From the Mesolithic to the Bronze Age: unraveling 5,000 years of European population history with paleogenomics

Íñigo Olalde Marquínez

TESI DOCTORAL UPF / 2016

DIRECTOR DE LA TESI

Dr. Carles Lalueza-Fox

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

Universitat Pompeu Fabra Barcelona

A mis padres
A Neus

"*It has often and confidently been asserted, that man's origin can never be known: but ignorance more frequently begets confidence than does knowledge: it is those who know little, and not those who know much, who so positively assert that this or that problem will never be solved by science*"

Charles R. Darwin, 1871. *The Descent of Man*

Acknowledgments

Me gustaría aprovechar estas líneas para agradecer a todas las personas que me han acompañado, tanto en lo profesional como en lo personal, a lo largo de esta etapa que está a punto de terminar. También quiero recordar a personas que formaron parte importante de mi vida en épocas anteriores. Al fin y al cabo, esta tesis doctoral es la culminación de muchos años de esfuerzo y vivencias compartidas que dejan lazos que la distancia difícilmente puede romper.

En primer lugar quiero agradecer a los arqueólogos que nos han cedido las muestras aquí analizadas. Gracias por vuestro imprescindible trabajo, sin el cual esta tesis nunca habría sido posible. En esta sección merecen una mención los aficionados a la espeleología. Sin ellos, La Braña 1 y La Braña 2 seguirían escondidos en el fondo de una cueva para siempre y no habríamos podido aprender todo lo que hemos aprendido. También, por qué no, gracias a todas las personas (sí, algún día fueron personas) que desinteresadamente e involuntariamente nos cedieron su ADN, y que nunca podrían imaginar que miles de años después llegáramos a conocer casi más de ellos que ellos mismos.

Thanks to Tom Gilbert and Hannes Schroeder for hosting me in Copenhagen and for our fruitful collaborations. It has been a pleasure to work with you. I would also like to thank David Reich for being a good host and for letting me look at the amazing dataset they are putting together with much effort. Thanks also to all the people in Reich's lab for making my time in Boston an incredible experience.

It was very inspiring to see great scientists being so passionate about science.

Muchas gracias a todas las personas que han formado y forman parte del IBE, que siempre será mi "hogar científico". Desde el primer día que llegué me di cuenta que era el sitio perfecto para empezar en el mundo de la investigación, gracias por todo el conocimiento científico transmitido. También, gracias por el buen ambiente, por las fiestas, por dejarme hacer de dj en el retreat y por los buenos momentos jugando volley. Os voy a echar mucho de menos.

Muchas gracias a Juan, Javi, Gabriel, Fede, Javi Quílez, Urko, Arcadi y Tomàs por vuestra ayuda con el paper de La Braña. A Irene Lobón por tu ayuda con el paper de Cova Bonica en pleno Agosto. A Lara por ayudarme con los PCAs. A Marcos por enseñarme bioinformática. También a Emiliano, Anna y Rita por gestionar de manera tan eficiente la burocracia. A Sergi Civit por tu ayuda con la estadística y por tu simpatía.

Me gustaría agradecer a mis compañeros del grupo de paleogenómica, pasados y presentes. Muchas gracias a Óscar, por enseñarme tanto y por integrarme en bioevo. Eres un tío genial y espero que todo te vaya muy bien. Como no, gracias a Fede por estar siempre dispuesto a ayudar y por ser un buen compañero, pero sobre todo por organizar aquel Crazy Friday. Muchísimas gracias a Carles, el ideólogo detrás de esta tesis doctoral. Muchas veces a lo largo de nuestra vida, los caminos que vamos tomando dependen de decisiones concretas de terceras personas. No sé si te acordarás, pero cuando te mandé el email para hacer las prácticas del máster en tu

laboratorio, me respondiste que no estabas buscando a gente en ese momento. Al día siguiente, tras leer mi expediente, me mandaste otro email diciendo que me pasara por el PRBB, que ya buscaríamos algo. Esta tesis demuestra que mandar ese segundo email fue una buena decisión, sobre todo para mi, tanto en lo profesional como en lo personal. Gracias por ser un buen mentor, por apretar cuando había que apretar, y por ser flexible cuanto tocaba ser flexible. Gracias también por tu confianza al ponerme al frente de proyectos tan ambiciosos, cuando ni yo mismo tenía claro si podía llevarlos a cabo.

En un plano más personal, gracias a la gente del PRBB. A Joan Pau por invitarme a jugar en el Camprodon team. Gracias a los Bonasters, sois unos cracks. A Hernani por abrirme las puertas de su casa en Copenhague. A Haizea, Xeli, Juanjo y David por las escapadas de fin de semana que tanto he disfrutado, es genial teneros como amigos, y en general a toda la buena gente que conocí durante el Erasmus: Elena, María, Gökhan, Meltem, Laris, Hassan…

A pesar de que apenas hemos podido vernos estos años, gracias a Arrocha, Imanol, Iraitz y Corti. Vosotros formáis parte de una de las mejores épocas de mi vida y se os echa de menos. También a las navarricas Eukene, Maru, Nerea y Amaya. Gracias a mis amigos de Vitoria por todos los buenos ratos juntos. En especial a Alain y Enjuto por el tiempo que vivimos en carrer Salvà, qué bien lo pasamos!! También a Alberto por enseñarnos Panamá y por más de 25 años de amistad.

Moltíssimes gràcies a la meva família vilanovina, per acollir-me i fer-me sentir un més des del primer dia. Significa molt per a mi i sempre

us ho agrairé. Gracias a Pilar por tratarme como a un nieto y permitirme disfrutar, a 500 kilómetros de Vitoria, de todas las ventajas de tener una abuela.

Como no, me gustaría expresar mi más sincero agradecimiento a mi familia, en Tenerife y en Vitoria. Puede que sea porque ya llevo 10 años fuera de casa, pero no os podéis imaginar lo mucho que he disfrutado todos los momentos juntos. A mis tíos: Tito, Bego, Toño, Itxaso, Blanca, Pedro, Rafa y Pilar. A mis primos: Rubén, Zuri, Susi, Alberto, Javi, Elena, David y Alberto. Os quiero mucho, sois la mejor familia que alguien puede tener. Mención destacada merece mi primo/hermano Javi. Gracias por tu complicidad y por las noches jugando a la play hasta las 4 de la mañana que tanto me han ayudado a desconectar. Aunque estemos lejos, siempre voy a estar ahí para lo que necesites. A mis abuelos, de los que tanto sigo aprendiendo. A mi abuela Conchi, gracias por ser entrañable y por tu alegría al verme entrar por la puerta siempre que voy a verte. A mi abuelo Honorio, cuyo recuerdo sigue y seguirá intacto, por ese carácter especial que ha dejado huella y por ser un ejemplo de honestidad y lealtad. Muchas gracias de corazón a mi abuela Jose, por tu bondad infinita y por darme en media hora todo el cariño que alguien puede necesitar en meses.

Esta tesis está especialmente dedicada a mis padres, Honorio y Lurdes. Todo lo que soy es gracias a vosotros, así que todos mis logros son vuestros. Gracias por darme todo lo necesario para ser feliz. Gracias, también, por vuestro apoyo incondicional, que me ha permitido hacer siempre lo que más me apetecía en cada momento, y gracias por hacer de mi educación vuestra prioridad. Es por eso que

hoy estoy escribiendo estas líneas. Me resulta muy difícil expresar con palabras toda mi gratitud, y espero al menos poder devolveros en forma de cariño y admiración parte de todo lo que me habéis dado.

Por último pero no menos importante, sino al contrario, es imprescindible dedicarte unas palabras a ti, Neus, mi compañera de viaje durante estos 4 años. Para mí, el doctorado siempre estará lleno de buenos recuerdos, en gran parte porque está asociado al tiempo en el que nos conocimos. Tengo que pedirte perdón y agradecerte a la vez. Perdón por las vacaciones estropeadas por culpa de papers que enviar, y en general por el tiempo que te mereces y que no te he podido dedicar. Espero que en un futuro estos sacrificios se vean recompensados de alguna manera. Gracias por darme la estabilidad necesaria en un período con tantos altibajos, por tu comprensión y ayuda constante, pero sobre todo por ser una persona maravillosa que hace feliz a todos lo que te rodean. No sé en qué lugar acabaremos los próximos años, pero no me importa nada porque contigo iría al fin del mundo.

Íñigo Olalde Marquínez

This research was supported by a PhD fellowship from the Basque Government “Programa Predoctoral de Formación de Personal Investigador No Doctor del Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco”.

Abstract

The study of genetic information recovered from ancient remains is a powerful tool for learning about human population history. Until recently, ancient DNA works have been hampered by technical limitations related to DNA preservation and have largely relied on single genetic loci. With the advent of high-throughput sequencing technologies, most of these limitations have been overcome, making it feasible to study millions of loci at ancient population scales. This has completely transformed the field, by greatly enhancing our ability to detect genetic signals and characterize past processes. Within the context of this paleogenomic revolution, I present genome-wide data from 55 ancient Europeans, including remarkable samples such as La Braña 1 Mesolithic forager and Cova Bonica Cardial farmer. We use this information to gather new evidence about the mechanisms underlying the transition to farming, which radically changed the fate of human societies, and the origin and expansion of the Late Neolithic Bell Beaker phenomenon. In addition, we provide valuable insights into the timing of recent adaptations.

Resumen

El estudio del material genético recuperado de restos antiguos es una poderosa herramienta para investigar la historia demográfica de las poblaciones humanas. Hasta hace muy poco, los trabajos de ADN antiguo estaban seriamente afectados por limitaciones técnicas relacionadas con la degradación del ADN, lo que restringía los análisis a marcadores uniparentales. Con el desarrollo de las tecnologías de secuenciación masiva, la mayoría de estas limitaciones han sido superadas, haciendo posible el estudio de millones de marcadores genéticos en cientos de individuos antiguos. Gracias a estos avances, el campo del ADN antiguo ha sufrido una completa transformación, reflejada en una capacidad sin precedentes para caracterizar acontecimientos pasados. En el marco de esta revolución paleogenómica, presento en esta tesis datos genómicos de 55 individuos europeos antiguos, entre los que se incluyen muestras emblemáticas como el hombre mesolítico de La Braña y la mujer neolítica de Cova Bonica. Esta valiosa información nos ha permitido desarrollar una visión mucho más completa y robusta sobre la transición al Neolítico, que cambió radicalmente el destino de las sociedades humanas, y sobre el origen y expansión del fenómeno campaniforme. Además, hemos profundizado en el conocimiento de adaptaciones recientes en poblaciones europeas.

Preface

The study of human prehistory has been traditionally addressed from the field of archaeology. More than 100 years of excavations have enlightened our knowledge of the past, setting the ground for the development of other disciplines. However, despite the abundance of archeological data in areas like Western Europe, many important events remain controversial, and the use of complementary approaches is therefore necessary.

Since the pioneering work of Cavalli-Sforza and colleagues almost 40 years ago, population geneticists have studied the patterns of genetic variability in present-day populations to make inferences about past demographic and selective processes. Recognizing its validity, this strategy poses inherent limitations. Our genomes have been shaped by a large number of both deep and recent evolutionary events, whose signatures are often difficult to disentangle.

Within this context, the relevance of ancient DNA research, which provides snapshots of past genetic diversity at different time points, becomes evident. Recently, thanks to the availability of high-throughput sequencing technologies, the recovery of hundreds of ancient genomes is routinely achieved, allowing the study of human population history with unprecedented resolution.

The work I present here is part of this ongoing paleogenomic revolution, and has made a valuable contribution to its development. It focuses on the recent (Holocene) human evolutionary time in

Europe, a period that witnessed some of the most transformative demographic and cultural changes in the history of our species.

Index

Abbreviations

A: Adenine

aDNA: Ancient DNA

AMH: Anatomically modern human

bp: Base pairs

BP: Years before present

C: Cytosine

cal BCE: Calibrated years before the common era

DNA: Deoxyribonucleic acid

G: Guanine

Gb: Gigabases

kb: Kilobases

kya: Thousand years ago

LGM: Last glacial maximum

Mb: Megabases

mtDNA: Mitochondrial DNA

RNA: Ribonucleic acid

SGS: Second-generation sequencing

SNP: Single nucleotide polymorphism

PCR: Polymerase chain reaction

T: Thymine

TGS: Third-generation sequencing

U: Uracil

1. INTRODUCTION

In the following sections, I will describe the technological breakthrough that has made this thesis possible (section 1.1), the field of genetics to which the work I hereby present belongs (section 1.2) and the period of human history under study (section 1.3).

1.1. Second generation sequencing

In 1944, Avery and colleagues (Avery et al. 1944) identified DNA as the hereditary material, whose double helical strand structure was deciphered by Watson and Crick in 1953 (Watson and Crick 1953). Only 60 years later, the genome of thousands of organisms has been sequenced, transforming nearly every field of biological research. How this happened is a remarkable example of scientific and technological innovation.

In 1977, Sanger and colleages (Sanger et al. 1977) and Maxam and Gilbert (Maxam and Gilbert 1977) developed two different techniques to sequence DNA. Due to its high efficiency and less handling of toxic radioisotopes, the method developed by Sanger and colleages and based on chain termination became the primary technology of the so-called “first generation”. Sanger sequencing remained as the prevailing method for the next 30 years, but increasing needs for automation and higher accuracy led to important improvements, like the introduction of automatic sequencing machines by Applied Biosystems, which used the same basic Sanger

method. These advances led to the completion of the first human genome sequence by the Human Genome Project in 2004 (International Human Genome Sequencing Consortium 2004), after 10 years and at a cost of $3 billion (Van Dijk et al. 2014).

The low throughput, low speed and high cost of first-generation methods stimulated the development and commercialization of second-generation sequencing technologies (SGS). This was accelerated by several funding programs, such as the one launched by the National Human Genome Research Institute (NHGRI) with the goal of reducing the cost of human genome sequencing to $1,000 (Schloss 2008).

In these SGS methods, bacterial cloning of DNA fragments is not needed and libraries can be prepared in a cell free system. For each original DNA template, tens of thousands of identical strands are attached to a given location and read in a process consisting of successive washing and scanning operations. The sequencing process involves sequentially flooding in reagents, such as labeled nucleotides, incorporating nucleotides into the DNA strands, stopping the incorporation reaction, washing out the excess reagent, scanning to identify the incorporated bases and finally treating the newly incorporated bases to prepare the DNA templates for the next ‘wash-and-scan’ cycle (Metzker 2010). Given that DNA anchor surfaces can have very high density of DNA fragments, millions of sequencing reactions can be produced in parallel, as compared to few hundreds in first-generation methods, resulting in very high overall throughput and low cost per identified base. Furthermore, base

interrogation is performed cyclically and thus no electrophoresis is required.

The first commercially successful SGS technology was the Roche 454 pyrosequencing method, released in 2005. One year later, Solexa (later purchased by Illumina) sequencing platform was commercialized, followed by SOLiD technology in 2007 and Ion Torrent in 2010 (Liu et al. 2012). These platforms represent the major SGS technologies, and have introduced several improvements in sequencing chemistry and base-calling software since the original releases. In general, 454 produces longer reads (up to 1 kb) but with low throughput and high reagent cost, whereas SOLiD generates very short reads (up to 75 bases) but with low error rates and high throughput (Van Dijk et al. 2014). Ion Torrent is the only system with semi-conductor technology, which results in higher speed and smaller instrument size. Illumina, currently the leader in the SGS industry, offers the highest throughput and the lowest per-base cost, with read lengths up to 300 bp.

Over the last decade, SGS technologies have revolutionized the field of genomics by drastically decreasing the prize of sequencing an average human genome, which has now reached the long-awaited $1,000 threshold (Figure 1). As a consequence, sequencing entire genomes is now possible for small laboratories, but at the same time the development of bioinformatic tools is required to manage the huge amount of data generated. The range of SGS application is rapidly growing and includes, among others, large-scale human genetic variation studies, e.g. the 1000 Genomes Project (The 1000

Genomes Project Consortium 2015), RNA-sequencing, DNA methylation studies, metagenomics, disease gene identification, forensic genetics, agrigenomics and paleogenomics, with this thesis as a clear example.

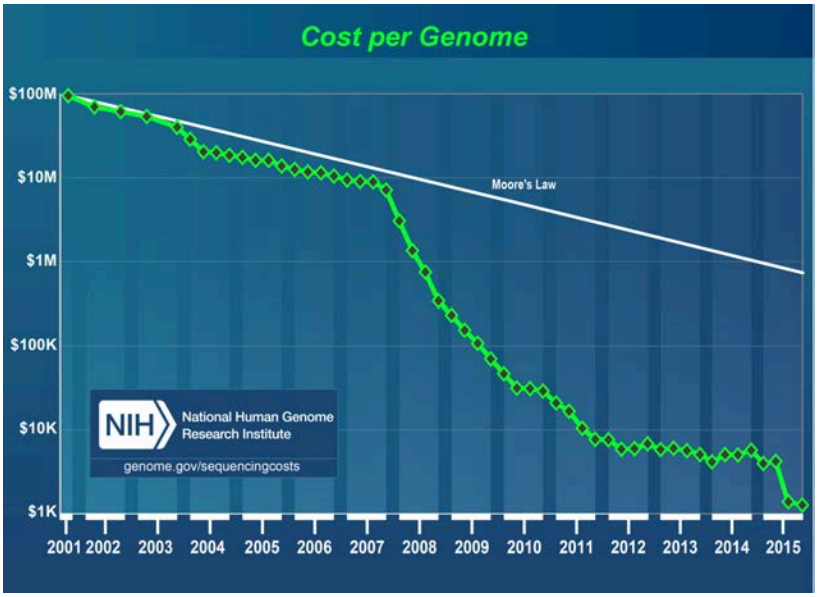


Figure 1. Prize of sequencing one human genome.
https://www.genome.gov/sequencingcosts/

In spite of the recent adoption of SGS technologies, a new generation of methods, called third-generation sequencing (TGS), is being developed in order to overcome the limitations of current technologies. Some of them are already available on the market, like the Pacific Biosciences (PacBio) technology, which is based on the direct observation of a single molecule of DNA polymerase as it synthesizes a strand of DNA (Schadt et al. 2010). Therefore, biases introduced during the PCR amplification step are avoided. More importantly, this method does not halt the sequencing reaction after each base incorporation, exploiting the high catalytic potential of

DNA polymerase. This results in faster run times and extremely long reads of about 20 kb. It remains to be explored, however, how TGS methods could be applied to the sequencing of ancient genomes.

1.2. Ancient DNA

Ancient DNA (aDNA) is the genetic material extracted from degraded biological samples, including preserved bones, teeth, hair, seeds, or other tissues. After the dead of an organism, DNA repair mechanisms cease functioning and DNA immediately begins to be digested by nucleases, either from the organism itself or from infiltrated microorganisms like bacteria or fungi. The rate of decay is determined by the environment (Lindahl 1993; Smith et al. 2003), with cold, dry and low-radiation conditions favoring DNA preservation (Figure 2). Under ideal conditions, DNA is not expected to survive more than one million years. In fact, the oldest aDNA recovered to date is that of a Pleistocene horse found in permafrost and dated to 560–780 kya (Orlando et al. 2013).

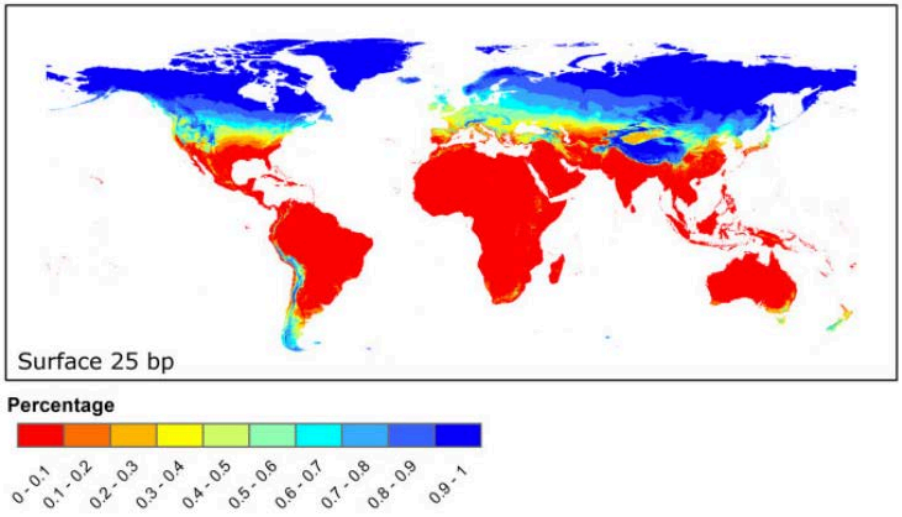


Figure 2. Expected survival of DNA at the surface after 10,000 years for 25 bp fragments. (Hofreiter et al. 2014)

However, the retrieval of ancient DNA is not temporally-bound, as a specimen's age is not linearly correlated to the amount of surviving DNA, which is characteristic of the inherent variability of DNA degradation and percentages of surviving DNA across specimens and spatio-temporal contexts (Marciniak et al. 2015).

Post-mortem molecular degradation processes can be broadly categorized as: (i) lesions that reduce the size of DNA, such as single- and double-strand breaks, (ii) Blocking lesions that obstruct the movement of DNA polymerases along a template strand preventing DNA replication, such as crosslinks and oxidation products of pyrimidines (Pääbo 1989; Hoss et al. 1996), and (iii) Miscoding lesions, such as cytosine deamination, that can be amplified using the polymerase chain reaction (PCR), but will cause incorrect nucleotides to be incorporated when DNA is replicated (Hofreiter et al. 2001a).

These different types of DNA damage mechanisms continuously acting on aDNA sequences result in characteristic patterns observed in aDNA extracts. Understanding and quantifying these patterns are crucial steps for aDNA research, since they greatly impact both laboratory procedures and bioinformatics pipelines. At the same time, molecular damage is useful because it allows us to distinguish and even separate real aDNA molecules from contaminant sequences, which should not display such patterns if they originate from a modern source.

1.2.1. Characteristic features of aDNA extracts

a) Fragmentation

Since the very early days of aDNA research it was clear that most recoverable DNA fragments from ancient specimens were very short (Pääbo 1989).

