sarsa, en el noreste de la Península Ibérica y cuyo ADN mitocondrial fue publicado por OL-ALDE *et al.* (2015). En la información suplementaria de la publicación los autores describen una escasa cobertura del genoma mitocondrial de esta muestra, en el que faltan muchas de las posiciones definitivas

del haplotipo, de hecho, pudieron genotipar solo 12 sitios polimórficos, 10 de los cuales con una cobertura inferior a 3 lecturas, lo que podría dar lugar a errores de genotipado que explicarían su posicionamiento en la filogenia, al margen de los otros clados.

El individuo de Chan do Lindeiro se sitúa dentro del clado mayoritario, junto a individuos que fueron asignados al haplo-

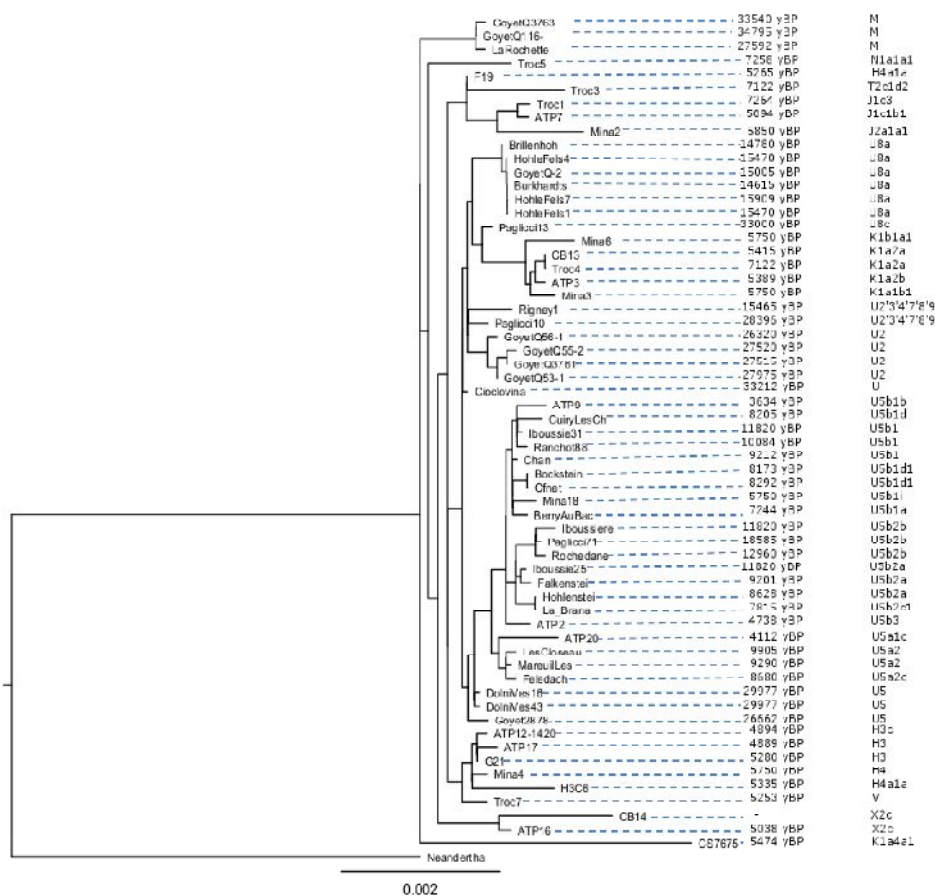


Fig 3. Analisis filogenético de los genomas mitocondriales bajo el método de Maxima verosimilitud (RaxML, STAMAKIS 2014). Elba se encuentra bajo la denominación de “Chan”. Al lado de cada individuo se incluye su datacion 14C calibrada y su asignacion haplotipica (GUNTHER *et al.*, 2015; HAAK *et al.*, 2015; OL-ALDE *et al.*, 2015; POSTH *et al.*, 2015).

tipo U5b1, si bien ninguna de las secuencias incluidas en este buscado es idéntica a la de Chan do Lindeiro. La otra muestra mesolítica ibérica (La Braña), aparece asociada a una muestra de Hohlenstein (Alemania), en un buscado hermano al de Chan do Lindeiro cuyas muestras fueron todas asignadas al haplotipo U5b2. Dos muestras Paleolíticas de Dolni Vestonice (República Checa) y una de Goyet (Bélgica) son basales a todo el buscado, las mismas muestras fueron asignadas por Haplogrep al haplogrupo U5, forma ancestral del resto de haplotipos de este grupo.