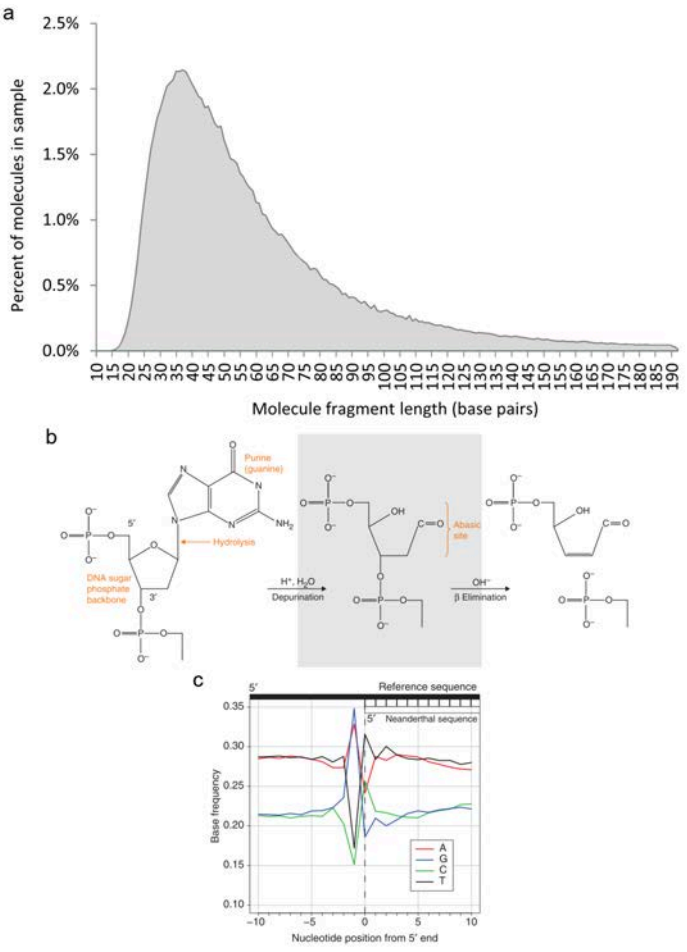


Figure 3. DNA Fragmentation pattern. a) Fragment length distribution of a permafrost-preserved woolly mammoth mummy. Modified from Marciniak et al. (2015). **b)** Depurination as a likely cause of DNA fragmentation. (Dabney et al. 2013b). **c)** Excess of A and G upstream of the 5' end of aDNA sequences, supporting that depurination is a major mechanism of DNA fragmentation. (Dabney et al. 2013b)

This has been extensively confirmed with SGS technologies that allow a better characterization of the length distribution of aDNA extracts (Green et al. 2010). Although the degree of fragmentation depends on environmental conditions, the average fragment length in ancient samples is usually below 70 bp (Figure 3a). One possible mechanisms for DNA fragmentation is depurination (Lindahl 1993), in which the *N*-glycosyl bond between a sugar and an adenine or guanine residue is cleaved, resulting in an abasic site. The DNA strand is then fragmented through β elimination, leaving 30-aldehydic and 50-phosphate ends (Figure 3b). Using data from SGS experiments, it is possible to study the frequency of each base at each position of aDNA sequences, but also the frequencies at genomic positions upstream of the 5' end and downstream of the 3' end of the fragments, after aligning those sequences to a reference genome. Consistent with depurination driving post-mortem DNA fragmentation, it has been shown (Briggs et al. 2007) that purines are overrepresented at genomic positions preceding read starts (corresponding to abasic sites in aDNA molecules) (Figure 3c).

b) Increased rate of deamination at the ends of aDNA molecules

The most common miscoding lesion in aDNA is the hydrolytic deamination of cytosine to uracil (Gilbert et al. 2007) (Figure 4a), which directs the incorporation of adenine during DNA replication. Importantly, it has been shown that these substitutions primarily occur close to fragment ends, where up to 40% of cytosines exhibit the signature of deamination, in contrast to only 2% of internal cytosines showing such misincorporations (Briggs et al. 2007; Brotherton et al. 2007).

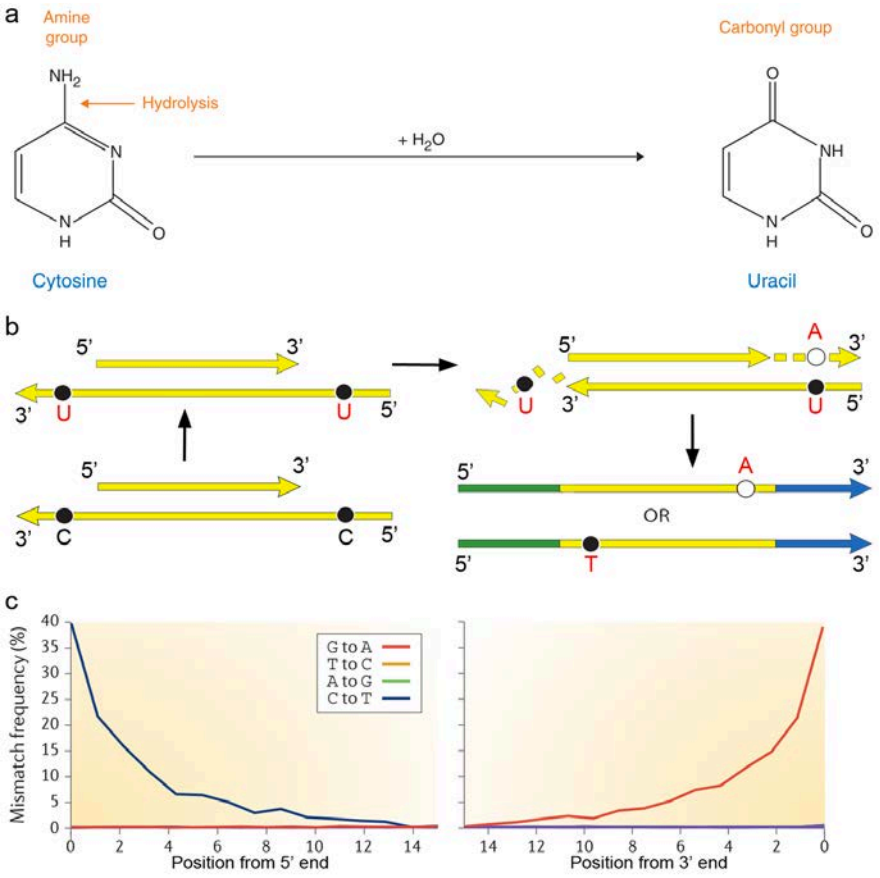


Figure 4. Deamination pattern in aDNA. a) Hydrolytic deamination of cytosine to uracil. (Dabney et al. 2013b). **b)** Blunt end repair with T4 DNA polymerase removes 3' overhangs and fills in 5' overhangs. Modified from Briggs et al. (2007). **c)** High rate of C to T substitutions at 5' ends and G to A substitutions at 3' ends. Modified from Stoneking and Krause (2011).

This can be explained by the occurrence of single-stranded overhanging ends, where the rate of cytosine deamination is about two orders of magnitude faster than in double-stranded DNA (Lindahl 1993).

Depending on the protocol used for library preparation, deaminated cytosines can produce different signals in SGS reads. Most methods for creating double stranded SGS libraries involve blunt-end repair

by T4 DNA polymerase (Briggs et al. 2007), which removes overhanging 3’ ends and fills in overhanging 5’ ends. As a consequence, uracils on overhanging 3’ ends are degraded, whereas uracils on overhanging 5’ ends serve as a template during the 3’ fill-in step (Figure 4b), resulting in high rates of C to T changes at 5’ ends and high rates of apparent G to A changes at 3’ ends observed in SGS reads (Figure 4c). The 3’ end in SGS reads therefore presents the reverse complementary pattern to the 5’ end but both are caused by deaminated cytosines at the 5’ end of original aDNA molecules.

c) Percentage of endogenous DNA

DNA extracted from ancient remains is a mixture of endogenous (from the target organism) and exogenous DNA, which mainly derives from plants, fungi, bacteria and other microorganisms that colonize the sample post mortem. Another source of exogenous DNA is contamination, which occurs when DNA from present-day environment is introduced into the sample at any stage of excavation or laboratory analysis. The percentage of endogenous DNA in ancient samples can vary dramatically, but is usually very low (rarely exceeding 5%) (Green et al. 2010; Reich et al. 2010; Carpenter et al. 2013). Notable exceptions include remains preserved under cold conditions, such as Ötzi the Iceman with 78% endogenous (Keller et al. 2012), the Saqqaq Paleo-Eskimo with 84% (Rasmussen et al. 2010) or woolly mammoths found in permafrost with 80% (Miller et al. 2008).

The low percentage of endogenous DNA material has important implications for aDNA experiments. In order to obtain significant

genetic information from the target individual, large amounts of sequencing data must be produced. For instance, to obtain the draft Neandertal genome 1.5 billion reads were generated from samples with less than 5% endogenous (Green et al. 2010). As a consequence, accessing high coverage genomes is prohibitive for most laboratories even with falling sequencing costs. To overcome this issue, several targeted capture approaches that enrich the endogenous content of SGS libraries have been developed (Carpenter et al. 2013; Fu et al. 2013a; Castellano et al. 2014; Gansauge and Meyer 2014; Haak et al. 2015; Orlando et al. 2015). Some of these methods have been applied in the works presented in this thesis and will be reviewed in section 2.1.3.

1.2.2. Introduction to the field of ancient DNA

In 1984, aDNA research began with the publication of DNA sequences from the quagga (Higuchi et al. 1984), an equid that became extinct in the 19th century, and shortly after from an Egyptian human mummy (Paabo 1985). These studies used bacterial cloning to amplify small DNA sequences retrieved from skin fragments. While the validity of the first study has been confirmed (Vilstrup et al. 2013), the second is widely accepted to be the result of contamination.

The development of PCR, which allows to target and replicate specific DNA sequences (Mullis and Faloona 1987), together with the finding that DNA survives in calcified material (Hagelberg et al. 1989), resulted in a rapid increase and diversification of aDNA

studies. These works relied on the amplification of specific sequences with PCR, followed by traditional Sanger sequencing, either directly or after cloning the PCR product into a vector. However, the degradation of aDNA molecules coupled with the sensitivity of PCR to contamination contributed to the publication of the so-called "antediluvian" DNA sequences from Miocene plants (Golenberg et al. 1990; Soltis et al. 1992) and Cretaceous dinosaur bones (Woodward et al. 1994), either proven impossible to reproduce or shown to derive from an identifiable source of contamination (Austin et al. 1997; Hebsgaard et al. 2005). These problems seriously undermined the field's reliability, and the aDNA community adopted a strict list of criteria for authenticating data and ensuring reliability (Cooper and Poinar 2000).

Despite inherent limitations, ancient DNA sequences obtained via PCR have allowed valuable phylogenetic studies of both extinct and extant species. PCR-based approaches have been successfully used to reconstruct the mtDNA phylogenetic relationships between living species and their extinct relatives, such us Balearic Islands cave goats (Ramírez et al. 2009), giant lemurs (Orlando et al. 2008) or New Zealand moas (Cooper et al. 2001). Other studies have tackled the genetic history of domestication, being cattle and pigs the most extensively studied species (Bailey et al. 1996; Larson et al. 2007). The study of hominin evolution has also benefit from PCR amplification of target DNA fragments. In 1997, the publication of the first mtDNA sequences from a Neandertal (Krings et al. 1997) showed that they fell outside the variation of modern humans. This was confirmed by additional mtDNA sequences from other

Neandertal specimens distributed all over Europe (Krings et al. 1999; Krings et al. 2000; Ovchinnikov et al. 2000; Lalueza-Fox et al. 2005; Caramelli et al. 2006; Lalueza-Fox et al. 2006; Orlando et al. 2006). Importantly, the recent human population history of Europe, which is the focus of this thesis, has received great attention, specially the Mesolithic-Neolithic transition (Sampietro et al. 2007; Bramanti et al. 2009; Malmström et al. 2009; Haak et al. 2010; Brandt et al. 2013).

Although some PCR-based aDNA works have analyzed nuclear regions (Krause et al. 2007; Lalueza-Fox et al. 2007; Lalueza-Fox et al. 2008; Ludwig et al. 2009), the vast majority of studies have targeted mtDNA regions in order to maximize the chances of recovery. This is because cells contain 100 to 10,000 copies of mtDNA and therefore it constitutes a primary source of DNA in ancient specimens. However, the insights gained using a single-locus marker like mtDNA are limited, and complex evolutionary events are unlikely to be resolved without hundreds of thousands of unlinked loci, which are far beyond the reach of PCR-based approaches.

1.2.3. Paleogenomics era

The development of SGS technologies and their application to the study of ancient remains has revolutionized the field of aDNA, making it feasible to move from single genetic loci to complete genome sequences, and reaching time depths that researchers hardly imagined ten years ago.

Besides the general advantages of SGS methods outlined in section 1.1, high-throughput sequencing offers additional properties specially suitable for aDNA studies. In fact, aDNA research is probably one of the fields most dramatically transformed by SGS. First, given that extracted DNA molecules are turned directly into a DNA library and sequenced over their full length, the detection of short DNA fragments, which make up the vast majority of fragments in aDNA libraries, is highly increased. With previous PCR-based approaches, DNA sequences needed to be long enough (usually longer than 100 bp) to enable the hybridization of both primers and thus most of the sequences present in ancient extracts were untargeted. Second, SGS technologies allow a better characterization of DNA preservation, which is crucial for deciding the most efficient sequencing strategy and also for distinguishing endogenous from contaminant sequences. Fragment length distribution, percentage of endogenous DNA and rates of deamination can be easily obtained after aligning the sequences to a reference genome. Importantly, the high rate of deamination at the ends of aDNA sequences is used as one of the main authentication criteria, but it could not be studied with PCR-based methods because the ends of sequenced reads represented the primer sequences and not the original ends of aDNA molecules.

Over the last ten years, SGS data from hundreds of ancient samples have been published. For some of them the complete genome has been obtained, and for a few exceptional samples even a high coverage genome is available. In 2006, the first aDNA study using SGS (Poinar et al. 2006) produced ~13 Mb of mammoth nuclear

DNA, followed by 1 Mb of Neandertal DNA (Green et al. 2006). A reanalysis of the later study suggested that more than 50% of the sequences were modern human contaminants (Wall and Kim 2007), highlighting the importance of data authentication in aDNA studies. The field quickly moved to complete or almost complete genomes at low coverage, such as the mammoth genome (Miller et al. 2008) or the 1.3-fold coverage Neandertal genome sequenced as part of the Neandertal Genome Project (Green et al. 2010). This landmark study revealed that Neandertals share more alleles with present-day humans from Eurasia, Melanesia and the Americas than with sub-Saharan Africans, suggesting an interbreeding event before the divergence of all non-African modern humans. This and other interbreeding events have been confirmed with more Neandertal genomic data, including the high coverage genome of a Neandertal from the Altai Mountains (Prüfer et al. 2014; Kuhlwilm et al. 2016). Importantly, paleogenomics has provided the first hominin group identified via genetic analysis rather than by anatomical description, the Denisovans from southern Siberia. While their mtDNA is an outgroup to Neandertals and modern humans (Krause et al. 2010b), the nuclear genome placed Denisovans as a sister group of Neandertals and revealed a 4-6% contribution to the genome of present-day Melanesians (Reich et al. 2010; Meyer et al. 2012; Sawyer et al. 2015).

Remarkably, three paleogenomic studies stand out because of the extremely old DNA recovered. The oldest is the 1.12-fold coverage genome sequence from a Middle Pleistocene horse recovered from permafrost (Orlando et al. 2013), which demonstrated that DNA can

survive for several hundreds of thousands of years under favorable conditions. The other studies were able to recover an almost complete mtDNA (Meyer et al. 2014) and additional mtDNA and nuclear sequences (Meyer et al. 2016) from Middle Pleistocene hominins from Sima de los Huesos in Nothern Spain. These studies have important implications for understanding the hominin evolutionary history, and showed once again discrepancies between mtDNA and nuclear histories. The Sima de los Huesos hominins carried mtDNAs more closely related to those of Denisovans in Asia than Neandertals, even though their nuclear genomes showed that they are more closely related to Neandertals. To explain such discordance, the authors postulate, among other possibilities, a turnover in the Neandertal mtDNA gene pool later in their history that introduced the mtDNAs seen in Late Pleistocene Neandertals.

With regard to modern human paleogenomes, the first complete genome sequence was published in 2010 (Rasmussen et al. 2010). It corresponded to a 4,000-year-old paleo-Eskimo from Greenland's Saqqaq culture, whose genome was sequenced to 20-fold coverage from a tuft of hair. This was followed in 2012 by the complete genome of the famous 5,300–year-old Tyrolean Iceman (also known as Ötzi) (Keller et al. 2012). Ever since, many ancient individuals have yielded whole genome data, mainly from Eurasia (Sánchez-Quinto et al. 2012; Skoglund et al. 2012; Fu et al. 2014; Gamba et al. 2014; Lazaridis et al. 2014; Raghavan et al. 2014; Seguin-Orlando et al. 2014; Skoglund et al. 2014a; Fu et al. 2015; Jones et al. 2015; Olalde et al. 2015; Cassidy et al. 2016; Martiniano et al. 2016; Schiffels et al. 2016), including the first complete European hunter-

gatherer genome (Olalde et al. 2014) presented in this thesis, but also from the Americas (Rasmussen et al. 2014; Raghavan et al. 2015; Rasmussen et al. 2015a) and Africa (Llorente et al. 2015). Recently, thanks to the introduction of technical improvements at different stages of the SGS pipeline and the use of target enrichment approaches, human aDNA studies are shifting from the analysis of one or a few exceptionally well-preserved individuals towards the recovery of genome-wide data from dozens of individuals (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015; Fu et al. 2016), allowing powerful population-based analyses.

Although they have not been explored in this thesis, two promising paleogenomic applications are worth mentioning. The first is the recovery of epigenetic information from aDNA data. Depth of coverage patterns and post-mortem deamination have been used to infer nucleosome and methylation maps, respectively (Briggs et al. 2010; Gokhman et al. 2014; Pedersen et al. 2014), providing insights into the evolution of epigenetic regulation. The other is the reconstruction of ancient pathogen genomic sequences, which is already improving our knowledge of pathogen evolution and adaptation (Bos et al. 2011; Martin et al. 2013; Wagner et al. 2014; Rasmussen et al. 2015b).

1.3. Holocene history of Europe

1.3.1. Upper Paleolithic context

Anatomically modern humans (AMH) spread out of Africa some 75-50 kya and arrived in Europe around 45 kya (Trinkaus 2005; Benazzi et al. 2011; Higham et al. 2011), encountering Neandertals after more than 400,000 years of separated evolutionary history (Hublin 2009). The nature of this interaction has been surrounded by substantial controversy, but increasing archeological and genetic evidence (Rougier et al. 2007; Green et al. 2010; Fu et al. 2015) support several interbreeding events between these hominin groups. Around 40 kya, Neandertals and their associated material culture disappeared from the fossil record (Higham et al. 2014), leaving AMH as the only inhabitants of Europe.

The Upper Paleolithic archaeological and palaeontological records of Europe are the richest and best-studied in the world for this period. As a consequence, archaeologists have identified different material cultures that varied in temporal scale and geographical distribution: Aurignacian, Gravettian, Solutrean, and Magdalenian. However, whether these cultural innovations represent movements of people is widely debated (Bar-Yosef 2007). Recent aDNA work (Fu et al. 2016) suggests that, at least in some cases, population transformation played an important role.

This period of AMH occupation, associated with an increase in symbolic and technological complexity, was marked by the Last

Glacial Maximum (LGM) that lasted between 26.5-19 kya (Clark et al. 2009).

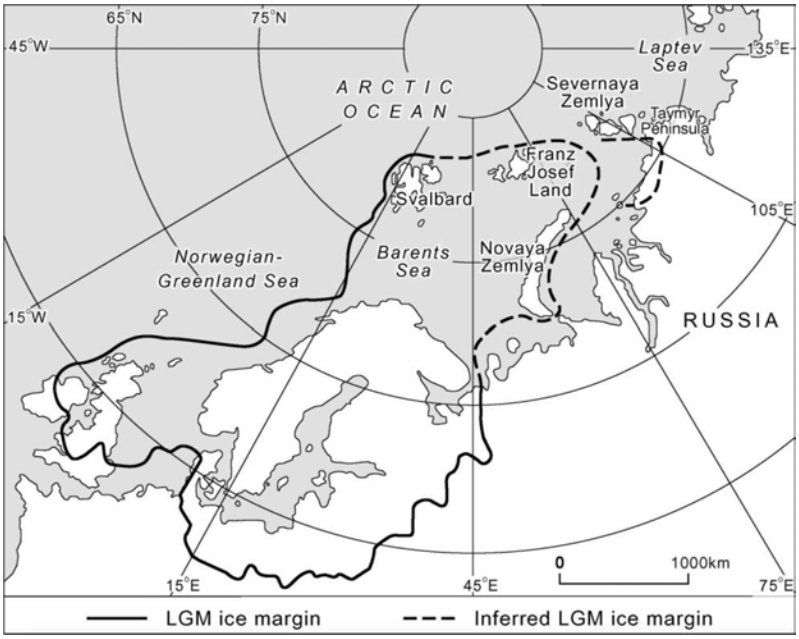


Figure 5. Location of the Eurasian High Arctic and ice-sheet limits during the LGM. (Siegert and Marsiat 2001)

During its cold peak, large parts of the continent were covered with ice sheets (Figure 5), driving local extinctions and pushing humans into southern refugia such as the Iberian, Italian and Balkan peninsulas, which acted as reservoirs of human diversity during this critical event. When climatic conditions improved and ice sheets began to retreat, central and northern latitudes were subsequently recolonized from these southern refugia (Gamble et al. 2004). While the distribution of some mtDNA haplogroups (e.g. H1, H3) have been interpreted as signals of these expansions (Achilli et al. 2004), growing aDNA evidence heavily undermines the role of H

haplogroups in the post-glacial repopulation of Europe (Fu et al. 2016).

The end of the last glacial period marks the boundary between the Pleistocene and the Holocene and was accompanied by the transition to a new period of human prehistory, the Mesolithic.

1.3.2. Mesolithic-Neolithic transition

The term Mesolithic ("Middle Stone Age") is used to designate those hunter-gatherer populations that lived between the end of the Pleistocene about 11 kya and the beginning of the Neolithic period, whose time of onset varies across continental regions.

During the Mesolithic, hunter-gatherer groups began to inhabit forest land and open tundra that had earlier been covered by ice. Increasingly warmer temperatures also triggered changes in the fauna. While some megafauna representatives such as the woolly mammoth, the woolly rhinoceros and the cave bear became extinct (Barnosky et al. 2004), other herbivores such as reindeer migrated north in search for more suitable habitats. Hunting of species such as aurochs, red deer and wild boars became a specialized task for the Mesolithic hunters, whose diet heavily relied on aquatic resources (Zilhão 2000). This was accompanied by some technological innovations, most notably the microlithization of flint tools, used as projectile points and also as components in other complex tools. In general, Mesolithic populations seem to be composed by small and highly mobile groups (Whittle and Cummings 2007), which resulted

in a significant degree of cultural homogeneity reflected in material objects, such as figurines and ornaments made with atrophic red deer canines that can be found in sites all over Europe.

This homogeneity is also evident from their mtDNA haplogroup composition. Ancient mtDNA sequences recovered from more than 70 European Mesolithic hunter-gatherers (Bramanti et al. 2009; Krause et al. 2010a; Hervella et al. 2012; Sánchez-Quinto et al. 2012; Bollongino et al. 2013; Der Sarkissian et al. 2013; Fu et al. 2013b; Lazaridis et al. 2014; Olalde et al. 2014; Skoglund et al. 2014a; de-la-Rua et al. 2015; Haak et al. 2015; Fu et al. 2016) belonged almost exclusively to haplogroup U, currently found at frequencies between 1 and 7% in most modern European populations, but at up to 20% in Baltic populations and around 40% in Saami (Pinhasi et al. 2012). Considering the vast geographic area involved, from the Iberian Peninsula in the West to Samara (Russia) in the East, this pattern suggests a relatively small population size probably resulting from a bottleneck during the LGM and subsequent demographic expansions. The abundance of U haplogroups in pre-Neolithic Europeans is consistent with coalescence time estimates for haplogroup U, which was described as one of the oldest Eurasian branches based on modern mtDNA data (Richards et al. 2000), and is thus a good candidate for the initial arrival of AMH in Europe. Despite the remarkable uniformity at the basal haplogroup level, some small differences are discernible. For instance, U5a haplogroups show higher frequencies in the East, whereas U5b haplogroup are more common in the West (Brandt et al. 2015), matching the proposed place of origin of these U5 subclades (Malyarchuk et al. 2010).

Interestingly, most of the non-U mtDNA sequences so far reported in Mesolithic hunter-gatherers were obtained using PCR-based approaches, for which contamination cannot be completely monitored and therefore they should be viewed cautiously. In fact, one Magdalenian individual from El Mirón cave in Spain was reported as belonging to haplogroup H with a PCR-based methodology (de-la-Rua et al. 2015), and later undoubtedly classified as U5b using a SGS-based mtDNA capture approach (Fu et al. 2016).

Given that single-locus systems like mtDNA and the Y-chromosome are unable to capture the full underlying population history, the use of hundreds of thousands of unlinked nuclear markers is needed in order to better study the European pre-Neolithic genetic substratum. In 2012, Skoglund and colleagues (Skoglund et al. 2012) published for the first time partial genomic data from 3 Scandinavian hunter-gatherers. Since the beginning of my phD work in the second half of 2012, genome-wide data from additional Mesolithic foragers have become available. Their implication for our understanding of this period will be conveniently discussed in sections 4 and 5.

The scenario radically changes with the arrival of the Neolithic, defined by a subsistence strategy based on domestic animals and crops. Farming arose in the Fertile Crescent of West Asia about 11 kya, when hunter-gatherers groups began to establish sedentary settlements and developed a food producing economy (Childe 1950). The Neolithic spread into Europe via Anatolia, reaching Greece by 7,000 cal BCE (Perlès 2001). From there, it subsequently spread west and northwest, following two different routes: the Mediterranean

coast to the west, associated to the Impressed/Cardial culture, and the Danube River to the center and north of Europe, associated with the Linear Pottery Culture (LBK) (Figure 6).

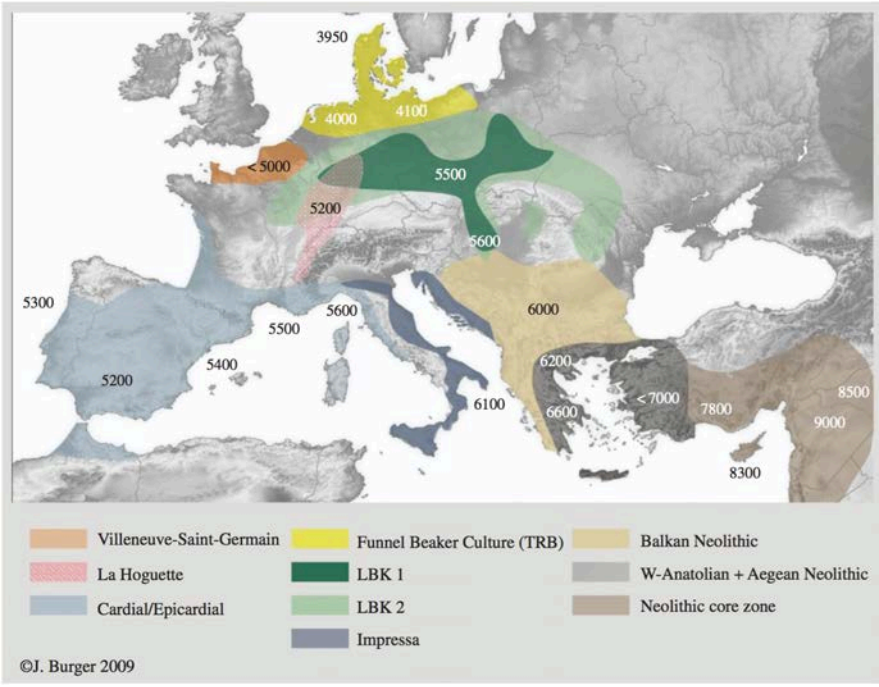


Figure 6. Arrival dates and approximate geographical expansions of defined Early Neolithic cultures. (Gerbault et al. 2011)

Overall, the expansion was probably complex and varied in intensity from one area to another, depending primarily on ecological conditions favorable to agriculture and also on local human population densities. In the Iberian Peninsula, the earliest Neolithic sites are represented by Cardial settlements located along the Mediterranean façade and dated to about 6,000 cal BCE (Zilhão 2000). Later, the agricultural frontier advanced towards the interior following important routes of communication like the Ebro River (Arias 2007). In some regions, such as southern Scandinavia and the

eastern part of the Baltic Sea, a hunting and gathering economy (although sometimes incorporating Neolithic pottery to their material culture) continued long after farming was introduced in the rest of Europe (Malmer 2002). By 4,000 cal BCE, the Neolithisation of continental Europe was complete and the hunting and gathering way of life had disappeared forever in this region of the world.

The Neolithic represents one of the biggest and most dramatic changes in human history, to the point that all modern European societies derive from that cultural, social and ecological turning point. Material culture evolved with the use of pottery and polished tools, developed to serve the new range of human activities. The establishment of large settlements allowed an increase in population size and density, and societies became more complex and stratified. The increase in population density, together with living arrangements in close proximity to domestic animals resulted in a larger number of potential infectious diseases, which could not be sustained in low density hunter-gatherer groups (Diamond 1997; Antia et al. 2003; Wolfe et al. 2007). Examples of zoonoses include measles that originated in cattle (Moss and Griffin 2006), influenza from ducks and pigs, smallpox that may have been transferred from cattle or camels and some other, more controversial examples such as tuberculosis (Wolfe et al. 2007). Furthermore, the human diet shifted from an almost total dependence on animal preys to a diet primarily based on cereals (Richards et al. 2003) and, later on, on domesticated animals and their derived products. A carbohydrate based diet is thought to have contributed to the increasing prevalence of malnutrition and oral diseases such as caries amongst agriculturist, as

compared to previous hunter-gatherer populations. Consequently, the transition to farming is associated with a decline in health (Spencer Larsen 1995; Aufderheide et al. 1998), illustrated by the drop in average adult height of European males from 179 cm during the Pleistocene to 150 cm during the Neolithic period (Hermanussen 2003). Parallel to changes in diet and health, the Neolithic probably involved changes in social behaviour that allowed humans to live in large sedentary groups based on mutual cooperation and increasingly large social differences.

This sharp change in lifestyle during the Mesolithic-Neolithic transition likely imposed strong selective pressures on genes associated to diet, immunity and behaviour, resulting in many genetic adaptions like the well-known lactase persistence phenotype related to milk consumption during adulthood (Itan et al. 2009).

The processes involved in the Neolithisation of Europe are crucial to understand the genesis of modern European populations and consequently are a subject of intense debate. For at least a century, two opposed scenarios have been proposed. According to the 'demic diffusion' model, Neolithic was carried throughout Europe by the actual movement of farming peoples from the Middle East intruding into a landscape of foragers (Childe 1925; Ammerman and Cavalli-Sforza 1984; Renfrew 1996; Chikhi et al. 1998), whereas the 'cultural diffusion' model supports a spread of ideas that led to the acculturation of indigenous populations (Barker 1985; Whittle 1996). However, these models occupy only a small fraction of the total possibility space and intermediate scenarios have also been proposed

to account for regional differences in a process covering a vast geographical area. The two alternative models produce very different predictions regarding the genetic affinities of both ancient and modern Europeans, and therefore genetic evidence has been used to test different scenarios. The demic diffusion model was first supported by the presence of a southeast-northwest allele frequency gradient in classical genetic markers across present-day Europe and the Near East (Menozzi et al. 1978; Ammerman and Cavalli-Sforza 1984) that strongly resembled the radiocarbon map for the Neolithic expansion. Nevertheless, genetic clines tell us nothing about when they were generated, and the observed genetic gradient could be the outcome of other human dispersals following a similar route. In fact, despite the amount of mtDNA, Y-chromosome and genome-wide genetic data accumulated for present-day European populations since the 1980s (Richards et al. 1996; Richards et al. 2000; Belle et al. 2006; Balaresque et al. 2010), the relative weight of migration versus acculturation in the spread of farming was far from being elucidated.

After the publication of the first mtDNA sequences from Neolithic farmers, interpretations varied from mainly pre-Neolithic ancestry (Haak et al. 2005) to mainly Neolithic ancestry (Sampietro et al. 2007) in modern Europeans. Finally, when a large number of Neolithic and Mesolithic mtDNA sequences became available (Bramanti 2008; Bramanti et al. 2009; Malmström et al. 2009; Haak et al. 2010; Deguilloux et al. 2011; Gamba et al. 2012; Hervella et al. 2012; Brandt et al. 2013; Brotherton et al. 2013; de-la-Rua et al. 2015; Szécsényi-Nagy et al. 2015), it was clear that the haplogroup composition of Mesolithic hunter-gatherers and early Neolithic

farmers was highly different. Hunter-gatherers, as previously mentioned, presented almost exclusively U haplogroups, whereas the most common haplogroups in Neolithic farmers were N1a, T2, K, J, HV, V,W and X. These results suggested that, at least from a matrilineal perspective, the Neolithisation process was associated with significant genetic influx from the Middle East, which was also consistent with aDNA Y-chromosome results (Haak et al. 2010; Lacan et al. 2011a; Lacan et al. 2011b; Szécsényi-Nagy et al. 2015). Furthermore, the presence of U mtDNA and Y-chromosome I haplogroups in early farmers could indicate some degree of admixture with local hunter-gatherers.

Despite the utility of uniparental markers, a definitive assessment of these questions must rely on nuclear genetic data. In fact, published genome-wide data from several early European farmers have certainly extended and consolidated previous findings, and will be discussed in sections 4 and 5.

1.3.3. The onset of metallurgy

During the Middle Neolithic about 5,000-3,000 cal BCE, there seems to be a period of stability in Europe, without major inputs from surrounding areas as judged from archaeological and genetic data. The different Early Neolithic cultures disintegrated into several smaller cultural groups with rather regional dispersal (Brandt et al. 2015), and presented a similar mtDNA composition compared to that of earlier farmers. Interestingly, genome-wide data from Middle Neolithic individuals (Skoglund et al. 2014a; Haak et al. 2015;

Cassidy et al. 2016) has revealed an increase in hunter-gatherer ancestry in multiple regions of Europe. This resurgence of the pre-Neolithic genetic component could be explained by local admixture between farmers and hunter-gatherer groups that persisted in different regions after the initial arrival of Neolithic people, or instead by the expansion across Europe of a population with high levels of hunter-gatherer ancestry. An interesting site that could provide valuable insights into these processes is the Blätterhöhle cave site in western Germany. Dietary stable isotope and aDNA analyses (Bollongino et al. 2013) identified two distinct groups in the Middle Neolithic occupation phase: one with a high freshwater fish diet and exclusively mtDNA U haplogroups, and the other with terrestrial diet and typical Neolithic mtDNA lineages, but also some U lineages. Genome-wide data from this and other sites will help to elucidate the genetic relationships and mobility patterns of these distinct populations that lived in Europe thousands of years after the initial advent of farming.

The third millennium BCE, which roughly coincides with the Late Neolithic and Early Bronze age, represents one of the most important turning points in the history of Europe. It was a period of profound social and economic transformations over wide areas in western Eurasia (Anthony 2007; Kohl 2007), triggered by the incipient use of metal, the appearance of novel social structures with increased stratification and a complex pattern of interaction networks. This period was characterized by the emergence of two pan-European cultural phenomena: the Corded Ware complex in the East and the

Bell Beaker complex in the West, with an overlapping zone in Central Europe.

The formation of the Corded Ware complex is strongly linked with the transformations taking place in the Pontic steppe during the late fourth millennium BCE. About 3,000 cal BCE, in the area north of the Black Sea, tribal groups associated to the Yamnaya culture adopted a nomadic form of pastoralism, made possible by the use of wagons and domesticated horses (Kristiansen 2011; Anthony and Ringe 2015). They moved between summer and winter pastures, making extensive use of grazing animals and grasslands in a mixed economy with agriculture playing a minor role. The formation of a new economy was associated with new definitions of family and inheritance to control property and mobile wealth. The monogamous family became the central social and economic institution, embedded in a patrilineal kinship system (Kristiansen 2011). Furthermore, warfare was both glorified and institutionalized with the consolidation of the ideological hegemony of the warrior elite (Vandkilde 2006). This new social organization was ritually expressed in the burials, consisting of individual burials of family groups with rich personal grave goods that were covered by a low mound or kurgan. Later graves were added on top of each other and covered with a new mound (Kristiansen 2011).

From the steppe, this mobile agro-pastoral economy expanded into Central and Northern Europe, where is named Corded Ware/Single Grave Culture (Figure 7). The expansion was rapid, as shown by the decimation of European forest lands in less than hundred years and

the appearance of large open areas for grazing the herds (Andersen 1995; Kremenetski 2003), and has been linked to the introduction of Indo-European languages in Europe (Anthony 2007).

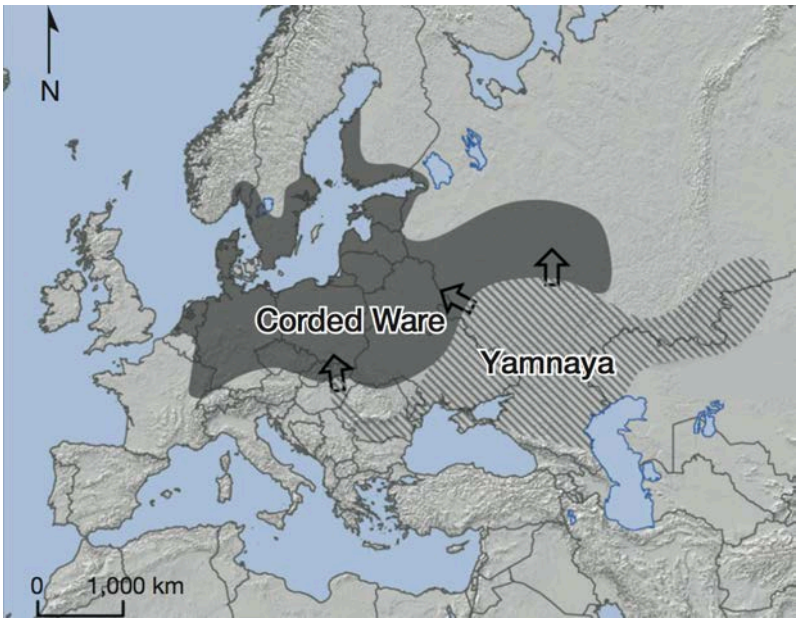


Figure 7. Geographic distribution of Yamnaya and Corded Ware cultures. Modified from Allentoft et al. (2015).

The relation between the Corded Ware complex and the steppe has recently been demonstrated with aDNA (Allentoft et al. 2015; Haak et al. 2015). Genome-wide data from Corded Ware samples showed a population turnover in Central Europe during the first half of the third millennium BCE, with Corded Ware individuals tracing ~75% of their ancestry to a population related to Yamnaya steppe herders. The Y-chromosome composition also supports this connection and presents a marked uniformity dominated by R1b and R1a haplogroups, which is consistent with a patrilineal organization of these societies.

The Corded Ware culture is complemented in the West by the Bell Beaker complex, named after its characteristic pottery. This culture shows a well-defined cultural package (Case 2004), involving different objects usually found in individual graves (Figure 8a):

-Pottery vessels with the shape of an inverted bell and decorated with horizontal bands of incised, excised or impressed patterns (Figure 8b). It has been suggested that most of them were designed for the ritualistic consumption of alcohol, most notably beer (Sherratt 1989).

-"Tanged daggers" made of cooper, often found in male graves and originally fitted into wooden handles. Bell Beaker are closely linked to metallurgy of copper and gold, and the use of metal in these societies became widespread for the first time in Europe (Czebreszuk 2004a). In fact, the introduction of metallurgy in northern Europe is likely associated to their expansion.

-Barbed flint arrowheads, also selectively found in male graves.

-Stone wrist-guards, traditionally associated with the protection of the archer's wrist from the bowstring. They are quite widespread in Iberia and the Atlantic Bell Beaker sites.

-Palmela points common in Iberian sites. Although some of them are considered to be arrowheads, the size and weight of many of them suggest they were used as javelin points.

-V-perforated buttons made of different materials, e.g. horn, bone and amber.

The traditional Bell Beaker package is associated, as we can see, to a particular sphere of life related to war and hunting, thus representing the emergence of a warrior-like social elite or aristocracy that became consolidated during the Bronze Age period. Its presence as burial practice, mainly in male graves, indicates that the objects were associated to the ancestors and the otherworld.

Figure 8. Bell Beaker funerary rituals. a) Reconstruction of a Bell Beaker burial. Museo Arqueológico Nacional, Madrid, Spain. **b)** Bell-shaped beaker from Salisbury in England**.** (Bogucki and Crabtree 2004)

The origin and impact of the Bell Beaker expansion have been the subject of numerous archaeological debates that are still ongoing. Some scholars proposed a continuous sequence of cultural developments in the Lower Rhine region from the Corded Ware pottery to the Bell Beaker ceramics (Lanting and Van der Waals 1976). However, refined radiocarbon dates later showed that Bell Beaker sites in Southwestern Europe, notably those at the Tagus estuary dating to 2,900-2,800 cal BCE, predated sites from Central Europe (Bailly and Salanova 1999; Müller and van Willigen 2001; Cardoso 2014), indicating a likely origin in Iberia (Harrison and Martin 2001). There, a characteristic type of Bell Beaker decoration developed (the so-called Maritime type decorated with bands filled with impressions made with a comb or cord) and very quickly spread to some Atlantic regions. The ecologically rich Tagus estuary has numerous walled enclosures and settlements from this period, indicating a rich demography. This is in agreement with numerous and spectacular tombs along with rich assemblages of material culture in the area. Together with this local origin, potential contacts with North Africa, probably based on trading of objects such as ostrich shell and ivory, has been postulated (Case 2004). To place the ancestral roots of this pottery outside Europe is, however, controversial.

From the Iberian Peninsula, the Bell Beaker complex quickly spread over large regions of western, north and central Europe (Figure 9), ranging from the Vistula river to the Atlantic coast and from Sicily to the British Isles, where it persisted for about 1,000 years. For a prehistoric culture, this represents an unprecedented geographic

scale, roughly equivalent to the area now occupied by the countries of the European Union (Czebreszuk 2004b). From an archaeological point of view, the Bell Beaker phenomenon is quite uniform -specially in its initial stages- although it became subsequently diversified through the assimilation and incorporation of local elements from previous Middle Neolithic traditions. This diversification is reflected in the pottery but also in the grave structure. Nevertheless, it cannot be ignored that is a perfectly recognizable and thus highly uniform archaeological tradition extending over a vast geographic area.

Figure 9. Areas with high concentration of Bell Beaker sites in Europe. (Czebreszuk 2004b).

The processes underlying the Bell Beaker expansion are controversial, and the arguments have again circled around the demic vs cultural diffusion dichotomy (Gallay 2001). While some researchers argue that the Bell Beaker expansion may have involved the largest sweep in the genetic composition and demographics of the post-farming European prehistory, others favor a model of diffusion of a specific cultural "package" adopted by local communities (called the "pots-not-people" paradigm). Still, others support mixed models, with colonization in the first settlements in a given area and subsequent acculturation of indigenous groups (Lemercier 2012), which could have been facilitated by social prestige associated to a permanent upper class of warrior aristocracies.

The publication of mtDNA sequences from ~40 individuals from Central and Northern Europe (Melchior et al. 2010; Lee et al. 2012; Brandt et al. 2013; Brotherton et al. 2013) has contributed to the ongoing debate about the origin and expansion of the Bell Beaker complex. These studies showed a high frequency of H lineages (~48 %), the most frequent haplogroup in West European populations today, together with the absence of I and U2 lineages present in Corded Ware samples. When compared to other ancient and modern populations, Bell Beakers showed affinity to some Iberian Neolithic populations and to present-day Iberians due to high frequencies of sub-haplogroups H1 and H3, suggesting genetic influx from the West into Central Europe. Recently, genome-wide data from 19 Bell Beaker samples from Germany and the Czech Republic (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015) revealed a clear signal of steppe ancestry, although less intense than in Corded Ware

people. Since only Central European individuals were analyzed, whether this signal is present in other areas is currently unknown. Thus, a comprehensive genetic survey covering the entire range of this pan-European culture is needed in order to fully address questions about its origin and expansion.

2. METHODS

2.1. Laboratory procedures

2.1.1. Sample retrieval and DNA extraction

For many ancient specimens, limited material is available for genetic analysis. For this reason, identifying the best skeletal elements with regard to aDNA preservation could greatly increase the rate of success in aDNA analysis. Recently, the dense part of the petrous bone (Figure 10a) has been shown to yield systematically higher endogenous DNA than teeth (4- to 16-fold increase) and other bones (183-fold increase) (Gamba et al. 2014; Pinhasi et al. 2015), allowing the study of samples from temperate or warm regions that are otherwise not suitable for aDNA analysis. When teeth are the available material, targeting the cementum layer of the roots (Figure 10b) is preferred, as it contains 14 times more endogenous DNA than the inner dentine (Adler et al. 2011; Damgaard et al. 2015).

Despite these useful recommendations, the percentage of endogenous DNA in an ancient sample is unpredictable and shows extreme variation within site (Reich et al. 2010) and even within the same bone (Green et al. 2010). This could be due to the presence of physical microenvironments with different conditions (e.g. humidity, pH) that facilitate or inhibit the growth of microorganisms involved in DNA degradation.

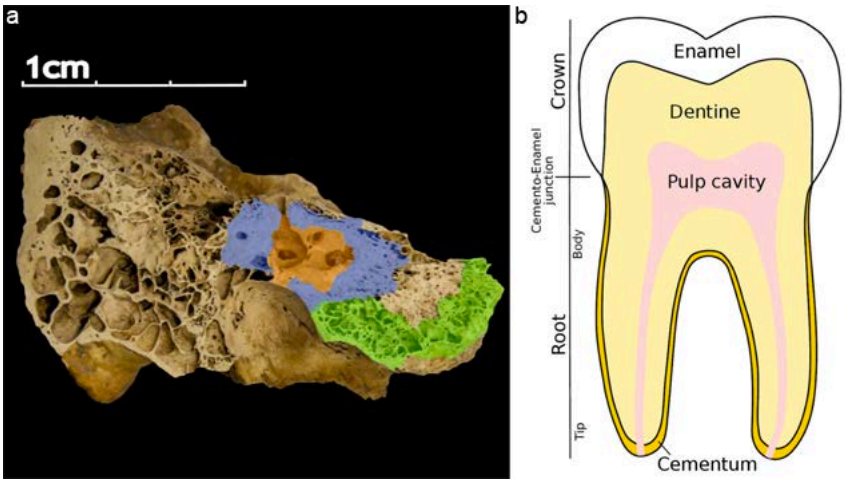


Figure 10. Skeletal elements with high percentage of endogenous DNA. a) Medial view of a cut of a left petrous bone. The dense part within the otic capsule (highlighted in orange) presents the best DNA preservation. Modified from Pinhasi et al. (2015). **b)** Diagrammatic representation of a human molar showing the DNA-rich cementum layer. (Adler et al. 2011)

Previous studies have detected high levels of pre-laboratory derived DNA contamination in human remains (Hofreiter et al. 2001b; Gilbert et al. 2005; Sampietro et al. 2006). Therefore, in order to avoid contamination, excavation and sampling should be performed under controlled conditions whenever possible. The use of sterile lab gear, including coveralls, face mask, face shield and gloves is highly recommended, as well as avoiding sample washing and cleaning steps (Fortea et al. 2008). When working with teeth and bones, removing the outermost surface is a standard procedure to exclude surface DNA contamination, and drilling at low speeds protects DNA from heat damage (Adler et al. 2011).

Most of the protocols developed for the extraction of aDNA consist of two steps: DNA release and DNA purification (Heintzman et al. 2015). In the case of hard tissues such as bones or teeth (the most

common types of samples in paleogenomic studies), during the DNA-release step the solid matrix is disrupted in order to release the DNA molecules into an aqueous solution. This is performed using a strong solution of proteinase K and ethylenediaminetetraacetic acid (EDTA), breaking down collagen and hydroxyapatite, respectively (Rohland and Hofreiter 2007). During the purification step, DNA is separated from other organic and inorganic compounds, using phenol–chloroform-based phase separation (Barnett and Larson 2012) or silica spin columns (Rohland 2012). In the later protocol, the product of the DNA release step is incubated with silica particles in suspension and a binding buffer. Silica particles adsorb DNA molecules and the binding buffer prevents silica particles from adsorbing coextracts that may have accumulated in the sample, such as humic acids and salts that could potentially inhibit subsequent enzymatic reactions.