La mayoría de los individuos neolíticos incluidos en esta filogenia se agrupan en subclados compuestos exclusivamente de muestras postmesolíticas. ATP2, 9 y 20 (Burgos, España) y Mina18 (Soria, España) son la excepción a esta tendencia, ya que se agrupan en el buscado U, próximas a Chan do Lindeiro y La Braña.

Caracterización fenotípica

La cobertura de las posiciones diagnóstico para los marcadores moleculares asociados a rasgos fenotípicos fue igual o inferior a 0.1 lecturas por posición. Esta cobertura indica un escaso éxito en el proceso de captura de las secuencias nucleares e impide llegar a algún tipo de conclusión en cuanto al aspecto o características metabólicas del individuo de Chan do Lindeiro, ya que la ausencia total o baja cobertura de los SNPs de interés impide su genotipado, así como la distinción entre secuencias endógenas de la muestra de otras contaminantes. Con tal motivo, los datos genéticos que se presentan en esa monografía asociados a rasgos fenotípicos y que se han utilizado en la reconstrucción del aspecto físico de Chan do Lindeiro (SERRULLA, presente

volumen), proceden de una comunicación personal de los resultados preliminares de un estudio de mayor envergadura, en el que se está llevando a cabo la secuenciación genómica nuclear de este individuo (FORTES *et al.*, en preparación). En base a los marcadores del sistema HirisPlex y 8-Plex que se han podido recuperar en el proyecto genómico, sabemos que el individuo de Chan do Lindeiro tenía ojos y cabello oscuros y era intolerante a la lactosa, si bien no se pudo determinar con certeza la coloración de la piel debido a la ausencia de cobertura en las posiciones de los marcadores asociados a este carácter. El mismo estudio ha confirmado a nivel genético el sexo femenino de la muestra.

DISCUSIÓN

En termino genéticos y demográficos, los principales eventos que afectaron al desarrollo de las poblaciones europeas desde la primera llegada del hombre moderno a Europa, hace unos 50000 años, pueden resumirse en los siguientes episodios: la repoblación de Europa desde los refugios glaciales tras el último máximo glacial (UMG, hace aproximadamente 24000 años); la recolonización de las áreas desiertas tras el final del Dryas reciente en la era postglacial (hace unos 11500 años); la expansión del Neolítico desde el Oriente Próximo (hace aproximadamente 8000 años); y migraciones posteriores a menor escala, como las invasiones desde la estepa rusa en la edad del Bronce. Todos estos fenómenos supusieron cuellos de botella poblacionales y alteración de las frecuencias alélicas debido a la subsecuente deriva genética. La posibilidad de recuperar ADN a partir de muestras antiguas está permitiendo cuantificar de forma directa el impacto

genético y demográfico de estos eventos y al hacerlo, está llevando a revisar teorías asentadas en evidencias arqueológicas que tratan de explicar las consecuencias demográficas y culturales de estos episodios en las sociedades primitivas. Mas aun, la combinación de datos paleogenéticos y dataciones de carbono de los restos humanos esta permitiendo describir a una escala cada vez mas fina los distintos tiempos y modos de transición poblacional en las distintas áreas geográficas en Europa.

En general, los datos genéticos coinciden con los arqueológicos en la descripción de poblaciones bastante homogéneas en el periodo que va desde el UMG hasta el comienzo de la expansión neolítica. Sin embargo, poco se sabe a nivel genético de los pobladores preUMG, estudios recientes ha descubierto una variabilidad genética mayor de la esperada en muestras de esa época, con la presencia del haplotipo M, hoy día ausentes en Europa (POSTH *et al.*, 2016) y de formas ancestrales del haplotipo N (FU *et al.*, 2015). Según POSTH *et al.* (2016) esta variabilidad se habría perdido en torno al Dryas reciente (hace 14500-115000 años), en los cuellos de botella asociados a los procesos de retirada y expansión desde los refugios glaciales. De hecho, la mayoría de las muestras de cazadores-recolectores del Paleolítico superior y mesolítico europeo pertenecen a algún subtipo de un único haplogrupo, el U (BRAMANTI *et al.*, 2009; HAAK *et al.*, 2015 ; FU *et al.*, 2013; POSTH *et al.*, 2016).