Since maximizing DNA extraction efficiency is of vital importance, several methodological advances have been introduced after extensive testing. Importantly, length distributions of DNA sequences reported in SGS studies consistently show a mode of ~40 bp or larger (Green et al. 2010; Orlando et al. 2013), suggesting that shorter sequences are not efficiently recovered by common extraction protocols. Recently, modifications in the volume and composition of DNA-binding buffers (Dabney et al. 2013a; Allentoft et al. 2015) have improved by twofold to fivefold the retrieval of 35–50-bp molecules, which represent the vast majority of DNA sequences in ancient specimens. Furthermore, a pre-digestion step has been shown

to significantly increase the relative proportion of endogenous DNA recovered (Damgaard et al. 2015). During this pre-digestion, surface contaminants are supposedly released into solution first and discarded, whereas endogenous DNA, more protected within the bone's microniches, will not be released until the full digestion step.

2.1.2. Library preparation

The most widely used SGS technologies require the modification of extracted DNA molecules before sequencing. Modification usually means the attachment of artificial DNA segments (adapters) to both ends of the fragments so that they are recognized by the sequencing platform; this process is termed "library preparation".

Although the first aDNA studies using SGS methods applied standard library construction protocols developed for modern DNA (Green et al. 2006), researchers soon realized that these resulted in the lost of substantial DNA fragments. Therefore, the use of protocols that are tailored to aDNA characteristic features is highly recommended. Since different methodologies produce different misincorporation patterns, the choice of the experimental procedure to follow has important consequences for quantifying post-mortem DNA damage and authenticating aDNA data.

Currently, two main library construction methods are available for sequencing on the Illumina sequencing platform: Double-stranded library preparation (used in the works presented in this thesis) and single-stranded library preparation.

a) Double-stranded library preparation

First, during the end-repair step, 3' overhangs are removed and 5' overhangs are filled in to create blunt ends (Figure 11). Next, short adapters are ligated to the ends of double stranded DNA molecules, either directly using two different adapters (blunt-end ligation) (Figure 11a) or using a single, Y-shaped adapter with a T-overhang that is ligated to both ends of DNA fragments that have been manipulated to carry A-overhangs (A-tailed ligation) (Figure 11b). As adapter ligation is random, with blunt-end ligation 50 % of the fragments will not contain both of the different adapters and hence they will not be sequenced (Meyer and Kircher 2010). In contrast, the use of Y-shaped (complementary at the T-tailed end but with non-complementary arms at the other end) adapters ensures that aDNA strands are flanked by distinct non-complementary adaptor sequences at each end (Bentley et al. 2008). However, this strategy is biased against the ligation of DNA templates starting with thymine residues or uracils (thymine analogues) (Seguin-Orlando et al. 2013), which comprise a substantial portion of aDNA extracts. Finally, PCR is used to amplify the adapter-ligated DNA molecules and to add a platform-specific adapter (Figure 11). During this step, the number of PCR cycles should be minimized in order to maintain a high molecular complexity, and PCR polymerases should be carefully chosen to avoid length and GC biases (Dabney and Meyer 2012). When multiple libraries are pooled for sequencing in the same sequencing run, sample-specific indexes can be attached to each fragment of the library during amplification (Craig et al. 2008;

Kircher et al. 2012). These indexes will be later used to computationally assign each read to its corresponding sample.

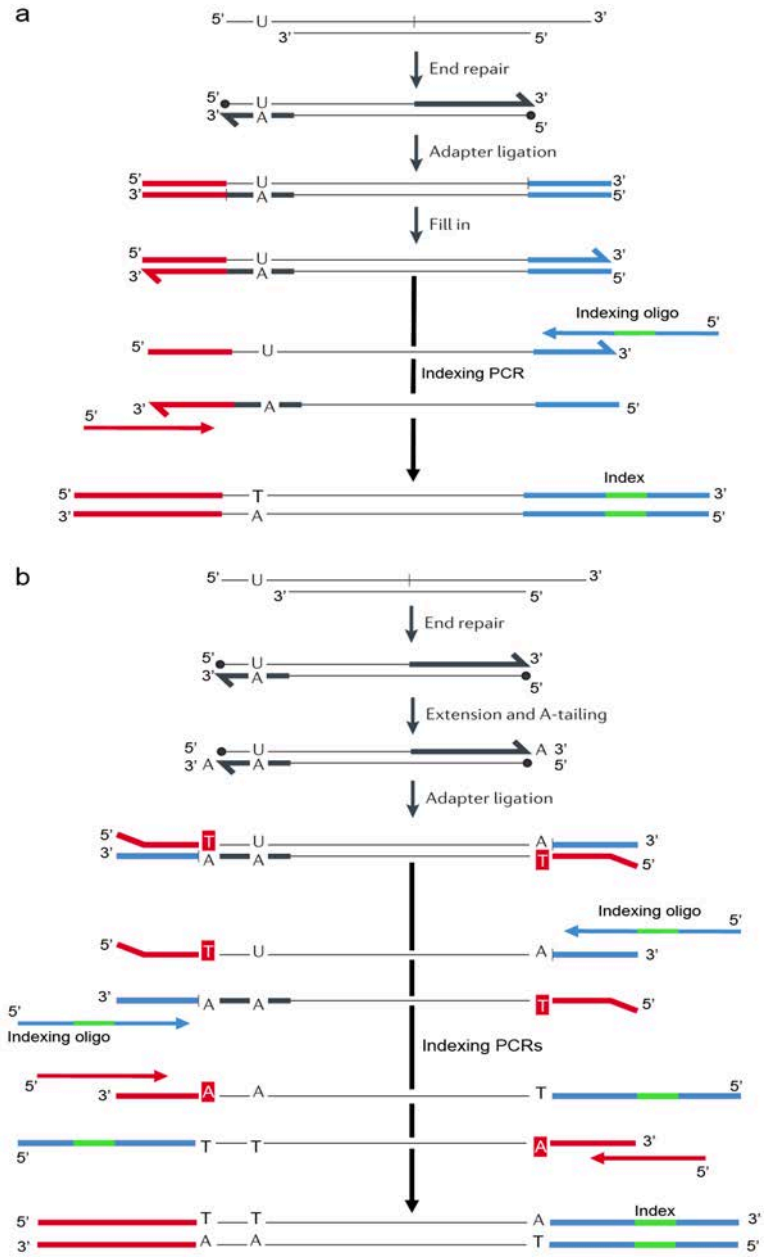


Figure 11. Double-stranded library preparation. a) Blunt-end ligation. **b)** Y-shaped ligation. Single-strand DNA breaks are shown as vertical lines. Modified from Orlando et al. (2015).

b) Single-stranded library preparation

In poorly preserved ancient samples, many fragments of surviving aDNA may be single-stranded, while others may contain nicks or end modifications that inhibit adaptor ligation. These molecules are not incorporated in double-stranded library preparation (Gansauge and Meyer 2013), limiting our ability to fully characterize the true complexity of aDNA extracts. A recently developed method (Meyer et al. 2012; Gansauge and Meyer 2013) overcomes this limitation by incorporating both single-stranded and double-stranded molecules into DNA libraries (Figure 12). First, double-stranded DNA fragments are converted into single-stranded DNA through heat denaturation and residual phosphate groups from the 5’ and 3’ ends are removed. Then, a biotinylated single-stranded adapter is ligated to the 3’ ends of the DNA strands, which are immobilized to streptavidin-coated beads. Next, a 5’-tailed primer complementary to the adapter is attached and used to copy the template strand, which becomes double-stranded. A second adapter is then joined to the opposite end of the DNA by blunt-end ligation with T4 DNA ligase (Figure 12). The library molecules are released from the beads by heat denaturation and used for PCR amplification and indexing. Although this method appears to recover a greater fraction of endogenous DNA material in highly degraded osseous samples (Meyer et al. 2012; Meyer et al. 2014), with conversion efficiencies of about 30–70% (Gansauge and Meyer 2013), its application to samples with better DNA preservation might not be as beneficial, in part due to higher cost and slower preparation times than double stranded-libraries (Bennett et al. 2014).

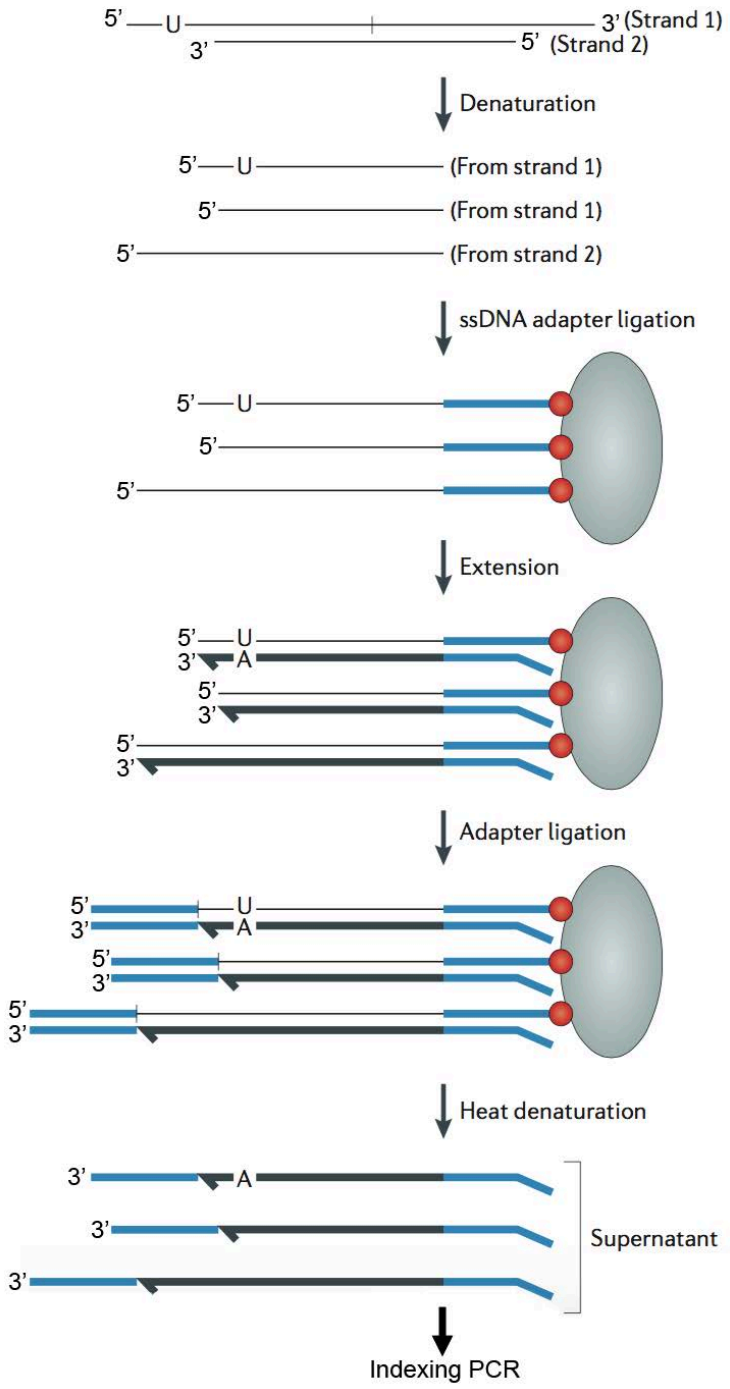


Figure 12. Single-stranded library preparation. Single-strand DNA breaks are shown as vertical lines, biotinylated adaptor groups are shown in red and streptavidin-coated beads are shown in grey. Modified from Orlando et al. (2015).

DNA fragments recovered from ancient samples present a high rate of cytosine deamination, specially at both ends of the molecules, resulting in C to T and G to A changes in the sequenced reads (section 1.2.1). Such nucleotide misincorporations lead to high amounts of errors in aDNA sequences that can potentially bias inferences and create misleading results.

Thanks to different molecular tools applied during library preparation, these errors can be removed from the final sequences. One approach is to use DNA polymerases that cannot replicate uracil (Rasmussen et al. 2010). During library PCR amplification, library fragments containing at least one uracil will not be amplified and therefore excluded from sequencing. However, this can produce a significant loss of endogenous DNA, given that in aDNA extracts up to 60% of all endogenous DNA fragments are estimated to contain at least one uracil (Briggs et al. 2010).

An alternative approach is to treat aDNA extracts with Uracil DNA Glycosylase (UDG) and Endonuclease VIII (endoVIII) before library preparation (Briggs et al. 2010). UDG removes uracil residues from DNA leaving abasic sites that are subsequently repaired with endoVIII by cleaving on the 5'- and 3'-sides (Figure 13c). This creates fragments that are shorter but still amenable to amplifying and sequencing.

The downside of removing aDNA damage is that DNA sequences can no longer be used to authenticate the samples. Hence, a common strategy is to first use an aliquot of the aDNA extract to prepare 'test'

libraries that can be evaluated for features expected from authentic aDNA. If these features are present, additional libraries free of errors can be built using UDG and endoVIII treatment. Alternatively, a single library can be prepared using partial UDG treatment (Rohland et al. 2015). This library preserves a damage signal at the terminal bases of DNA fragments (Figure 13b) and can be used both to test for authenticity and to perform genomic analyses, saving time and costs.

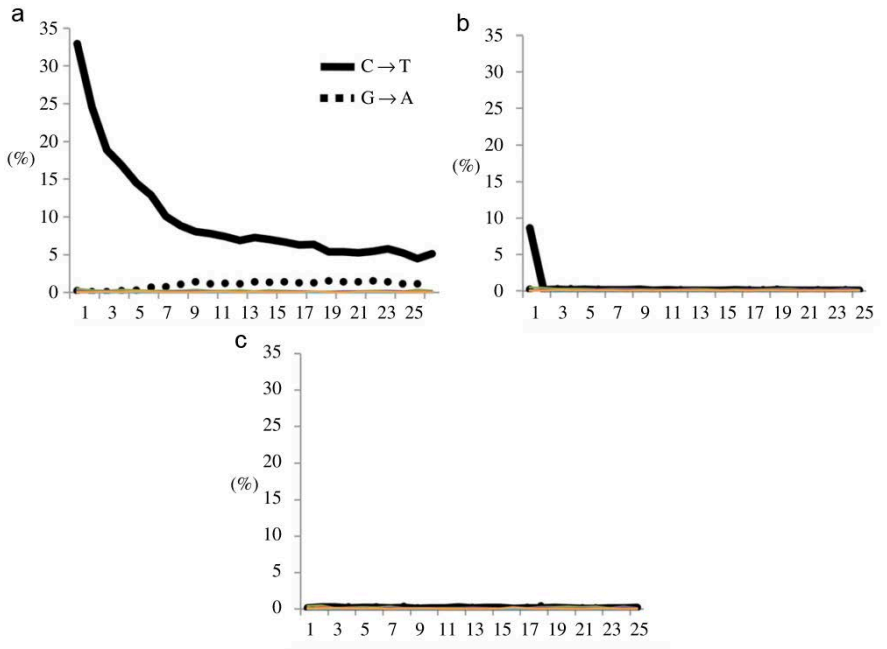


Figure 13. Damage profile of the terminal 25 nucleotide of a sample treated in three different ways during library preparation. a) No UDG treatment. **b)** partial UDG treatment. **c)** Full UDG treatment. Modified from Rohland et al. (2015).

2.1.3.Targeted capture approaches

Once DNA libraries have been prepared and amplified they can be directly sequenced, an approach called shotgun sequencing. This strategy provides valuable information about DNA degradation in the sample, such as the percentage of endogenous DNA or the fragment length distribution, and can be used to recover complete genomes of exceptionally well-preserved specimens (Rasmussen et al. 2010; Meyer et al. 2012; Olalde et al. 2014). However, since most ancient samples have very low proportions of endogenous DNA, the amount of sequencing required to obtain significant information is uneconomical and prohibitive for many researchers.

This problem can be overcome by using targeted capture methodologies that increase the endogenous content of aDNA libraries, either during library construction (Gansauge and Meyer 2014), or after (Maricic et al. 2010; Carpenter et al. 2013; Enk et al. 2014; Haak et al. 2015). In this thesis, three different enrichment approaches have been applied: mtDNA capture using PCR products, Whole-genome In-Solution Capture (WISC) and 1240k in-solution capture.

a) mtDNA capture using PCR products

In this method, complete mitochondrial genomes are captured using custom-made long-range PCR products (Maricic et al. 2010). First, two overlapping mtDNA fragments are amplified with PCR using DNA extracted from a present-day individual as a template. Then, the resulting PCR products are pooled in equimolar amounts and

sonicated to produce 150-700 bp sequences. These fragments are biotinylated, immobilized on streptavidin-coated magnetic beads and denatured to obtain the final mtDNA baits. Next, libraries prepared from aDNA extracts are amplified, denatured and incubated with the streptavidin beads coated with mtDNA baits. During incubation, mtDNA library fragments will hybridize to the baits. Finally, library molecules that do not hybridize such as those originating from environmental DNA and nuclear endogenous DNA are washed away and the mtDNA library pool is eluted. After a PCR amplification step, the enriched library is ready for sequencing.

b) WISC

Most enrichment strategies target a subset of the genome. When the goal is to target whole human genomes, artificially synthesizing probes covering 3 Gb of DNA is unfeasible. Thus, alternative approaches have been developed for whole-genome enrichment of aDNA libraries, such as the WISC methodology introduced by Carpenter *et al*. (2013) and a commercial alternative, MYbaits-WGE, which was developed by a company called MYcroarray (Ann Arbor, MI, USA). Both are based on a similar underlying molecular principle but with subtle differences in the protocol (Enk et al. 2014). These methods start with the construction of a genome-wide RNA bait library from modern DNA of a species (in our case a present-day human individual) closely related to the target ancient genome. This modern genomic DNA is sheared and the resulting fragments (~150-200 bp average size) are ligated to adaptors containing T7 RNA polymerase promoters (Figure 14). The RNA library is then generated by *in vitro* transcription in the presence of biotin 16-UTP,

so that the resultant RNA baits are biotinylated. These biotinylated RNA probes are hybridized in solution with the denatured aDNA. The hybridized molecules are retrieved using streptavidin-coated magnetic beads and the unbound library fraction is washed away (Figure 14). Finally, endogenous library fragments are released from the RNA baits with a RNase treatment and amplified before sequencing.

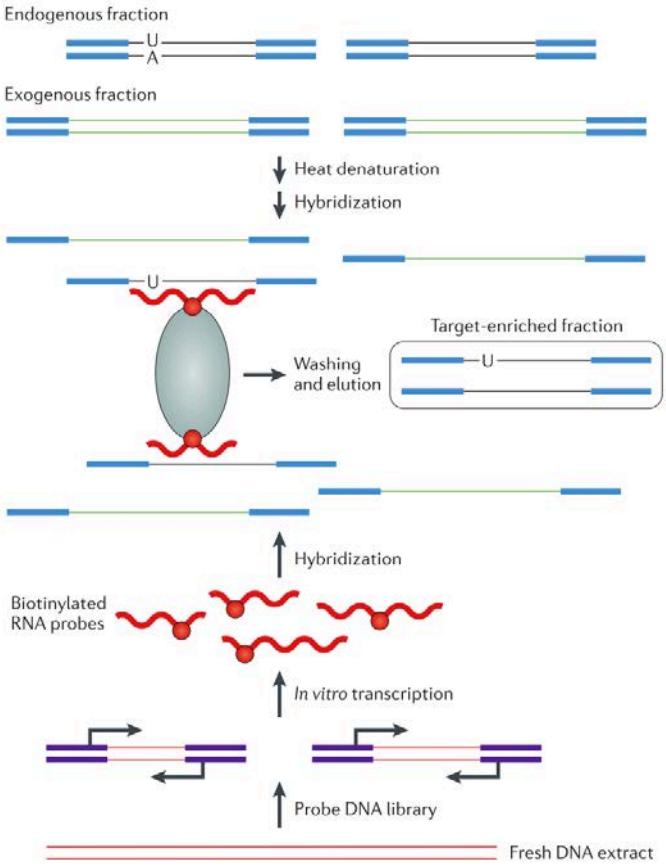


Figure 14. Whole-genome In-Solution Capture (WISC). (Orlando et al. 2015)

These approaches showed between 2- and 13-fold enrichment in non-clonal endogenous reads (Carpenter et al. 2013; Ávila-Arcos et al.

2015; Schroeder et al. 2015) for ancient libraries with vey low initial endogenous content, which allowed a higher resolution in population genetic analyses with less amount of sequencing generated. One caveat is the increase in the number of reads aligning in repetitive regions of the genome, despite the use of low-complexity DNA during hybridization. Furthermore, bait length distribution affects capture efficiency, with longer baits preferentially binding longer DNA fragments (Ávila-Arcos et al. 2015). Since this results in the loss of short endogenous DNA molecules, further research in this area could significantly improve the enrichment efficiency.

c) 1240k in–solution capture

Most endogenous DNA sequences retrieved with the previous methodology will map in regions of the genome that are conserved among all humans, and hence they will not be useful for population genetic analyses. In order to maximize the amount of informative DNA molecules that are sequenced, aDNA libraries can be enriched for subsets of the nuclear genome containing known polymorphic positions. In the 1240k capture, aDNA libraries are enriched for a targeted set of ~1,240,000 SNPs using in-solution capture with biotinylated probes (Mathieson et al. 2015). The targeted SNP set includes all the SNPs on the Affymetrix Human Origins array (Patterson et al. 2012), all SNPs on the Illumina 610-Quad array, all SNPs on the Affymetrix 50k array and a smaller number of SNPs chosen for other purposes such as SNPs identified as targets of selection. For each SNP target, four 52-nucleotide probes are used: two located on each side of the SNP, and two centered on the SNP containing one or the other alternate allele, respectively. During

hybridization, human DNA sequences that cover one of the polymorphic positions in the targeted SNP set will bind to the probes, whereas environmental DNA and human DNA mapping outside the targeted SNPs will be washed. By targeting sites in the genome that will actually be analyzed, this strategy is able to generate a 4-fold increase in median coverage of targeted SNPs as compared to a shotgun approach, while decreasing the mean number of reads generated per sample by 36-fold (Mathieson et al. 2015). When the objective is to discover new variants, this method is not appropriate as it only recovers know polymorphic sites.

2.1.4. Sequencing

All the steps of the Illumina technology take place in a flow cell, which can be partitioned in different lanes. The surface of the flow cell lanes is densely coated with forward and reverse primers, complementary to the adapter sequences introduced during library preparation (Buermans and den Dunnen 2014). The first step for loading the library onto the flow cell is denaturation of the double-stranded library fragments into single-stranded molecules, which are hybridized at one end to the surface primers. Next, for each original library fragment, a cluster of multiple copies is created using bridge amplification (Figure 15). Briefly, the free 3’ end of the fragments bends over and hybridizes to a complementary primer on the surface, forming a bridge structure that allows the synthesis of the complementary strand. The double-stranded bridge structure is denatured and a new amplification cycle begins. One flow cell

contains millions of spatially separated clusters, each with ~1,000 copies of an original library template, sufficient for reporting incorporated bases at the required signal intensity for detection during sequencing.

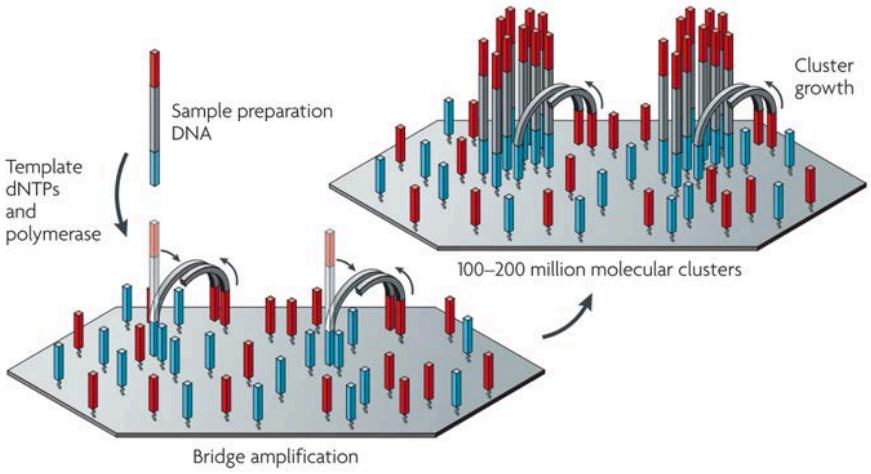


Figure 15. Cluster generation on the Illumina sequencing platform. Modified from Metzker (2010).

Illumina sequencing system is based on a cyclic method that comprises three steps (Figure 16a). First, DNA polymerase incorporates one fluorescently labeled nucleotide, which represents the complement of the template base, to the growing DNA strand. Nucleotides carry a base-unique fluorescent label and a reversible terminator attached to the 3' end that prevents multiple extension events. During the second step, unincorporated nucleotides are washed away and imaging is performed to determine which nucleotide was incorporated. Finally, the fluorescent dye and the 3' terminator are removed to prepare the DNA strand for a new cycle.

Since one nucleotide is incorporated in each cycle, the number of cycles determines the length of the resulting reads (Figure 16b).

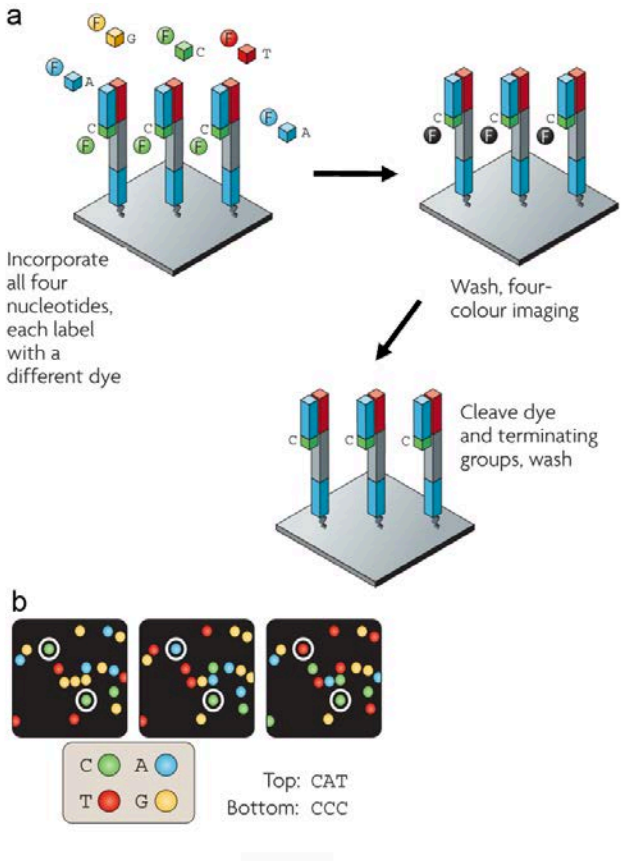


Figure 16. Sequencing on the Illumina platform. a) One cycle of nucleotide incorporation, imaging and cleavage**. b)** Three sequencing cycles highlighted for two clusters. Modified from Metzker (2010).

Sequencing on the Illumina platforms can be run in single-end mode or paired-end mode. If single-end sequencing is performed, only one end of the DNA fragment is sequenced using the primer site present in one of the adapters (Figure 17). During paired-end sequencing, once the sequencing from one end is finished, the complementary strand is synthesized and the sequencing priming site in the second

adapter is used to sequence the other end of the DNA fragment. When libraries with different index sequences are pooled and sequenced in the same run, the index is read in a third sequencing reactions, using a different priming site in the second adapter (Figure 17).

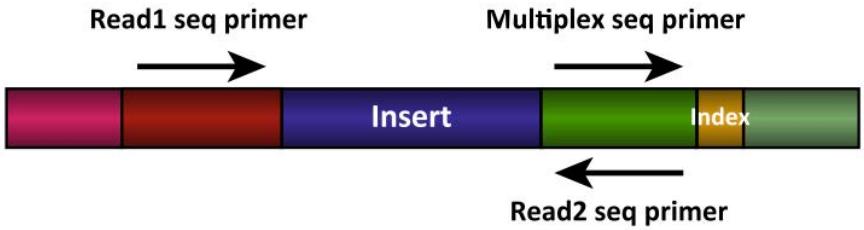


Figure 17. Primer sites used for sequencing both ends of library fragments and the index sequence. Modified from Van Dijk et al. (2014).

2.2. Bioinformatic processing

2.2.1. Removing adapter sequences

Due to post-mortem degradation, DNA molecules extracted from ancient remains are highly fragmented. As a consequence, the length of library inserts (original DNA fragments) is often shorter than the read length, leading to the sequencing of adapter fragments. If these adapter sequences are not removed from the reads before mapping, they can produce missed alignments or an increased rate of mismatches if the reads are mapped to the reference genome.

Many different programs exist for removing adapter sequences from high-throughput sequencing data (Falgueras et al. 2010; Martin 2011; Kircher 2012; Lindgreen 2012), each with different features and additional processing tools. When processing single-end reads, adapters are identified by aligning the expected adapter sequence to the reads. When the adapter sequence present in the read is very short identification becomes difficult, and therefore a minimum length of 5 bp (Kircher 2012) can be imposed to avoid false adapter trimming. Moreover, higher sequencing error rates at the ends of reads further complicates adapter identification. When analyzing paired-end data, adapter identification is simpler and much more sensitive because more information is available (Lindgreen 2012). The paired reads can be aligned and the overlapping region identified, revealing where the insert ends and adapter sequence begins. Depending on the length of the insert, three different situations can occur (Figure 18). If the insert is longer than twice the read length, there will be no adapter

contamination and no overlap between the forward and reverse read (Figure 18a). If the insert is shorter than twice the read length but still longer than the actual read length, there will be overlap between the paired reads at their 3' ends and no adapter contamination (Figure 18b). Finally, if the insert is shorter than the read length, adapter sequences of equal length will be present at the 3' ends of both reads, with a complete overlap in the remaining sequence (Figure 18c). In the last two cases, the forward and reverse reads are collapsed into one read by choosing the most likely nucleotide in case of mismatch, which significantly increases sequence quality (Kircher 2012).

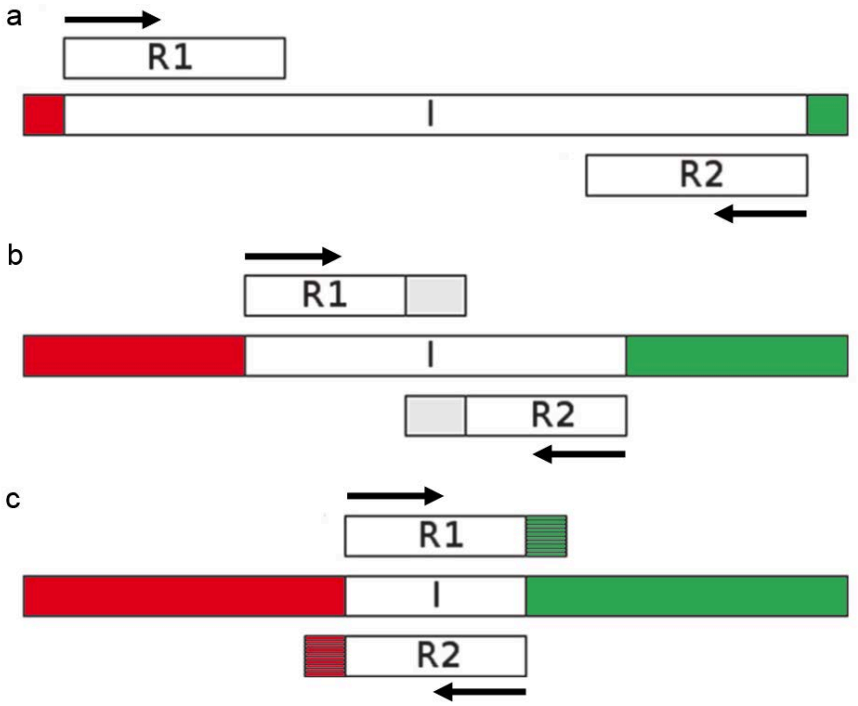


Figure 18. Paired-end sequencing with different insert sizes. Inserts are denoted I and paired-end reads R1 and R2. Read length denoted L_R, insert length denoted L_I. Arrows indicate the direction of sequencing. Adapter sequences are colored in red and green. **a)** $L_I \geq 2*L_R$, **b)** $L_R < L_I < 2*L_R$, **c)** $L_I < L_R$. Modified from Lindgreen (2012).

2.2.2. Mapping

Once adapter sequences have been trimmed, sequencing data is ready to be assembled. Due to the fragmented nature of aDNA, using overlapping short reads to create a *de novo* assembly is highly problematic, and hence most paleogenomic studies map reads to a previously assembled reference genome, either from the same species as the target individual or from closely related species.

An accurate alignment is critical for a reliable identification of genuine aDNA sequences, which in turn affects the quality of the inferences obtained from genetic information. Several software solutions have been developed for mapping SGS reads to reference genomes (Li and Durbin 2009; Li et al. 2009b; Lunter and Goodson 2011; Langmead and Salzberg 2012), but few of them incorporate position-specific scoring matrices that can accommodate aDNA features such as deamination near the ends of the fragments (Briggs et al. 2009; Green et al. 2010; Kerpedjiev et al. 2014).

Currently, the most widely used aligner in paleogenomic projects is BWA (Li and Durbin 2009), which was developed for mapping modern undamaged reads but can be tuned to handle aDNA data. Before mapping, reads shorter than 25-30 bp are usually discarded due to the inherent difficulty of aligning such sequences to reference genomes full of repetitive sequences. Furthermore, to cope with high rates of deamination and increase the number of true endogenous aDNA sequences recovered, two modifications in the default alignment parameters are recommended (Schubert et al. 2012). The

first modification is to disable the seed region, which comprises the 5' end of the sequences and is used to identify possible alignments along the genome. As misincorporations tend to appear towards the ends of DNA fragments, seeding approaches should be avoided (Schubert et al. 2012). The second modification is to increase the maximum edit distance allowed between the read and the reference genome, which improves the recovery of reads with several damage-related misincorporations that would otherwise be lost.

Optimizations in extraction and library preparation protocols over the last few years are progressively enabling the retrieval of very short aDNA reads. Thus, new strategies to confidently align those sequences while discriminating endogenous fragments from contaminating environmental DNA will undoubtedly improve our ability to characterize ancient genomes.

2.2.3. Duplicate removal

After library preparation, DNA libraries are PCR-amplified in order to reach the amount of DNA required for sequencing. In aDNA studies, the number of PCR cycles applied is often higher than in modern DNA experiments due to poor preservation and low DNA content. This leads to an excess of clonal molecules in the sequencing reads, which must be identified and removed as they can severely bias downstream analyses such as coverage estimation or polymorphism detection.

The most common approach for identifying PCR duplicates of the same original DNA fragment is based on the detection of reads mapping on the same strand with the same outer alignment coordinates. This task can be performed with standard bioinformatics tools such as the *rmpdup* tool from Samtools (Li et al. 2009a) or the *MarkDuplicates* program from Picard tools (http://broadinstitute.github.io/picard/). Once PCR clones are detected, these tools select the read with the highest sum of base quality scores, while the rest are either removed from the alignment file or flagged as PCR duplicates.

2.3. Assessing data authenticity

As mentioned in previous sections, DNA from ancient specimens is present in low amounts and is highly degraded. In contrast, modern DNA is ubiquitous and generally undamaged, making modern DNA contamination a major issue for aDNA studies. Furthermore, despite taking rigorous precautions during laboratory procedures, contamination is often inevitable because many ancient remains have been extensively handled during excavation, archaeological research and museum storage. In hominin paleogenomic studies contamination is particularly problematic since endogenous and modern contaminant sequences derive from the same or closely related species. Data authentication is therefore an essential step in the pipeline of every paleogenomic project.

The following sections provide a brief overview of commonly used procedures to evaluate the authenticity of aDNA sequences.

2.3.1. Deamination pattern and read length distribution

The characteristic misincorporation patterns caused by cytosine deamination at overhanging ends of aDNA molecules (Figure 4) can be used to support specimen authenticity (Hofreiter et al. 2001b; Briggs et al. 2009). Deamination rates are easily computed from alignment files and plotted with *MapDamage 2.0* software (Jónsson et al. 2013); if no increase in deamination is observed towards read ends, this could represent an initial indication of extensive modern

DNA contamination in the data. For instance, Meyer et al. (2014) observed that libraries with modern DNA contamination presented less than 5% deamination, while libraries without contamination exhibited higher rates. Nevertheless, this criterion is not straightforward. Although a linear correlation between time and deamination rate across different preservation conditions has been reported (Sawyer et al. 2012), other studies have failed to observe such a clear pattern (Orlando et al. 2013; Rohland et al. 2015). This is likely due to the fact that age is not the only factor driving the amount of deamination observed in an ancient sample. Other factors such as general climatic conditions or specific microenvironments within the remains play an important role in DNA degradation. As a consequence, recent samples from hot environments might present a similar degree of preservation than older samples from temperate regions.

A further strategy to assess the presence of contamination using deamination patterns is to compare analytical results when considering all reads or only damaged reads. The program *PMDtools* (Skoglund et al. 2014b) computes a postmortem degradation score for each read, which can be used to keep the most damaged fraction. If damaged reads (likely endogenous) show, for instance, different population genetic affinities than the full population of reads, it might indicate that sequences from a distinct source are present in the undamaged fraction of reads. This approach is only possible for samples with high amounts of data, as true endogenous sequences without misincorporations will be lost if damage restriction is applied.

Similarly, fragment length distribution has also been suggested as an authentication criterion. True ancient molecules are expected to be very short (Figure 3a), with significant variability between samples due to different preservation conditions. In fact, endogenous and contaminant sequence can show similar read length distributions (Krause et al. 2010a).

Consequently, indirect measures of contamination like those provided by deamination rates and fragment length distributions should be viewed as additional evidence, rather than as strict criteria.

2.3.2. Mitochondrial contamination

The mtDNA is found in high number of copies within cells, and hence many paleogenomic studies are able to recover significant amounts of mitochondrial DNA data. Moreover, since only one haplotype is expected in each individual, the presence of different alleles at a given position can be used to directly estimate a contamination value.

Informative sites for such estimation are those where sequences are private to the studied individual and absent or in low frequency in putative contaminants. If the ancient individual and modern contaminants are distantly related, high number of informative sites will be available. For example, modern human contamination in Neandertal mtDNA data can be computed by counting the number of sequences carrying the modern human allele at positions where the Neandertal mtDNA consensus differs from all modern human

mtDNAs (Green et al. 2010; Prüfer et al. 2014). In the case of modern human paleogenomic studies, a very similar strategy identifies positions where the mtDNA consensus sequence of the ancient individual is different from more than 99% of 311 worldwide modern human mtDNAs (Krause et al. 2010a). At those positions, sequences from contaminant individuals are particularly likely to differ from the ancient mtDNA consensus. A more powerful method (Fu et al. 2013b) takes the aDNA reads, the consensus ancient sequence and a panel of putative contaminant mtDNAs as input data and derives a probabilistic model. Using Markov chain Monte Carlo (MCMC), it provides a maximum likelihood estimation of the mtDNA contamination rate.

2.3.3. X-chromosome contamination in males

The ratio of mtDNA to nuclear DNA is highly variable among individuals (Schwarz et al. 2009). As a result, extrapolating mtDNA contamination estimates to the nuclear genome is problematic. For instance, if a contaminant source presents low amounts of mtDNA relative to the endogenous individual, the mtDNA contamination value can underestimate the true level of contamination in the nuclear genome.

Male individuals are haploid for most X-chromosome regions and thus discordant alleles between overlapping reads result from sequencing or mapping errors, and are distributed randomly along the genome. In the presence of contamination, the number of discordant alleles at polymorphic sites should increase, while this value at

adjacent non-polymorphic positions should not vary, as the ancient individual and the contaminant share the same allele. Therefore, differences in the discordance rate between polymorphic and non-polymorphic sites in the X-chromosome can be used to estimate nuclear contamination. If allele frequencies in the putative contaminating population are introduced in the model, a maximum likelihood estimation of the contamination rate can be obtained (Rasmussen et al. 2011).

2.4. Learning about population history from genetic data

The ultimate force that generates genetic variation is mutation, whereas recombination creates new allele combinations and statistical independence between different loci, which can be seen as independent stochastic outcomes of the same ancestral process. The patterns of genetic variation in extant and past populations are influenced by various factors, such as migration, isolation, admixture, genetic drift and natural selection. The study of those patterns is therefore a powerful tool to reconstruct past demographic and evolutionary events.

In humans, about 85% of all genetic variation is harbored within subpopulations, with the remaining variation found between different subpopulations (Barbujani and Di Benedetto 2001). However, this does not preclude the classification of individuals into different groups using genetic information. In fact, it is the cumulative effect of small allele frequency differences over a large number of loci what allows us to make inferences about human population history with high resolution (Edwards 2003).

In the following sections, I will briefly review different methods applied in this thesis for studying the genetic structure of human populations.

2.4.1. Principal component analysis and clustering algorithms

Principal component analysis (PCA) was one of the earliest statistical approaches to be used for the analysis of multilocus genetic data (Menozzi et al. 1978), and has become an extremely popular tool in the field of paleogenomics. The objective of this method is to extract the most important information in multivariate data where a graphical representation is impossible due to the high dimensionality. Specifically, PCA transforms a series of correlated variables (in our case genetic variants across thousands of loci) into a smaller number of linearly uncorrelated variables called principal components that represent relatedness between individuals. For example, when population structure conforms to an isolation-by-distance model with two spatial dimensions of genetic differentiation, genetic distances will be correlated to geographic distances and PCA will be able to recapitulate the geographic locations of individuals (Novembre et al. 2008) (Figure 19).

This method has several advantages; the projection of samples onto the principal components reveals population structure in a visual manner, it does not assume any specific genetic model and it is computationally fast. Furthermore, PCA projections inform about coalescent rates between individuals (McVean 2009). Despite evident limitations (i.e. different scenario can lead to similar coalescent rates and hence similar projections), this property allows us to make inferences about underlying demographic processes.

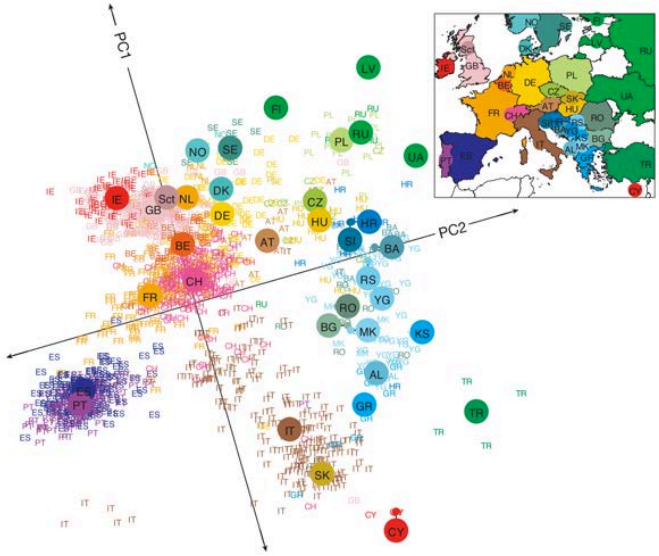


Figure 19. PCA of 1,387 European individuals from the Population Reference Sample (POPRES) project genotyped at 500,568 SNPs. (Novembre et al. 2008)

Another popular tool for analyzing population structure is the model-based clustering method called ADMIXTURE (Alexander et al. 2009). In the model underlying this approach, individuals are assumed to have inherited part of their ancestry from two or more hypothesized ancestral populations that instantly diverged, characterized by a set of allele frequencies at each locus. The goal is to estimate for each individual the proportion of the different ancestral components and the allele frequencies of each ancestral population.

Both PCA and ADMIXTURE are powerful tools for detecting genetic structure and admixture and can be used to formulate interesting hypotheses. However, these methods are very sensitive to the sampling scheme and more importantly, they do not directly reveal the actual history that produced the observed patterns (in fact,

the patterns could be compatible with multiple population histories), nor they provide any formal tests of admixture between populations (Engelhardt and Stephens 2010; Patterson et al. 2012).

2.4.2. *f*-statistics framework

This set of techniques relies on the study of allele frequency correlation patterns across populations and provides formal tests for whether admixture occurred (Patterson et al. 2012). They have been extensively used in paleogenomic studies over the last decade, uncovering several previously unknown admixture events (Green et al. 2010; Reich et al. 2010; Haak et al. 2015). A drawback of these methods is that individuals must be first classified into populations, which can be to a certain extent arbitrary and lead to inaccurate results.

The 3-population test is the least model-based and allows us to detect the presence of admixture in a population *X* from populations *A* and *B*. If population *X* has ancestry from *A* and *B*, the statistic $f_3 = (p_X - p_A)(p_X - p_B)$ averaged across all loci is expected to be negative, where p_X, p_A and p_B are the frequencies of one of the two alleles at each locus for population *X*, *A* and *B*, respectively. The intuition behind this test is the fact that if *X* is admixed from *A* and *B*, allele frequencies in *X* will tend to be intermediate between allele frequencies of *A* and *B*, and thus the product will be negative. Nevertheless, if *X* has experienced strong post-admixture drift, allele frequencies in *X* will be affected by random noise, masking the signal so that the statistic might not be negative. To test for statistical

significance, standard errors can be computed using weighted block jackknife (Busing et al. 1999). A special case is the outgroup f_3-statistic $f_3(A, B ; O)$, which can be used to compute the shared genetic drift between populations *A* and *B* since their divergence from the outgroup *O* (Reich et al. 2009; Raghavan et al. 2014).

The 4-population test is a more powerful approach, but it requires knowledge of the history of divergence between populations. Given the unrooted topology (*A*,*B*)(*C*,*D*) describing the relationship between populations *A*, *B*, *C* and *D*, the following statistic averaged across all loci can be computed, $f_4(A, B ; C , D) = (p_A - p_B)(p_C - p_D)$. If the simple population topology holds (no asymmetrical gene flow after divergence), allele frequency differences between *A* and *B* should not be correlated with allele frequency differences between *C* and *D* and the statistic is expected to be 0 (with the standard error computed as above). In contrast, if there was gene flow between *A* and *C* (or *B* and *D*) allele frequency differences will be correlated and the statistic is expected to be significantly positive. Similarly, gene flow between *A* and *D* (or *B* and *C*) will result in a negative statistic. The *D*-statistic test, also known as ABBA-BABA test, is a very similar version of the 4-population test but uses a different normalization (Patterson et al. 2012). It was first introduced for the study of Neandertal introgression into modern humans (Green et al. 2010) and further developed in later works (Reich et al. 2010; Durand et al. 2011).

The f_4-statistics can also be used to estimate ancestry proportions in an admixed population (Patterson et al. 2012). For example, suppose that we have the following topology ((*A*,*B*)(*C*))(*O*), with population

D being admixed from populations related to *B* and *C*, and *O* being an outgroup population. Then, the proportion of *B*-related ancestry in *D* can be estimated by taking the ratio of two f_4-statistics $f_4(A, O; D, C) / f_4(A, O; B, C)$, averaged over all loci. This statistic can be thought as how strong is the skew in the allele frequency of *D*, with respect to the skew observed for 100% admixture (*B*).