El individuo de Chan do Lindeiro también pertenece a este haplogrupo U, lo que es esperable considerando su antigüedad y origen geográfico. Dentro del haplogrupo U, Chan do Lindeiro se clasifica en el subhaplotipo U5b1, cuyo origen se sitúa en la Península Ibérica hace unos 16-20000 años

(MALYARCHUK *et al.*, 2010). El otro individuo mesolítico estudiado a nivel de mitocondrial completo en España (La Braña-Arintero, León, SÁNCHEZ-QUINTO *et al.*, 2012), pertenece al haplotipo U5b2, que tiene un tiempo de coalescencia algo más antiguo, en torno a los 20-24000 años (MALYARCHUK *et al.*, 2010) y en la actualidad tiene su mayor representación en centroeuropa. Un tercer subhaplogrupo del U5b es el U5b3, cuyo origen se sitúa también en la Península Itálica (PALA *et al.*, 2009). Considerando la distribución geográfica y los tiempos de coalescencia parece que una diversificación inicial del U5b ocurrió en el sur y centro de Europa, seguida de una expansión de subclusters particulares del U5b hacia el este europeo (MALYARCHUK *et al.*, 2010). Esta diversificación inicial seguida de la segregación geográfica se ha relacionado con los cuellos de botella y efectos fundadores en torno al UMG y Dryas reciente (FU *et al.*, 2013; POSTH *et al.*, 2015). Así, Chan do Lindeiro a nivel de ADN mitocondrial sería una muestra representativa de la población de cazadores-recolectores del refugio Franco-Cantábrico y por tanto, de las poblaciones que desde esta área geográfica recolonizaron los territorios del centro y este europeo en la era postglacial a medida que los hielos se retiraban. El impacto demográfico de los portadores del haplogrupo U5 es reconocible también en la filogenia presentada por POSTH *et al.* (2016) y en la que acompaña a este estudio, ya que todas las muestras postUMG incluidas en el dataset pertenecen a dicho clado hasta la llegada del Neolítico.

El haplogrupo U5 se encuentra hoy en día en Europa con frecuencias en torno al 5-12%. La forma U5b sigue siendo la de mayor frecuencia en el área franco cantábrica

y también en las poblaciones europeas occidentales, en porcentajes que varían del 5 al 8% de la población (BRAMANTI *et al.*, 2009). Una excepción a este bajo porcentaje es la población Saami, en el norte de la Península Escandinava, donde la frecuencia poblacional del U5b alcanza el 26.5-56.8% (TAMBETS *et al.*, 2004). Tal elevado porcentaje podría ser el resultado de un aislamiento prolongado de esta población Saami, en la que habría persistido un haplotipo mitocondrial presente de manera prácticamente ubicua y en elevadas frecuencias poblacionales en la Europa mesolítica.

El análisis de ADN nuclear de muestras mesolíticas europeas confirma la afinidad genética de estos grupos de cazadores recolectores, posiblemente ligada a la condición nómada de estas poblaciones, que habría favorecido el flujo génico evitando la formación de estructuras genéticas claramente definidas por límites geográficos, al menos en la Europa occidental (SÁNCHEZ-QUINTO *et al.*, 2012; LAZARIDIS *et al.*, 2014; JONES *et al.*, 2015). El análisis de la variabilidad genética de estas muestras y su comparación con las poblaciones europeas actuales los sitúa al margen del pool genético actual, lo que descarta la existencia de continuidad genética desde el mesolítico a nuestros días, tanto en el sur (incluyendo cornisa cantábrica y País Vasco) como en el norte de Europa. La falta de continuidad parece deberse no solo a la llegada de los primeros agricultores desde Oriente Medio y al impacto demográfico que supuso la revolución neolítica (GUNTHER *et al.*, 2015, MATHIESON *et al.*, 2015), sino también a migraciones posteriores masivas que tuvieron lugar durante la edad del Bronce (HAAK *et al.*, 2015). En el caso de Chan do Lindeiro no disponemos todavía de marca-

dores genómicos suficientes para comparar su variabilidad genética con el del conjunto de muestras antiguas y poblaciones actuales, sin embargo, disponemos ya de información en cuanto a algunos factores fenotípicos relacionados con su aspecto y características metabólicas. Chan do Lindeiro no presenta las mutaciones asociadas con la tolerancia a la lactosa y tendría el cabello oscuro, coincidiendo en estos rasgos con la mayoría de las muestras mesolíticas estudiadas en Europa, incluyendo el individuo del Paleolítico superior de Bichon, Suiza (JONES *et al.*, 2015) y La Braña (OLALDE *et al.*, 2014). Sin embargo, al contrario de La Braña y en común con Bichon, Chan do Lindeiro tendría ojos de coloración oscura, siendo una de las pocas muestras en Europa que testimonian la presencia de los polimorfismos asociados con esta tonalidad al menos hace ya 9000 años.