Recently, Haak and colleagues (Haak et al. 2015) developed an extension of the *f*-statistic framework, implemented in the *qpWave* and *qpAdm* programs (https://github.com/DReichLab). The aim of these methods is to identify populations that might have contributed ancestry to a given population and to estimate mixture proportions. Indeed, they have been successfully applied in paleogenomic studies (Haak et al. 2015; Mathieson et al. 2015) to obtain proportions of different ancestral components in ancient West Eurasian individuals. The idea is to relate a *Test* population (our population of interest) to a set of *Outgroup* populations via a set of *Reference* populations. There are three main prerequisites. First, the *References* should not be equally related to the panel of *Outgroups*, i.e. they must share different amounts of genetic drift with them as a result of their deep evolutionary history. Second, the *References* must be clades, with respect to the *Outgroups*, of the populations that contributed ancestry to *Test*. Third, the *Outgroups* must be related to *Test* via the *References*, not directly. Given this setup, we can write f_4-statistics of the form $f_4(Test, O_1; O_2, O_3)$, where O_1,O_2 and O_3 are different *Outgroups*, as a weighted sum of the statistics of the *Reference* populations:

$$f_4(Test, O_1; O_2, O_3) = \sum_{i=1}^{N} \alpha_i f_4(Ref_i, O_1; O_2, O_3)$$

Given a set of n *Outgroups*, there are $n\binom{n-1}{2}$ possible arrangements of the *Outgroups*, resulting in many different equations that can be used to estimate the α_i mixture proportions. The main advantage is that these statistics are affected only by drift shared between the *References* and the *Outgroups*, not by drift specific to the *Reference* populations or the *Test*. Therefore, we do not need to use as *Reference* the exact population that was involved in the admixture event. Instead, we can use a very drifted population, as long as it forms a clade with the admixing population. Another aspect of this methodology that makes the inferences robust is the fact that the phylogeny relating the different *Outgroups*, which can be complex, does not need to be known.

2.4.3. Tree-based modeling of population relationships

A different class of methods relies on the relationships between populations, with the underlying assumption that those relationships conform to a simple bifurcating tree (Cavalli-Sforza and Edwards 1967; Nielsen et al. 1998). Under these models, populations become instantaneously isolated from each other and genetic drift is the force that increases the differentiation between them. With this approach a large number of populations can be included in the model, but an important limitation is evident; since migration and gene flow after

divergence are widespread in most species, a simple tree is not a good representation of population histories.

To address this issue, several methods that fit models of both population splits and gene flow to genetic data have been developed, such as *qpgraph* (Reich et al. 2009), *MixMapper* (Lipson et al. 2013) and *Treemix* (Pickrell and Pritchard 2012). While the first method requires manual exploration of model space, the last two automatically explore the space of possible models. *Treemix*, as an example of these approaches, begins by inferring a maximum likelihood tree of population relationships without admixture from genome-wide allele frequency data. Then, pairs of populations showing a poor fit to the tree model are identified. This is done by looking at the residuals between the empirical allele frequency covariances and the ones predicted by the tree. Finally, migration edges between populations with maximum residuals are added to the model in order to improve the fit. This method has been shown to recapitulate known relationships between well-studied populations (Pickrell and Pritchard 2012) and is useful to investigate the processes by which population structure arose.

3. OBJECTIVES

The general objective of this work is to produce genome-wide aDNA data to learn about the recent human population history of Europe. Specifically:

Genomic data from two Mesolithic hunter-gatherers from La Braña-Arintero cave in Spain are used to:

1. Characterize the pre-Neolithic genetic substratum of the Iberian Peninsula.

2. Study the genetic affinities between European hunter-gatherers.

3. Test competing hypotheses regarding the nature of the Neolithic expansion into Europe.

4. Identify adaptive changes that occurred during the transition from a foraging to a farming lifestyle.

We sequence the complete genome of an Early Neolithic Cardial female from Cova Bonica in Spain in order to:

1. Study the genomic landscape of the Iberian Peninsula before and after the advent of farming.

2. Investigate the genetic relationships between the Mediterranean and Central European routes of Neolithic expansion into Europe.

Genome-wide capture data from Bell Beaker individuals covering the entire geographic distribution of this Late Neolithic cultural complex is used to:

1. Understand population transformations in Europe during the onset of metallurgy.

2. Shed light into the origin and expansion of the Bell Beaker phenomenon.

3. Investigate the role of Bell Beakers in the formation of present-day European genomic landscape.

4. RESULTS

4.1. Genomic affinities of two 7,000-year-old Iberian hunter-gatherers

Federico Sánchez-Quinto, Hannes Schroeder, Oscar Ramirez, María C. Ávila-Arcos, Marc Pybus, Iñigo Olalde, Amhed M.V. Velázquez, María Encina Prada Marcos, Julio Manuel Vidal Encinas, Jaume Bertranpetit, Ludovic Orlando, M. Thomas P. Gilbert, and Carles Lalueza-Fox

Current Biology. 2012; 22:16

DOI: http://dx.doi.org/10.1016/j.cub.2012.06.005

Sánchez-Quinto F, Schroeder H, Ramirez O, Avila-Arcos MC, Pybus M, Olalde I, Velazquez AM, Marcos ME, Encinas JM, Bertranpetit J, Orlando L, Gilbert MT, Lalueza-Fox C. Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. Curr Biol. 2012 Aug 21;22(16):1494-9.
doi: 10.1016/j.cub.2012.06.005

4.2. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European

Iñigo Olalde, Morten E. Allentoft, Federico Sánchez-Quinto, Gabriel Santpere, Charleston W. K. Chiang, Michael DeGiorgio, Javier Prado-Martinez, Juan Antonio Rodríguez, Simon Rasmussen, Javier Quilez, Oscar Ramírez, Urko M. Marigorta, Marcos Fernández-Callejo, María Encina Prada, Julio Manuel Vidal Encinas, Rasmus Nielsen, Mihai G. Netea, John Novembre, Richard A. Sturm, Pardis Sabeti, Tomàs Marquès-Bonet, Arcadi Navarro, Eske Willerslev, and Carles Lalueza-Fox

Nature. 2014; 507(7491): 225-8

DOI: 10.1038/nature12960

Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CW, DeGiorgio M, Prado-Martinez J, Rodríguez JA, Rasmussen S, Quilez J, Ramírez O, Marigorta UM,Fernández-Callejo M, Prada ME, Encinas JM, Nielsen R, Netea MG, Novembre J, Sturm RA, Sabeti P, Marquès-Bonet T, Navarro A, Willerslev E, Lalueza-Fox C. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature. 2014 Mar 13;507(7491):225-8. doi: 10.1038/nature12960.

4.3. A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures

Iñigo Olalde, Hannes Schroeder, Marcela Sandoval-Velasco, Lasse Vinner, Irene Lobón, Oscar Ramirez, Sergi Civit, Pablo García Borja, Domingo C. Salazar-García, Sahra Talamo, Josep María Fullola, Francesc Xavier Oms, Mireia Pedro, Pablo Martínez, Montserrat Sanz, Joan Daura, João Zilhão, Tomàs Marquès-Bonet, M. Thomas P. Gilbert, and Carles Lalueza-Fox

Molecular Biology and Evolution. 2015; 32(12): 3132–3142

DOI: 10.1093/molbev/msv181

Olalde I, Schroeder H, Sandoval-Velasco M, Vinner L, Lobón I, Ramirez O, Civit S, García Borja P, Salazar-García DC, Talamo S, María Fullola J, Xavier Oms F, Pedro M, Martínez P, Sanz M, Daura J, Zilhão J, Marquès-Bonet T, Gilbert MT, Lalueza-Fox C. A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. Mol Biol Evol. 2015 Dec;32(12):3132-42. doi: 10.1093/molbev/msv181

4.4. The Bell Beaker Complex and the shaping of the Western European genomic landscape

Iñigo Olalde, Alissa Mittnik, Swapan Mallick, Nadin Rohland, Eadaoin Harney, Kristin Stewardson, David Billoin, Anthony Denaire, Joan Frances Farré, Michiel Gazenbeek, Philippe Lefranc, Olivier Lemercier, Arnaud Lefebvre, Tona Majó, Pierre-Jérôme Rey, Joël Serralongue, Philipp W. Stockhammer, Luc Vergnaud, Joao Zilhao, Johannes Krause, Nick Patterson, Wolfgang Haak, Carles Lalueza-Fox, and David Reich

In preparation

The Bell Beaker Complex and the shaping of the Western European genomic landscape

Iñigo Olalde[1], Alissa Mittnik[2,3], Swapan Mallick[4,5,6], Nadin Rohland[4], Eadaoin Harney[4,5,6], Kristin Stewardson[4,5,6], David Billoin[7], Anthony Denaire[8], Joan Frances Farré[9], Michiel Gazenbeek[10], Philippe Lefranc[11], Olivier Lemercier[12], Arnaud Lefebvre[13], Tona Majó[14], Pierre-Jérôme Rey[15], Joël Serralongue[16], Philipp W. Stockhammer[17,18], Luc Vergnaud[19], Joao Zilhao[20,21,22], Johannes Krause[2,3,23], Nick Patterson[5], Wolfgang Haak[24], Carles Lalueza-Fox[1]*, and David Reich[4,5,6]*

*Co-senior authors
†Correspondence should be addressed to: D. R. (reich@genetics.med.harvard.edu)

[1] Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain
[2] Institute for Archaeological Sciences, University of Tubingen, D-72070 Tübingen, Germany
[3] Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany
[4] Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA
[5] Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA
[6] Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA
[7] INRAP Grand-Est-Sud et UMR 6298 ArTeHiS, Dijon, France
[8] Antea-Archéologie (Habsheim, France) et UMR 7044, Strasbourg, France
[9] Ajuntament de Cerdanyola, Barcelona, Spain
[10] INRAP Grand-Est-Nord, Metz, France et UMR 6130 CEPAM Nice, France
[11] INRAP Grand-Est-Sud, Strasbourg, France et UMR 7044, Strasbourg, France
[12] Département d'Histoire de l'Art et Archéologie Université de Bourgogne; UMR 6298 ArTeHiS (UB, CNRS, Ministère de la Culture, INRAP)
[13] INRAP Grand-Est-Nord, Metz, France et UMR 5199 PACEA-LAPP Bordeaux, France
[14] ARCHAEOM, Departament de Prehistòria, UAB, Barcelona, Spain
[15] UMR 5254 Edytem, Le Bourget du Lac, France
[16] Service d'Archéologie Départemental de Haute-Savoie, Annecy, France

[17] Cluster of Excellence “Asia and Europe in a Global Context”, Heidelberg University, Heidelberg, Germany
[18] Institute for Prehistory and Early History and Near Eastern Archaeology, Heidelberg University, Heidelberg, Germany
[19] Antea-Archéologie, Habsheim, France
[20] Seminari Estudis i Recerques Prehistòriques (SERP; SGR2014-00108). Dept. Prehistòria, H. Antiga i Arqueologia. Facultat de Geografia i Història. Universitat de Barcelona. 08001, Barcelona, Spain
[21] Centro de Arqueologia. Universidade de Lisboa (UNIARQ). Faculdade de Letras. Alameda da Universidade. 1600-214, Lisboa, Portugal
[22] Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain
[23] Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, 72072 Tübingen, Germany
[24] Australian Centre for Ancient DNA, School of Biological Sciences and The Environment Institute, University of Adelaide, Adelaide, South Australia 5005, Australia

The Bell Beaker Complex (BBC) was the first widely distributed archaeological culture of Western Europe, arising around 2,800 BCE in Iberia and spreading to the north and east before disappearing ~1,800 BCE. A major open question is the extent to which the BBC spread through movement of ideas or people. We assembled genome-wide data on 63 BBC individuals to provide a genomic characterization of the BBC across its geographic and temporal range. In contrast to people of the Corded Ware Complex (CWC) who were the counterpart of the BBC in the east and whose spread was achieved by mass migration and replacement of local populations, the spread of the BBC into central Europe was mediated at least in part by communication of ideas, as skeletons found in BBC contexts in Iberia are genetically very different from those in Central and Northern Europe. However, migration played a key role in other aspects of the BBC dispersal, as in Britain, where we report the first data from the pre-BBC period (n=3 samples), we show that its arrival was associated with a replacement of the local population as complete as what occurred with the advent of farming.

At the beginning of the third millennium BCE, two widespread archaeological cultures spread across Europe, replacing the more localized cultures that preceded them[1]. The Corded Ware Complex (CWC) in the east was associated with skeletons that are genetically homogeneous over a broad range, consistent with deriving most of their ancestry from Yamnaya steppe pastoralists[2–4]. Although the CWC itself did not disperse further than central Europe, the migration from the steppe had a long-term genetic impact in Europe, as today, about half the ancestry of northern Europeans and a large part of southern European ancestry comes from the steppe[2–4]. A potential candidate for the continuation of the steppe dispersal is the Bell Beaker Complex (BBC), which spread as dramatically in the west as the CWC did in the east, with the oldest sites being in Iberia after ~2800 BCE[5–7]. The BBC spread into Central Europe[5,8,9] by ~2500 BCE where it overlapped with the CWC, and from there it expanded to other areas including the British Isles. The BBC is defined by a mixture of artefacts including stylized bell-shaped pots, copper daggers, and equipment related to archery including arrowheads, stone wristguards, and v-perforated buttons[10]. Archaeologically, debate around the BBC has centered on whether the expansion of this culture was mediated by the movement of people, ideologies, or a mixture of factors[11–16]. Genome-wide data of BBC skeletons in Germany and the Czech Republic has revealed that they have large amounts of steppe ancestry, consistent with being a mixture of people related to the CWC, and people with ancestry similar to the preceding farming cultures of Europe[2–4]. However, a deeper understanding of

the ancestry of the BBC requires characterization of specimens throughout its temporal and geographic range.

Ancient DNA data set and authenticity

To understand the genetic structure of people assigned to the BBC and their relationship to preceding, contemporary, and subsequent peoples, we enriched ancient DNA libraries for sequences overlapping 1.2 million single nucleotide polymorphisms (SNPs) by hybridizing the DNA with previously synthesized 52 nucleotides probes targeting these positions (1240k capture), using previously described protocols[4,17]. We filtered out samples with low coverage (<30,000 SNPs) or evidence of contamination based on variability in mitochondrial DNA sequences or on the X chromosome (in males). We also identified first-degree relative pairs, and for each pair kept in the analysis dataset the individual with the larger number of SNPs. We generated new data for 44 ancient individuals (Supplementary Information, section 1), and combined it previously published BBC data[3,4] from 19 samples from Germany and Hungary. The combined dataset included specimens from Iberia (n = 17), Eastern France (n = 7), Italy (n = 1), Britain (n = 11), Hungary (n = 2), Germany (n = 21) and the Czech Republic (n = 4) (Supplementary Data Table 1) (Fig. 1). We combined this with 186 specimens from other ancient populations (Supplementary Data Table 2), of which 183 have been previously published[3,4,18–26] (we increase coverage for 10 of them), while 3 are newly reported in this study and represent English samples from the Middle Neolithic (3,360-3,100 BCE). When analyzing only ancient individuals, we used 1,054,671 autosomal SNPs. When analyzing ancient individuals together with 2,572

modern individuals genotyped on the Affymetrix Human Origins array[20] we analyzed 591,642 SNPs.

Y chromosomes analysis

We determined Y-chromosome haplogroups for the 29 male BBC individuals (Supplementary Information, section 5). The BBC individuals from the Iberian Peninsula carried Y sequences that were common across Europe during the earlier Neolithic period[2,4,18,24,27,28]: I2a2 (n = 2 excluding relatives) and G2 (n = 1) (Supplementary Table 2). In contrast, BBC samples outside Iberia largely carried R1b lineages (86%, n = 22), which have been associated with the arrival of eastern migrants of ultimate steppe origin into Europe at the end of the Neolithic period and beginning of the Bronze Age[2,3]. In cases where we could determine the R1b subtype, we found that all (n=10) were positive for the R1b-S116/P312 mutation, the dominant R1b type in Western Europe today with frequency peaks in Ireland and the Franco-Cantabrian region[29]. Its widespread presence in Central and Northern European samples of our study suggests a role for the BBC in the dissemination of this lineage throughout Western Europe.

Insights into the BBC origin and expansion from genome-wide data

We carried out Principal Component Analysis (PCA) of all ancient individuals and present-day West Eurasians, which revealed high genetic heterogeneity within the BBC, contrasting with the homogeneity of the CWC. While most of the Iberian samples and one sample from eastern France cluster with preceding Middle Neolithic Europeans and contemporaneous Chalcolithic populations from El

Mirador and El Portalón caves in Atapuerca, Spain (Fig. 2a), Central and Northern European Bell Beaker groups plot near the CWC, with the remaining samples at intermediate positions along the CWC-Early European farmer axis of variation. We obtained qualitatively consistent inferences using ADMIXTURE model-based clustering[30] (Fig. 2b). The affinity to the CWC in a subset of the BBC people from Northern and Central Europe suggested that the main cause of genetic differentiation between Bell Beaker groups is the presence of variable amounts of steppe-related ancestry, which we verified using the statistic f_4*(Chimp, Test; Yamnaya_Samara, Anatolia_Neolithic)*. This statistic measures the degree of allele sharing of a *Test* ancient DNA sample with *Yamnaya* steppe pastoralists compared to *Anatolia_Neolithic* farmers that we use to represent the type of ancestry that was prevalent in Europe prior to the migrations from the east[4]. There is striking variation in the proportions of steppe ancestry among BBC individuals not only at a pan-European scale (more in the east than in the west), but within regions (e.g. different sites within the Haut-Rhin region of France), and within sites (e.g. Budapest-Békásmegyer site in Hungary). This is also observed at the Arroyal I in Northern Spain, where two BBC have more affinity to Yamnaya than the other two (the latter are similar to other Iberian BBC samples and contemporary Chalcolithic Iberians in having little or no steppe affinity (Fig. 3a)). Thus, our results point to a model where the BBC context of many burials reflects people with mostly CWC-related ancestry.

To increase statistical power for subsequent analyses, we grouped individuals based on their geographic location and degree of steppe

affinity (Supplementary Table 4 and Supplementary Information, section 8) and used *qpAdm*[2] to fit the ancestry of BBC associated skeletons into models with estimated mixture coefficients (Supplementary Information, section 9). First, we modelled the BBC groups from Iberia and France that had no evidence of steppe ancestry based on the qualitative analyses (*BB_Iberia_Cer, BB_Iberia_Alm, BB_Iberia_Arr1* and *BB_France_Heg*) as a mixture of *Anatolia_Neolithic* populations and Western hunter-gatherers (WHG) representing the indigenous people of Europe[4]. All these groups fit the model ($P > 0.18$), with *BB_France_Heg* showing less WHG ancestry than Iberian groups (Table 1). In light of these findings, we tested for genetic continuity in Iberia between BBC (taking *BB_Iberia_Cer* as representative) and preceding Middle Neolithic and Chalcolithic populations. Symmetry tests using f_4-statistics (Fig. 4a-b) indicate that they are consistent with descending from a common ancestral population with no external immigration, an inference that is also supported by shared genetic drift analysis using outgroup-f_3 statistics (Fig. 4c). The finding that the Iberian BBC are genetically continuous with the preceding and contemporary non-BBC cultures is consistent with the proposed origin of the BBC in the Iberian Peninsula[5–7,31].

We next examined the ancestry of BBC samples with steppe affinities. PCA (Fig. 2a), ADMIXTURE (Fig. 2b) and the statistic f_4(*Chimp*, *Test*; *Yamnaya_Samara*, *Anatolia_Neolithic*) (Fig. 3a) show that many individuals in the study are genetically close to individuals attributed to the CWC, but differ from them in having less steppe-related ancestry (e.g. f_4(*Chimp*, *Yamnaya_Samara*;

Corded_Ware_Germany, *BB_France_Mon*) = -0.0015, Z = -4.48)) (Supplementary Table 5). We used *qpAdm*[2] to show that they can be well modelled as a mixture of *Corded_Ware_Germany* and a Neolithic farmer population without steppe-related ancestry (Fig. 3b, Supplementary Table 6), which could be either local (e.g. central European Middle Neolithic) or Iberian Bell Beaker; our analyses do not have enough resolution to distinguish these scenarios. Thus, a substantial proportion of the Central European BBC (and possibly up to 100%) was local in origin, showing how the movement of the BBC into central Europe to a large extent reflected movement of ideas.

A nearly complete turnover of the British population at the time of the BBC

A striking finding of this study concerns the ancestry turnover that occurred in association with the British BBC samples (n=11). We show that the British BBC had a similar proportion of steppe ancestry as the Central European BBC (Fig. 3; Supplementary Table 6) to which it had an archaeological affinity (both were associated with the "All Over Corded" subtype of the BBC). The presence of steppe ancestry in the British BBC sharply contrasts with the Neolithic Britons in our study (n=3) who have no evidence of steppe ancestry (Fig. 5c; Supplementary Information, section 10). Thus, the arrival of steppe ancestry in Britain is likely to have been mediated by a large-scale migration that also brought the BBC complex; a previous study showed that steppe ancestry arrived in Ireland by the Bronze Age[26], and here we show that at least in Britain it arrived by the earlier Bell Beaker period. Beyond Britain, we also document an association between the BBC and the earliest evidence of steppe

affinities in Northern Italy (Fig. 3 and 5b; Supplementary Information, section 8) and Northern Iberia, where two of the BBC samples from the Arroyal I site present a signal of steppe ancestry (Fig. 3 and 5a; Supplementary Information, section 8).

Conclusions

Our study has shown how, in sharp contrast to the arrival of the CWC that represented a genetically homogeneous population and whose arrival involved a significant population turnover, in the case of the BBC cultural diffusion played an important role. Some parts of the BBC expansion, however, were driven to a substantial extent by migration, most notably the BBC expansion in Britain where migrants bearing the BBC nearly completely replaced the local population. The fact that the BBC expansion to Britain was so strongly associated with migration suggests the possibility that the subtype of the BBC that expanded into Britain—the "All Over Corded" subtype that developed in the part of central Europe where the BBC and CWC overlapped and interacted with each other—may have been fundamentally different in its ideology from its Iberian predecessors. While the traditional way of viewing the All Over Corded subtype of the BBC is as a BBC culture, it is also possible to view it as a westward extension of the CWC, continuing in the northwest the dramatic population replacements its predecessors had achieved in central Europe (the CWC) and eastern Europe (the Yamnaya). An important direction for future work is to develop more sensitive methods to track the differences among Bell Beaker populations with similar amounts of steppe ancestry, which should make it possible to track the paths of migration of the late BBC within

Europe, and to understand the way in which particular BBC subtypes spread. It is also important to correlate these genetic results with other types of evidence—in particular about changes in technology, climate, and pathogen exposure—in order to obtain better insight about what made it possible for the central European BBC to largely replace the population of Britain that existed prior to their arrival.

References

1. Czebreszuk, J. in *Ancient Europe, 8000 B.C. to A.D. 1000: An Encyclopedia of the Barbarian World* (eds. Bogucki, P. I. & Crabtree, P. J.) 476–485 (Charles Scribner's Sons, 2004).
2. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).
3. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522,** 167–172 (2015).
4. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).
5. Bailly, M. & Salanova, L. Les dates radiocarbone du Campaniforme en Europe Occidentale : Analyse critique des principales séries de dates. *Mémoires la Société préhistorique française* **26,** 219–224 (1999).
6. Müller, J. & van Willigen, S. New radiocarbon evidence for European Bell Beakers and the consequences for the diffusion of the Bell Beaker phenomenon. in *Bell beakers today: pottery, people, culture, symbols in prehistoric Europe. proceedings of the International colloquium, Riva del Garda (Trento, Italy)* (ed. Nicolis, F.) 59–80 (2001).
7. Harrison, R. J. & Martin, A. M. Beakers and Social Complexity in Central Spain. in *Bell beakers today: pottery, people, culture, symbols in prehistoric Europe. proceedings of the International colloquium, Riva del Garda (Trento, Italy)* (ed. Nicolis, F.) 111–124 (2001).
8. Harrison, R. & Heyd, V. The Transformation of Europe in the Third Millennium BC: the example of 'Le Petit-Chasseur I +

III' (Sion, Valais, Switzerland). in *Praehistorische Zeitschrift* **82,** (2007).

9. Lemercier, O. Historical model of setting and spreading out of the Bell Beaker culture in Mediterranean France. in *Similar But Different: Bell Beakers in Europe* (ed. Czebreszuk, J.) 193–205 (2004).
10. Czebreszuk, J. *Similar But Different. Bell Beakers in Europe*. (Sidestone Press, 2004).
11. Grupe, G. *et al.* Mobility of Bell Beaker people revealed by strontium isotope ratios of tooth and bone: a study of southern Bavarian skeletal remains. *Appl. Geochemistry* **12,** 517–525 (1997).
12. Price, T. D., Knipper, C., Grupe, G. & Smrcka, V. Strontium Isotopes and Prehistoric Human Migration: The Bell Beaker Period in Central Europe. *Eur. J. Archaeol.* **7,** 9–40 (2004).
13. Vander Linden, M. What linked the Bell Beakers in third millennium BC Europe? *Antiquity* **81,** 343–352 (2007).
14. Fitzpatrick, A. P. *The Amesbury Archer and the Boscombe Bowmen: Bell Beaker burials on Boscombe Down, Amesbury, Wiltshire*. (Wessex Archaeology, 2011).
15. Bogucki, P. I. & Crabtree, P. J. *Ancient Europe, 8000 B.C. to A.D. 1000: An Encyclopedia of the Barbarian World*. (Charles Scribner's Sons, 2004).
16. Lemercier, O. Interpreting the Beaker phenomenon in Mediterranean France: an Iron Age analogy. *Antiquity* **86,** 131–143 (2012).
17. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* (2015). doi:10.1038/nature14558
18. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3,** 698 (2012).
19. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
20. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,**

409–413 (2014).

21. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346,** 1113–1118 (2014).

22. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).

23. Olalde, I. *et al.* A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Mol. Biol. Evol.* **32,** 3132–3142 (2015).

24. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl. Acad. Sci.* 201509851 (2015). doi:10.1073/pnas.1509851112

25. Jones, E. R. *et al.* Upper palaeolithic genomes reveal deep roots of modern eurasians. *Nat. Comm.* 1–8 (2015). doi:10.1038/ncomms9912

26. Cassidy, L. M. *et al.* Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc. Natl. Acad. Sci.* **113,** 1–6 (2016).

27. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5,** 5257 (2014).

28. Skoglund, P. *et al.* Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science* **201,** 786–792 (2014).

29. Valverde, L. *et al.* New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia. *Eur. J. Hum. Genet.* 1–5 (2015). doi:10.1038/ejhg.2015.114

30. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).

31. Case, H. Beakers and the Beaker culture. in *Similar but Different. Bell Beakers in Europe* (ed. Czebreszuk, J.) 11–34 (2004).

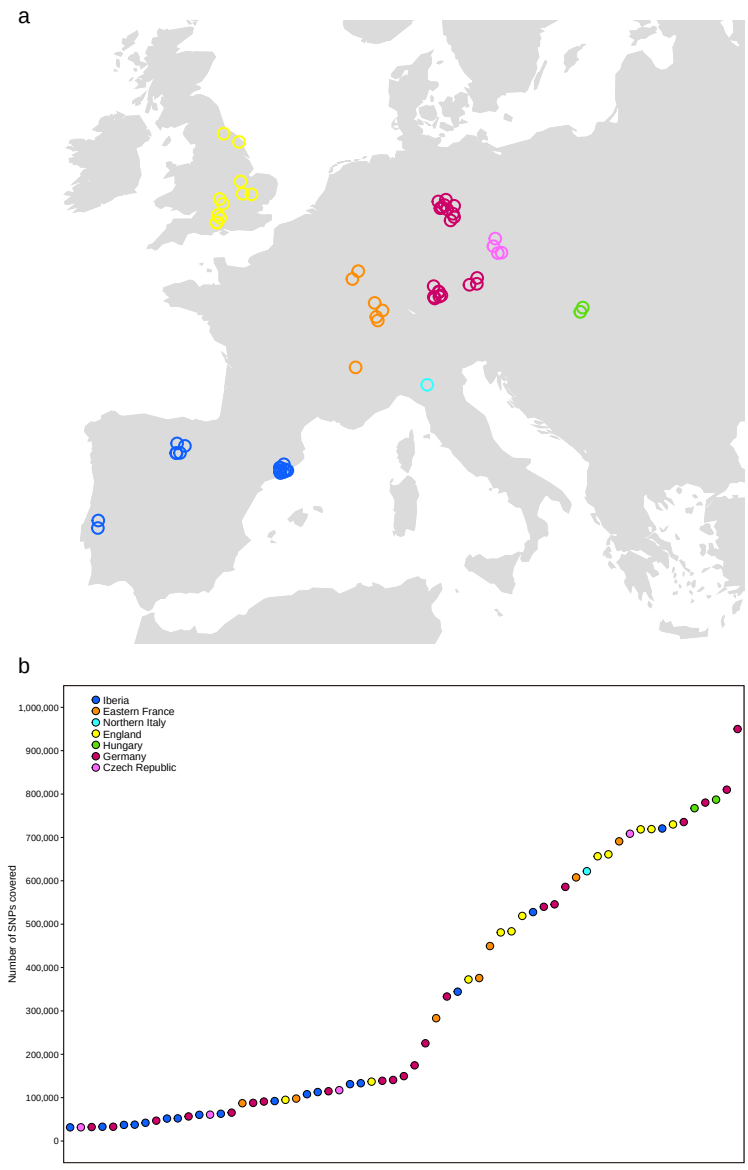


Figure 1. BBC individuals analysed in this study. a) Geographic distribution of sampling locations. Individuals excavated from the same site are jittered for clarity. **b)** Number of 1240k SNPs covered at least once.

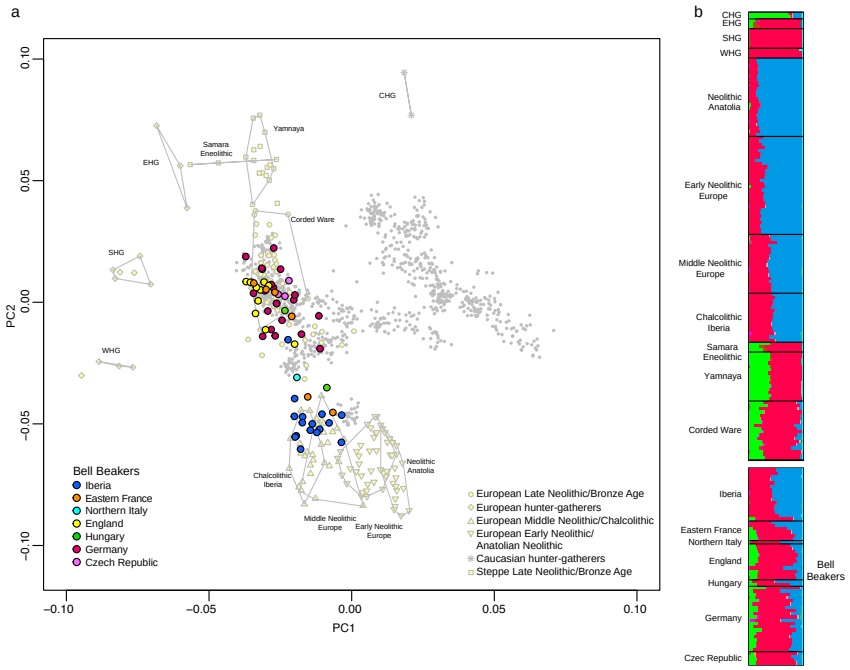


Figure 2. BBC population structure. a) Principal component analysis of BBC individuals, other ancient individuals and present-day West Eurasian individuals (grey dots). **b)** ADMIXTURE clustering analysis with k=11.

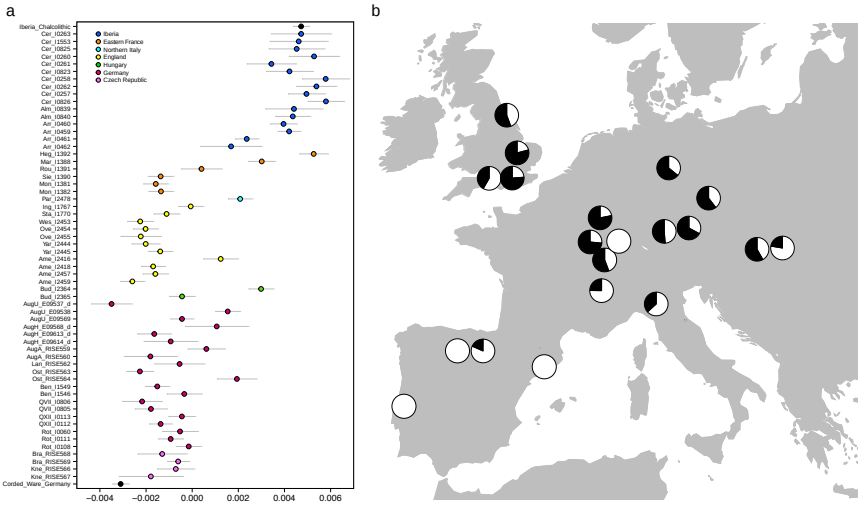


Figure 3. Steppe affinities across the Bell Beaker horizon. a) *f*-statistics of the form f_4(Chimp, Test; Yamnaya_Samara, Anatolia_Neolithic). Bell Beaker individuals are identified by their Site ID and Lab ID (Supplementary Data Table 1). Error bars represent 1.5 standard errors. **b)** Geographic distribution of Corded Ware-related ancestry in Bell Beaker groups (in black). Proportions were obtained from Supplementary Table 6.

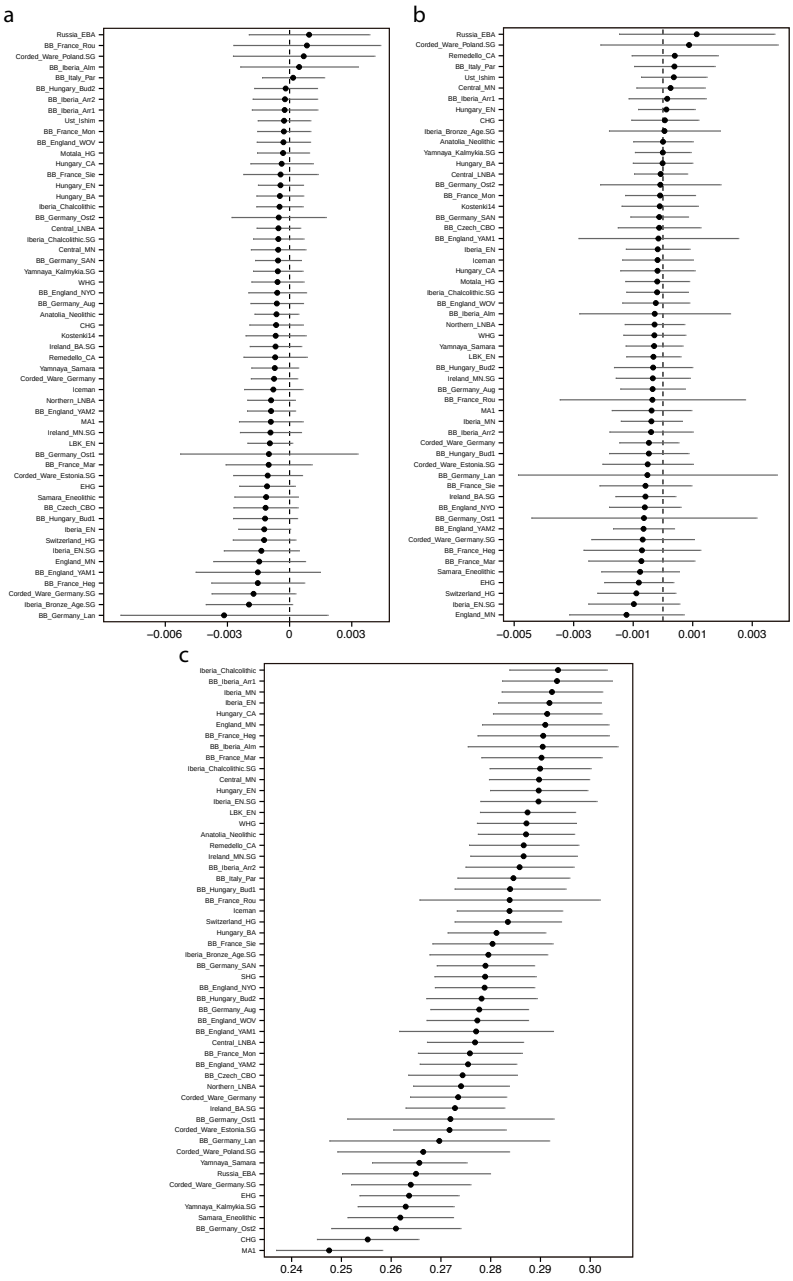


Figure 4. Continuity between Iberian BBC and preceding populations. a) Symmetry tests of the form f_4(*Chimp*, *Test*; *Iberia_MN*, *BB_Iberia_Cer*). **b)** Symmetry tests of the form f_4(*Chimp*, *Test*; *Iberia_Chalcolithic*, *BB_Iberia_Cer*). **c)** Shared genetic drift between *BB_Iberia_Cer* and ancient populations, as measured by f_3(*Mbuti*; *BB_Iberia_Cer*, *Test*).

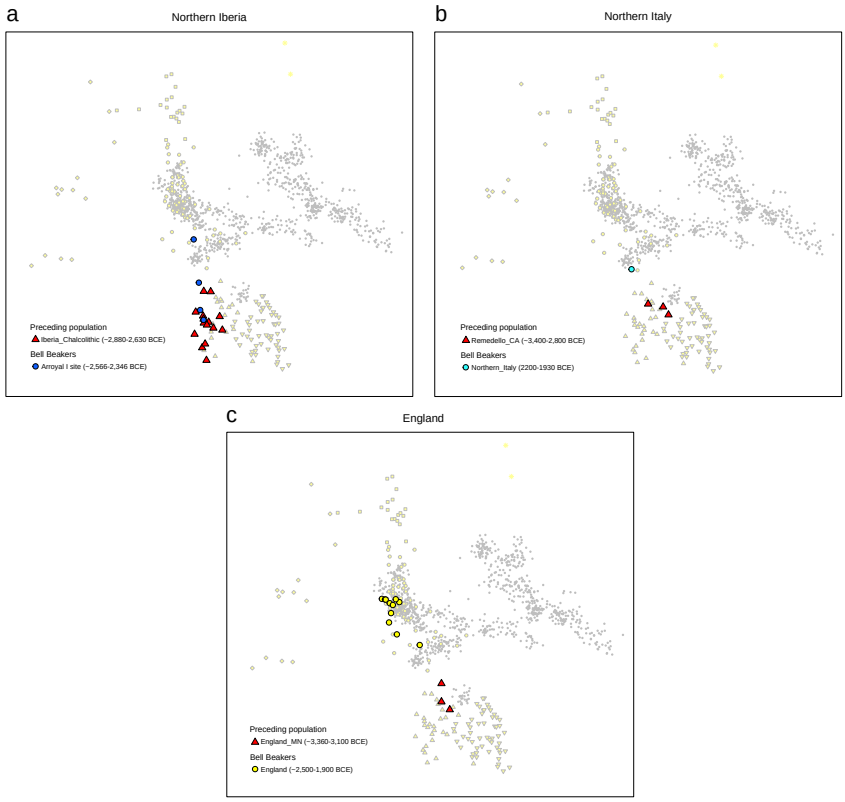


Figure 5. BBC introduced steppe ancestry in Northern and Southern Europe. Principal component analysis contrasting BBC individuals with preceding populations in **a)** Northern Iberia, **b)** Northern Italy and **c)** England.

Table 1. Modelling Bell Beaker groups without steppe ancestry as a mixture of *Anatolia_Neolithic* and *WHG*.

		Mixture proportions		
	P for rank=1	***Anatolia_ Neolithic***	***WHG***	**Std. Error**
BB_Iberia_Cer	1.78E-01	0.770	0.230	0.046
BB_Iberia_Alm	4.27E-01	0.554	0.446	0.096
BB_Iberia_Arr1	2.64E-01	0.685	0.315	0.047
BB_France_Heg	1.96E-01	0.921	0.079	0.059

5. DISCUSSION

In the following section, I will first comment on the different sequencing strategies used in section 4. Then, I will discuss the relevance of the results presented herein for our understanding of the human population history of Europe during the Holocene. This period witnessed crucial events that radically transformed human lifestyle and whose consequences we are still experiencing today (Bogucki and Crabtree 2004; Whittle and Cummings 2007). In this thesis, I have focused on two of these events: the transition to farming beginning around 8,500 years ago and the expansion of the Bell Beaker phenomenon, part of a broader set of transformations taking place in Europe during the third millennium BCE that paved the way to the emergence of Bronze Age societies.

From an archeological perspective, long-standing debates around both the Mesolithic-Neolithic transition and the Bell Beaker expansion have centered on whether these major cultural shifts represented movements of people (demic diffusion), ideas (cultural diffusion) or a mixture of both (Childe 1925; Barker 1985; Whittle 1996; Gallay 2001; Lemercier 2004). Academic views towards these controversies have been influenced by the political, social and intellectual context of the time, as well as by the historical and geographical context of the countries with more tradition of archeological research (Härke 1998).

The use of genetic data from ancient individuals allowed us to address these questions with a much more powerful and robust

approach, as DNA provides essential information for studying the origin and affinities of the people directly involved in a given demographic process.

5.1. Shotgun versus capture approaches

Throughout the works I have presented here, different sequencing strategies were applied. In sections 4.1 and 4.2 we performed shotgun sequencing on teeth samples from La Braña 1 and La Braña 2 individuals. Thanks to exceptional DNA preservation in a molar tooth from La Braña 1 (~50% of the sequences mapped to the human reference genome), we obtained 3.4x depth of coverage for the nuclear genome. Most ancient samples, however, contain small amounts of endogenous DNA and are not amenable to shotgun sequencing. In fact, other samples from La Braña 1 and La Braña 2 turned out to contain only trace amounts of endogenous DNA (as low as 0.01% in one femur fragment). This highlights the enormous variability in DNA preservation between skeletal samples separated by a few centimeters and under very similar environmental conditions.

In section 4.3, DNA was very poorly preserved in all samples except CB13 female from Cova Bonica. To recover her genome at low coverage, shotgun sequencing alone was not a feasible strategy and we applied a combination of shotgun and whole-genome capture following the WISC approach (section 2.1.3). Post-capture libraries showed a significant increase in the percentage of human DNA, which allowed us to access their full complexity with less amount of

sequencing. Despite this clear benefit, we observed that the total amount of human non-redundant sequences decreased in post-capture libraries with respect to pre-capture libraries, due to the loss of human endogenous sequences during the capture process. A plausible explanation relies on the length of the probes, too long to efficiently bind short endogenous fragments (Ávila-Arcos et al. 2015). Optimizations in bait length distribution and other critical aspects will help to improve the enrichment efficiency and extend the scope of this methodology.

In section 4.4, we recovered genome-wide data from Late Neolithic individuals using the 1240k capture (section 2.1.3). By targeting only sequences that contained useful information, we were able to analyze a large number of samples, some of them with low endogenous DNA content. To maximize the number of polymorphic position covered while minimizing the amount of sequencing required, libraries were sequenced to the point where the number of polymorphic position covered by at least one read did not significantly increased with more sequencing. Consequently, most SNPs positions were finally covered by a few reads. While this is enough for performing detailed populations analyses, some questions will require high coverage and diploid genotype calling, something that has not yet been fully explored with this technique.

5.2. Mesolithic-Neolithic transition

Mesolithic genetic substratum

In order to shed light on the process of Neolithisation, understanding the pre-Neolithic genetic substratum of Europe is of utmost importance. When this thesis was conceived, only three European hunter-gatherers had yielded genome-wide data (Skoglund et al. 2012). These Scandinavian individuals lived many centuries after the initial arrival of farmers to the area, and only very partial genomic data (less than 5% of the genome) was recovered. With partial genomic data from La Braña 1 and La Braña 2 Iberian hunter-gatherers (section 4.1), but specially with the complete genome of La Braña 1 (section 4.2), who became the first pre-Neolithic European with a complete genome sequence, we were in the position to study the genetic make-up of Mesolithic Europe with unprecedented resolution.

When we analyzed the genetic affinities of La Braña individuals, they fell outside present-day European genetic diversity but close to Northern Europeans, suggesting that these populations have retained more Mesolithic ancestry than other modern groups. Interestingly, despite being separated by more than 2,000 kilometers, La Braña 1 showed affinity to the three Scandinavian hunter-gatherers, which could be explained by the presence of a common genetic signature among European foragers. Such signature matched the well-known uniformity in maternal lineages among Mesolithic Europeans (Brandt et al. 2015), but for the first time we were able to prove it using hundreds of thousands of genetic markers that truly represent

the overall population history. This was later confirmed with genome-wide data from more hunter-gatherer individuals (Gamba et al. 2014; Lazaridis et al. 2014; Skoglund et al. 2014a; Haak et al. 2015; Mathieson et al. 2015). Thanks to these studies, we know that Mesolithic Europe was genetically structured into three main groups: Western hunter-gatherers (WHG), including La Braña and stretching from the Iberian Peninsula to the Carpathians, Eastern hunter-gatherers (EHG) from the western part of Russia and Scandinavian hunter-gatherers (SHG), who seem to be admixed from WHG and EHG populations.

With regard to uniparental markers, La Braña individuals belonged to mtDNA haplogroup U5b, with La Braña 1 further classified as U5b2c1. This was not unexpected, given that most of the pre-Neolithic Europeans belonged to haplogroup U. The presence of the same mtDNA in both La Braña 1 and La Braña 2, who were found in the same cave and yielded overlapping radiocarbon dates, suggests they could be closely related. However, to resolve their putative relatedness we must wait until more nuclear sequences from La Braña 2 are recovered. A more surprising finding came from the Y-chromosome analysis of La Braña 1. His Y-chromosome belonged to haplogroup C-V20, which had only been found in a few present-day Southern Europeans (Scozzari et al. 2012), but at the time not in any other ancient sample. Later studies found C-V20 in Upper Paleolithic males (Fu et al. 2016), which indicates that it was relatively common in the pre-Neolithic paternal gene-pool, otherwise dominated by Y-chromosome haplogroup I (Lazaridis et al. 2014; Haak et al. 2015; Jones et al. 2015; Fu et al. 2016).

Furthermore, the La Braña 1 genome allowed us to estimate for the first time the degree of heterozygosity of Mesolithic Europeans. When we compared our genome with present-day humans from different continents, La Braña 1 genome clearly presented lower heterozygosity than a panel of different European, African and Asian populations, corresponding to a reduction in effective population size of approximately 20% relative to extant Europeans. A plausible explanation for such a restricted population size, also observed in additional hunter-gatherers (Gamba et al. 2014; Lazaridis et al. 2014; Skoglund et al. 2014a), is the extreme conditions during the Last Glacial Maximum that likely involved local extinctions and bottlenecks. These Mesolithic foragers, however, seem to have developed strategies to avoid consanguineous matings, as they did not harbor long runs of homozygosity suggestive of recent inbreeding.

People, ideas or both?

In order to test different hypotheses regarding the spread of farming into Europe, we compared La Braña 1 pre-Neolithic genome with the two Neolithic genomes available at the time, e.g Scandinavian Gok4 (Skoglund et al. 2012) and Tyrolean Iceman (Ötzi) (Keller et al. 2012). Both PCA and allele sharing (section 4.2) analyses showed a sharp difference in their genomic affinities, with Neolithic farmers clustering with Southern Europeans and La Braña 1 falling outside modern European diversity but close to Northern Europeans, as previously mentioned. Such strong genetic differentiation between individuals living before and after the arrival of farming strongly supported a model of Neolithic expansion with extensive migration.