Los datos aportados por nuestro análisis genético de Chan do Lindeiro contribuyen a confirmar la prevalencia del haplotipo U5b en la Europa Occidental, posiblemente asociada al cuello de botella que siguió al Dryas reciente y a ampliar nuestro conocimiento de los rasgos físicos y metabólicos de estas poblaciones, confirmando la homogeneidad de las mismas en cuanto a su poca capacidad para digerir lácteos -carácter que se habría desarrollado más adelante, con la llegada de la agricultura-, y constatando una mayor variabilidad genética de la descrita hasta la fecha en cuanto a pigmentación ocular. Nuestro estudio incrementa en un 50% el número de genomas mitocondriales completos actualmente disponibles en muestras mesolíticas de la Península Ibérica y extiende el rango de antigüedad en 2000 años, confirmando la continuidad genética de las poblaciones mesolíticas de la Península Ibérica en

ese intervalo temporal, al menos en lo que se refiere a los linajes maternos.

AGRADECIMIENTOS

Este trabajo se ha realizado gracias a la colaboración del Grupo Espeleológico A Trapa (Vigo) con financiación de la Xunta de Galicia (Programa de Consolidación e Estructuración Redes, R2014/024)

REFERENCIAS BIBLIOGRÁFICAS

- BRAMANTI, B., THOMAS, M.G., HAAK, W., UNTERLAENDER, M., JORES, P., TAMBETS, K., ANTANAITIS-JACOBS, I., HAIDLE, M.N., JANKAUSKAS, R., KIND, C.J., et al. 2009. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326, 137–140.
- BRIGGS, A.W., STENZEL, U., JOHNSON, P.L., GREEN, R.E., KELSO, J., PRUFER, K., MEYER, M., KRAUSE, J., RONAN, M.T., LACHMANN, M., y PÄÄBO, S. 2007. Patterns of damage in genomic DNA sequences from a Neanderthal. *Proc. Natl. Acad. Sci. USA* 104, 14616–14621.
- FU, Q., MITTNIK, A., JOHNSON, P.L., BOS, K., LARI, M., BOLLONGINO, R., SUN, C., GIEMSCH, L., SCHMITZ, R., BURGER, J., et al. 2013. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23, 553–559.
- FU, Q., HAJDINJAK, M., MOLDOVAN, O.T., CONSTANTIN, S., MALLICK, S., SKOGLUND, P., PATTERSON, N., ROHLAND, N., LAZARIDIS, I., NICKEL, B., et al. 2015. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524, 216–219.
- GAMBA, C., JONES, E.R., TEASDALE, M.D., MCLAUGHLIN, R.L., GONZALEZ-FORTES, G., MATTIANGELI, V., DOMBORÓCZKI, L., KŐVÁRI, I., PAP, I., ANDERS, A., WHITTLE, A., DANI, J., RACZKY, P., HIGHAM, T.F., HOFREITER, M., BRADLEY, D.G., y PINHASI, R. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature Commun.* 5, 5257.
- GINOLHAC, A., RASMUSSEN, M., GILBERT, M.T. et al (2011) mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* 27(15), 2153-2155.
- GÜNTHER, T., VALDIOSERA, C., MALMSTRÖM, H. et al. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl. Acad. Sci. USA* 112(38), 11917–11922.
- 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, 207.
- HODGES, E., ROOKS, M., XUAN, Z., et al. 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature Protocols* 4, 960–974.
- JONES, E.R., GONZALEZ-FORTES, G., CONNELL, S. et al 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature Commun.* 6, 8912. DOI: 10.1038/ncomms99
- KETTUNEN, J., SILANDER, K., SAARELA, O., et al 2010. European lactase persistence genotype shows evi-

- dence of association with increase in body mass index. *Hum. Mol. Genet.* 19(6), 1129-1136.
- KRAUSE, J., BRIGGS, A.W., KIRCHER, M., MARICIC, T., ZWYNS, N., DERVIANKO, A., y PÄÄBO, S. 2010. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* 20, 231–236.
- LAZARIDIS, I., PATTERSON, N., MITTNIK, A., RENAUD, G., MALLICK, S., KIRSANOW, K., SUDMANT, P.H., SCHRAIBER, J.G., CASTELLANO, S., LIPSON, M., et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409–413.
- LAZARIDIS, I., NADEL, D., ROLLEFSON, G. et al 2016. The genetic structure of the world's first farmers. *Nature* 536, 419–424.
- LI, H. y DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 25, 1754-1760.
- MALYARCHUK, B., DERENKO, M., GRZYBOWSKI, T., PERKOVA, M., ROGALLA, U., VANECEK, T. y TSYBOVSKY, I. 2010. The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS One* 5(4), e10285.
- MARICIC, T., WHITTEN, M., y PÄÄBO, S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5(11), e14004.
- MARTIN, M. 2011. Cutadapt removes adapter sequences from high throughput sequencing reads. *EMBnet.journal* 17(1), 10-11.
- MATHIESON I., LAZARIDIS I., ROHLAND N. et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528, 499–503.
- MCKENNA, A., HANNA, M., BANKS, E., et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- MEYER, M., y KIRCHER, M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc.* doi:10.1101/pdb.prot5448.
- MILNE, I., STEPHEN, G., BAYER, M., et al. 2013. Using Tablet for visual exploration of secondgeneration sequencing data. *Briefings in Bioinformatics* 14(2), 193-202.
- OLALDE, I., ALLENTOFT, M.E. y SANCHEZ-QUINTO, F. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. 2014. *Nature* 507(7491), 225-228.
- OLALDE, I., SCHROEDER, H., SANDOVAL-VELASCO, M. et al. 2015. A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Mol. Biol. Evol.* 32(12), 3132–3142.
- ORLANDO, L., GINOLHAC, A., RAGHAVAN, M., VILSTRUP, J., RASMUSSEN, M., MAGNUSSEN, K., STEINMANN, K.E., KAPRANOV, P., THOMPSON, J.F., ZAZULA, G., et al. 2011. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Res.* 21, 1705–1719.
- PALA, M., ACHILLI, A., OLIVIERI, A., KASHANI, B., PEREGO, U., et al. 2009. Mitochondrial haplogroup U5b3: A distant echo of the Epipaleolithic in Italy and the legacy of the early Sardinians. *Am. J. Hum. Genet.* 84, 814–821.

- PINHASI, R., THOMAS, M.G., HOFREITER, M., CURRAT, M. y BURGER, J. 2012. The genetic history of Europeans. *Trends Genet.* 28(10), 496-505.
- PNEUMAN, A., BUDIMLIJA, Z.M., CARAGINE, T., et al. 2012. Verification of eye and skin color predictors in various populations. *Leg. Med.* 14, 78–83.
- POSTH, C., RENAUD, G., MITTNIK, A. et al 2016. Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Current Biology* 26, 827–833.
- ROHLAND, N., SIEDEL, H., HOFREITER, M. 2010. A rapid column-based ancient DNA extraction method for increased sample throughput. *Mol. Ecol. Resour.* 10(4), 677-683.
- SANCHEZ-QUINTO, F., SCHROEDER, H., RAMIREZ, O. et al. 2012. Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. *Current Biology* 22, 1494–1499.
- SPICHENOK, O., BUDIMLIJA, Z.M., MITCHELL, A.A., et al. 2011. Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Sci. Int. Genet.* 5, 472–478.
- STAMATAKIS, A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* 30(9), 1312-1323.
- TAMBETS, K., ROOTSI, S., KIVISILD, T., HELP, H., SERK, P., LOOGVALI, E.L., TOLK, H.V., REIDLA, M., METSPALU, E., PLISS, L., et al. 2004. The western and eastern roots of the Saami—the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. *Am. J. Hum. Genet.* 74, 661–682.
- THORVALDSDÓTTIR, H., ROBINSON, J.T., MESIROV, J.P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in *Bioinformatics* 14, 178-192.
- WALSH, S., LIU, F., WOLLSTEIN, A. et al. 2013. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* 7, 98–115.
- YE, K., GU, Z. 2011. Recent advances in understanding the role of nutrition in human genome evolution. *Adv. Nutr.* 2(6), 486-496.