Nevertheless, further evidence was needed to definitely solve the migration-acculturation controversy. After all, Gok4 was a Neolithic farmer from the periphery of Europe and Ötzi lived more than two millennia after the initial introduction of agriculture. When data from several Early Neolithic individuals from Central Europe were accumulated (Gamba et al. 2014; Lazaridis et al. 2014; Haak et al. 2015), it was clear that the Mesolithic-Neolithic transition in Central Europe was mediated by large-scale migrations of farmers with Near Eastern affinities. Importantly, in those studies the Mediterranean route of Neolithic expansion associated to the Cardial culture was heavily underrepresented, likely because of unfavorable conditions for aDNA preservation in the Mediterranean area. We thus decided to fill this gap and retrieve genomic information from some of the earliest farmers of the Iberian Peninsula, excavated from Early Neolithic Cardial sites along the Mediterranean and Atlantic facades (section 4.3). As expected, DNA preservation was poor, but using a combination between shotgun and capture approaches we were able to recovered five complete mtDNA genomes, which belonged to typical Neolithic lineages, and a complete low-coverage nuclear genome from a female individual (CB13) excavated at Cova Bonica (Vallirana, Barcelona).

Using this new genomic sequence, currently the only complete ancient genome from the Mediterranean region, we characterized the western end of the Neolithic expansion (section 4.3). The first important result was the strong genetic similarity between our Cardial female and other Early Neolithic individuals from Central Europe, pointing to a common origin for both routes of Neolithic expansion

into Europe. Like other early farmers, CB13's genetic background was very different from that of La Braña 1, who lived around 600 years earlier, and other pre-Neolithic hunter gatherers. In fact, if we take European individuals that lived during the Mesolithic-Neolithic transition and group them genetically, the clustering will be driven by lifestyle and not by geographic proximity. In light of these results, we can state with certainty that in most parts of Europe, it was the immigration of farmers what triggered the development of agricultural societies. This does not mean, though, that Mesolithic hunter-gatherers were completely replaced by incoming farmers. If we plot CB13 (and other Early farmers) together with hunter-gatherers and modern individuals in a PCA, her position is shifted towards the hunter-gatherer cline with respect to the modern Near Eastern cline (section 4.3), suggesting that early farmers had hunter-gatherer-related ancestry. This is confirmed with *ADMIXTURE* analysis and *f*-statistic and supported by the presence of typical hunter-gatherer uniparental markers in early farmers, e.g. mtDNA haplogroup U5 and Y-chromosome haplogroups I and C (Gamba et al. 2014; Haak et al. 2015; Mathieson et al. 2015). Interestingly, as we move to the West from Anatolia, the amount of hunter-gatherer-related ancestry in early farmers increases (Gamba et al. 2014; Haak et al. 2015; Mathieson et al. 2015). Therefore, it seems that Neolithic farmers admixed with autochthonous hunter-gatherers they encountered along their route of expansion. We even have direct evidence of the first contacts between the two groups: Hungarian KO1 sample, excavated in an Early Neolithic context but with 100% hunter-gatherer ancestry (Gamba et al. 2014). In the case of CB13,

we showed that her hunter-gatherer genetic signature was closer to foragers from Eastern Europe, e.g. Hungary, than to local Iberian foragers, e.g. La Braña 1 (section 4.3). This is not surprising given that she lived shortly after the initial arrival of farming in Iberia.

In summary, the works I have presented in sections 4.1, 4.2 and 4.3 describe a Mesolithic-Neolithic transition in Europe dominated by migration, but with significant incorporation of autochthonous hunter-gatherers through admixture. More aDNA data from currently underrepresented areas are needed to investigate regional differences and to elaborate a more complete view of the Neolithisation process.

Timing of selection events: adaptation to settled agricultural life

The shift from foraging to farming was one of the most dramatic lifestyle changes in the history of modern humans. It involved new environments and pathogens, as well as very different diets and social organizations that likely resulted in strong selective pressures and adaptation. Ancient genomes provide a direct way to study these adaptations by analyzing individuals that lived before and after a given selective event. This approach had been hampered by the lack of a pre-Neolithic genome, until we recovered the genome of La Braña 1 Mesolithic hunter-gatherer (section 4.2). We did not intend to use his genomic sequence to discover new signals of selection. Instead, we determined the allelic state of La Braña 1 at loci with known signals of recent selection in order to investigate whether they could be related to the Mesolithic-Neolithic transition. Although most of our observations have been confirmed with additional ancient

individuals, our approach had one evident caveat; one low-coverage individual might not be a good representative of an entire population.

The first obvious locus to check was the SNP (rs4988235) responsible for lactase persistence in Europe, one of the strongest signals of selection detected in present-day Europeans (Bersaglieri et al. 2004). La Braña 1 carried the ancestral allele and therefore he was likely lactose intolerant like other hunter-gatherers (Malmström et al. 2010), as expected by the absence of raw milk from domesticated animals in Mesolithic times. Interestingly, CB13 Neolithic genome (section 4.3) and all the Early and Middle Neolithic individuals sequenced to date (Burger et al. 2007; Gamba et al. 2014; Mathieson et al. 2015) also lacked the derived allele responsible for lactase persistence, contradicting an Early Neolithic date of the selective sweep at this SNP (Itan et al. 2009). The derived allele did not reach appreciable frequencies until the Bronze Age (Gamba et al. 2014; Allentoft et al. 2015; Mathieson et al. 2015), with the earliest observation in a German Bell Beaker sample dated to 2,450-2,140 cal BCE, 3,000 years after the initial arrival of agriculturalists. Thus, in just 4,000 years the allele went from very low frequency to almost fixation in many European populations. A more dense temporal and geographic sampling will surely help to explain these findings, but a possible explanation is that lactase persistence was not strongly selected until the spread of pastoralism practices during the Late Neolithic/Bronze Age.

Another suggested adaptation to the Neolithic diet is a high number of copies of the salivary amylase (*AMY1*) gene that facilitated high

starch consumption (Perry et al. 2007). We estimated a low number of copies (5) for La Braña 1 (section 4.2), compatible with a low-starch diet. Nevertheless, a later study estimated a high number copies (13) in a Mesolithic hunter-gatherer from Luxembourg (Lazaridis et al. 2014), which questions a Neolithic origin for the *AMY1* gene copy number increase.

We also examined La Braña 1 alleles at 10 non-synonymous variants with signals of recent positive selection in modern Europeans from the 1000 Genomes Project (Grossman et al. 2013). Surprisingly, we found that La Braña 1 harbored ancestral alleles at SNPs rs16891982 and rs1426654, located in *SLC45A2* and *SLC24A5* genes, respectively. The derived alleles at those positions are the major determinants of light skin pigmentation in Europeans (Lamason et al. 2005; Norton et al. 2007) and are almost fixed (97% *SLC45A2*) or fixed (100% *SLC24A5*) in present-day Europeans. Selection at those loci has been estimated to have begun between 11-19 kya, based on modern patterns of genetic variation (Soejima and Koda 2007; Beleza et al. 2013). The relatively dark pigmentation of La Braña 1 provided, to our knowledge, the first direct evidence that the adaptive spread of light skin pigmentation either was not complete by the Mesolithic, or occurred afterwards. Later aDNA studies corroborated our observations and demonstrated different timings for selective sweeps at *SLC45A2* and *SLC24A5* (Gamba et al. 2014; Lazaridis et al. 2014; Wilde et al. 2014; Allentoft et al. 2015; Mathieson et al. 2015). In the case of *SLC24A5*, the derived allele appears at very high frequencies in Early Neolithic Europe, but such a rapid increase was likely driven by migration from Anatolia (where the derived allele was fixed)

during the Neolithic expansion, and not by a selective sweep in Europe (Mathieson et al. 2015). In contrast, the derived allele of *SLC45A2* began to appear over the course of the Neolithic but always at lower frequencies than present-day Europeans, which suggests a strong selective sweep in Europe favoring lighter pigmentation during the last 5,000 years. An interesting question is what selective pressure was responsible for this recent sweep. Traditionally, it has been suggested that selection favored lighter pigmentation to maximize the production of vitamin D in regions with lower ultraviolet radiation (Jablonski and Chaplin 2000). This explanation does not seem completely satisfactory for such a recent selective event because no significant change in ultraviolet radiation occurred over the last millennia. A complementary hypothesis is that the shift to a Neolithic diet poor in vitamin D may have generated additional selective pressures favoring depigmentation in order to increase the skin vitamin D synthesis. To finish with the phenotypic characterization of La Braña 1, we determined that he very likely had blue eyes due to the presence of derived alleles for the rs12913832 SNP at *HERC2/OCA2* locus (Sturm et al. 2008), a trait shared with all the available Mesolithic hunter-gatherers (Mathieson et al. 2015; Fu et al. 2016). The frequency of the derived allele dropped during the Neolithic and Bronze Age and rose to high frequencies in extant Northern European, suggesting a complex pattern of selection.

Another possible target of selection during the transition to farming is the immune response, essential for fighting against new pathogens that came into contact with humans during that period. When we interrogated La Braña 1 genome at immunity-related loci with signals

of recent adaptation, we found the derived, putatively selected allele in several of them, including some Toll-like receptors genes involved in pathogen recognition (Netea et al. 2012). This could suggest that the Neolithic transition did not drive selection at those loci, but the interpretations here are not straightforward. Allele frequencies at those loci are not as extreme as those at pigmentation genes; the ancestral and derived alleles appear at moderate frequencies in present-day populations. More importantly, the mechanism of action, the exact infectious agents to which these genes respond and the interaction between different genes are not completely understood. With a better understanding of the genetic basis of the immune response, and with a more complete paleogenomic record, we will be able to better explain how humans adapted to these critical challenges.

5.3. Bell Beaker phenomenon

Similar to the Neolithic expansion, the Late Neolithic Bell Beaker phenomenon has also been a topic of intense debate among scholars. Our goal in section 4.4 was to contrast different hypotheses with the aid of genome-wide data from individuals excavated at Bell Beaker sites.

Discriminating between cultural and demic diffusion was, however, more challenging than in the case of the Mesolithic-Neolithic transition, due to the particular genetic structure of Late Neolithic/Early Bronze Age Europe. In fact, if we think about the genetic landscape at the onset of farming, the autochthonous European hunter-gatherers were genetically very different to the putative Near Eastern immigrants and thus, if migration was in play, we should clearly detect (as we do in sections 4.1, 4.2 and 4.3) an ancestry shift in Europe. Conversely, over the course of the Neolithic period, farmers from different areas in Europe formed a tight genetic cluster with the same two ancestry components: a Near Eastern-like component and a hunter-gatherer-like component. Therefore, assuming a Late Neolithic origin in Iberia (the most widely accepted hypothesis based on radiocarbon dates) for the Bell Beaker complex, a putative expansion from Iberia to Central Europe mediated by extensive migration will be difficult to track genetically, as Iberian Late Neolithic people did not harbor any characteristic genetic component that was not present in Central Europe. This does not mean that Iberian and Central European Neolithic farmers lacked any genetic differentiation, but that such differentiation was quite subtle

and difficult to detect using common methods such as PCA, *ADMIXTURE* or *f*-statistics that rely on unlinked genetic variants. For instance, a possible source of subtle differentiation could be different affinity to La Braña 1 Iberian hunter-gatherer, likely higher in Late Neolithic Iberia than in Central Europe. To identify fine-scale genetic structure, haplotype-based information obtained from phased diploid calls would be required (Leslie et al. 2015).

Another complication is that around the time of appearance of Bell Beakers ~2,800 BCE, the steppe migration had already arrived to Central Europe (Haak et al. 2015), and we know that it continued further to the West and South ultimately reaching most parts of Europe. Hence, during the third millennium BCE, we had the Bell Beaker expansion (either cultural or demic) going from West to East and a migration of people with steppe-related ancestry going from East to West, complicating the interpretation of genetic patterns observed in individuals dated to this period.

With these difficulties in mind, we attempted to investigate the history of this relevant cultural complex with the resolution provided by the genome-wide data we collected (section 4.4), which was otherwise much higher than any previous genetic study addressing the Bell Beaker complex.

The first interesting aspect of Bell Beakers that we analyzed was the proposed origin in Iberia (Bailly and Salanova 1999; Müller and van Willigen 2001). We found that most of our Iberian Bell Beakers were genetically very similar to preceding and contemporaneous Iberian

populations, e.g. Middle Neolithic individuals and Chalcolithic individuals without the Bell Beaker package. This is what we would expect to find if the Bell Beaker complex was originally developed by Iberian populations in the Late Neolithic. Still, other hypotheses are compatible with our data. For instance, an origin elsewhere and cultural transmission to Iberia would also produce the same pattern, but would conflict with the available radiocarbon evidence.

Next, we looked for genetic patterns that could inform us about the underlying mechanism of the Bell Beaker expansion. Our data showed a conspicuous genetic heterogeneity across the entire Bell Beaker range, caused by different amounts of steppe-related ancestry: more in Central and Northern Europe and almost absent in Iberia. The fact that we detected very different genomic affinities in individuals buried with Bell Beaker features suggests that the expansion was, at least in part, mediated by cultural diffusion. Interestingly, the genetic heterogeneity was also evident at regional scale and even within the same site, suggestive of recent or ongoing admixture between autochthonous Neolithic farmers and steppe immigrants. Related to this point, most Bell Beaker samples from Central Europe presented less steppe ancestry and more Neolithic farmer ancestry than Corded Ware people, who slightly preceded Bell Beakers in those areas and spread via massive migration from the steppe (Haak et al. 2015). This extra amount of Neolithic farmer ancestry could have an origin in Central Europe, where it was predominant before the arrival of steppe immigrants. Alternatively, it could have been introduced by Iberian Bell Beakers, who also carried this type of ancestry, through migration out of Iberia and

admixture with the Corded Ware. As mentioned before, distinguishing between the two hypotheses is quite difficult because Iberian and Central European farmers were genetically very similar (see, for instance, figures 2a and 2b in section 4.4).

A previous genetic study (Brotherton et al. 2013) analyzed the mtDNA of seven Bell Beaker individuals from Germany bearing haplogroup H and compared these data with the present-day mtDNA haplogroup H diversity. They proposed a strong genetic influx from Iberia into Central Europe during the Bell Beaker expansion, based on high frequencies of sub-lineage H1 and H3 for both present-day Iberians and the seven German Bell Beakers. However, caution should be taken when inferring complex population processes with data from only one mtDNA lineage and small sample sizes, as they might not reliably represent the real frequencies of the entire population. Moreover, the authors assumed that the mtDNA pool of Iberian populations has not significantly changed over the last 5,000 years, which is dubious, at best. In our dataset, for example, haplogroup H3 does not seem a good predictor of a putative migration of Bell Beakers out of Iberia, given that we found frequencies of 0% and 2% for Bell Beakers in Iberia (n=16) and outside Iberia (n=44), respectively.

Steppe-related ancestry is widespread in present-day Europe (Haak et al. 2015). While its arrival to Central Europe is well documented, how and when it spread to other areas is not currently understood. Comparing the genomic affinities of individuals newly reported in section 4.4 with data from previous works, we observed that the Bell

Beaker complex was responsible for the earliest appearance of steppe-related ancestry in three different regions: Britain, Northern Iberia and Northern Italy. With regard to the British Isles, we already knew that steppe ancestry was present in Ireland by the Early Bronze Age (Cassidy et al. 2016), and our data demonstrates that it had reached Britain by the earlier Bell Beaker period. Furthermore, the strong genetic discontinuity between Bell Beaker and preceding Middle Neolithic samples from England suggests a significant population turnover associated with the arrival of the Bell Beaker complex to Britain. In the case of Northern Iberia, we detected a signal of steppe ancestry in two Bell Beaker individuals (specially in one of them) from El Arroyal site. The interpretation here is not straightforward. One possibility is that these individuals were part of small-scale migrations into the Iberian Peninsula that helped to develop some of the features of the Bell Beaker complex (Case 2004), but without significant genetic impact in the general Iberian population. Alternatively, they could be pioneers of a population influx into Iberia that resulted in the presence of steppe ancestry a few centuries later during the Early Bronze Age (Günther et al. 2015).

Another observation that supports a role for the Bell Beaker phenomenon in the spread of steppe ancestry outside Central Europe is the Y-chromosome haplogroup composition. Our Bell Beaker individuals outside Iberia presented a striking uniformity dominated by R1b lineages (86%, n = 22), whose frequency sharply increased when the steppe migration arrived in Central Europe (Haak et al. 2015). Importantly, all the R1b Y-chromosomes with enough resolution belonged to R-S116 subtype that is dominant today in

Western Europe, and hence its current distribution could be related to the Bell Beaker complex. It is remarkable how the previous Neolithic lineages were almost completely replaced by a few R1b subtypes, likely reflecting new ways of social organization brought by herders of steppe origin, with some male lineages having higher reproductive success over many generations and rapidly displacing other lineages. This has been observed in other pastoral societies, such as Altai-speaking nomadic populations (Balaresque et al. 2015).

A possible confounding factor in our dataset stems from the fact that the individuals we sampled for aDNA analysis were excavated from graves with typical Bell Beaker-associated artefacts like bell-shaped beakers, tanged daggers or arrowheads. Those prestige goods were reserved for people with high social status and, as a consequence, our dataset might not be a good representation of the general population.

In my opinion, future efforts on the genetic characterization of the Bell Beaker phenomenon should focus on three aspects: gathering data from poorly represented areas in our study, recovering high-coverage genomes for fine-structure analyses and obtaining insights into temporal dynamics through the study of individuals dated to different phases within the Bell Beaker period at a given region.

5.4. Concluding remarks and future directions

The objective of this thesis was to generate genome-wide data from ancient Europeans to study recent demographic and evolutionary processes. Using the latest advances in SGS techniques applied to paleogenomic research, I have presented new aDNA sequences from 55 humans who lived between 6,000 and 2,000 cal BCE. This dataset includes the first complete pre-Neolithic European genome and the first complete Early Farmer genome from the Mediterranean, both of special importance given the unfavorable conditions for DNA preservation in the Iberian Peninsula. In light of the most recent aDNA works, recovering genome-wide information from 55 ancient individuals might seem like an easy task, but we must bear in mind that five years ago not a single ancient individual (modern human) from Europe had available genomic data.

These achievements are certainly of great value, but the true relevance of this thesis relies on its crucial contribution to our understanding of key demographic and evolutionary events. We were able to clarify long-standing archeological debates regarding the mechanisms underlying both the Mesolithic-Neolithic transition and the Bell Beaker phenomenon. We learned that the real dynamics of those processes did not adjust, in most cases, to one of the two polarized models of cultural and demic diffusion. Instead, our data supported intermediate models, with different contributions of migration and acculturation depending on the process and the area involved. Furthermore, we gained novel insights into the timing of recent selection events, such as dietary and pigmentation adaptations,

that helped Europeans to face challenges imposed by new lifestyles and environments.

Additionally, DNA sequences produced in this work are available for other researchers, and will surely contribute to future investigations addressing population genetics, selection and other topics. In fact, our data have already been used in many studies, most of them published in high-profile journals.

Thanks to the paleogenomic evidence accumulated in the past few years, our knowledge of human prehistory is more complete than ever before. The good news is that many interesting questions are still waiting to be addressed with aDNA. Although the major recent population events in Europe may have already been deciphered, they did not equally impact different regions across such a vast territory. Sampling currently underrepresented areas will hence provide us with a more accurate vision of such processes. If we move out of Europe, the picture becomes even more promising, since aDNA studies have barely focused on areas such as Africa, Asia or the Americas, where many important demographic and selective events remain obscure.

Finally, improvements in all the steps of aDNA research, including data management and interpretation, will maximize the number of ancient specimens suitable for paleogenomic study and provide high quality genomes for fine-scale genomic analyses.

For those who, like me, greatly enjoy the study of the past, future looks bright and exciting.

Contributions to other publications

Charlier P, **Olalde I**, Solé N, Ramírez O, Babelon JP, Galland B, Calafell F, Lalueza-Fox C. 2013. Genetic comparison of the head of Henri IV and the presumptive blood from Louis XVI (both Kings of France). *Forensic science international* 226:38–40.

Ramírez O, Gómez-Díaz E, **Olalde I**, Illera JC, Rando JC, González-Solís J, Lalueza-Fox C. 2013. Population connectivity buffers genetic diversity loss in a seabird. *Frontiers in zoology* 10:28.

Ramirez O, **Olalde I**, Berglund J, Lorente-Galdos B, Hernandez-Rodriguez J, Quilez J, Webster MT, Wayne RK, Lalueza-Fox C, Vilà C, Marques-Bonet T. 2014. Analysis of structural diversity in wolf-like canids reveals post-domestication variants. *BMC genomics* 15:465.

Gómez-Sánchez D, **Olalde I**, Pierini F, Matas-Lalueza L, Gigli E, Lari M, Civit S, Lozano M, Vergès JM, Caramelli D, Ramírez O, Lalueza-Fox C. 2014. Mitochondrial DNA from El Mirador Cave (Atapuerca, Spain) Reveals the Heterogeneity of Chalcolithic Populations. *PLoS ONE* 9:e105105.

Olalde I, Sánchez-Quinto F, Datta D, Marigorta UM, Chiang CWK, Rodríguez JA, Fernández-Callejo M, González I, Montfort M, Matas-Lalueza L, Civit S, Luiselli D, Charlier P, Pettener D, Ramírez O, Navarro A, Himmelbauer H, Marquès-Bonet T, Lalueza-Fox C. 2014. Genomic analysis of the blood attributed to Louis XVI (1754-1793), king of France. *Scientific reports* 4:4666.

Olalde I, Capote J, Del-Arco MC, Atoche P, Delgado T, González-Anton R, Pais J, Amills M, Lalueza-Fox C, Ramírez O. 2015. Ancient DNA sheds light on the ancestry of pre-hispanic Canarian pigs. *Genetics Selection Evolution* 47:1–5.

Olalde I, Lalueza-Fox C. 2015. Modern humans' paleogenomics and the new evidences on the European prehistory. *Science and Technology of Archaeological Research* 1:STAR20151120548.

Ramírez O, Burgos-Paz W, Casas E, Ballester M, Bianco E, **Olalde I**, Santpere G, Novella V, Gut M, Lalueza-Fox C, Saña M, Pérez-Enciso M. 2015. Genome data from a sixteenth century pig illuminate modern breed relationships. *Heredity* 114:175–84.

Bibliography

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* 75:910–918.

Adler CJ, Haak W, Donlon D, Cooper A. 2011. Survival and recovery of DNA from ancient teeth and bones. *J. Archaeol. Sci.* 38:956–964.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522:167–172.

Ammerman AJ, Cavalli-Sforza LL. 1984. The Neolithic transition and the genetics of populations in Europe. Princeton: Princeton University Press

Andersen ST. 1995. Pollen analytical investigations of barrows from the Funnel Beaker and Single Grave Cultures in the Vroue area, West Jutland, Denmark. *J. Danish Archaeol.* 12:107–131.

Anthony DW, Ringe D. 2015. The Indo-European Homeland from Linguistic and Archaeological Perspectives. *Annu. Rev. Linguist.* 1:199–219.

Anthony DW. 2007. The Horse, the Wheel, and Language. Princeton and Oxford: Princeton University Press

Antia R, Regoes RR, Koella JC, Bergstrom CT. 2003. The role of evolution in the emergence of infectious diseases. 426:8–11.

Arias P. 2007. Neighbours but diverse: social change in north-west Iberia during the transition from the Mesolithic to the Neolithic

(5500–4000 cal BC). In: Whittle A, Cummings V, editors. Going over: The Mesolithic-Neolithic Transition in North-West Europe. Oxford University Press.

Aufderheide AC, Rodriguez-Martin C, Langsjoen OM. 1998. The Cambridge encyclopedia of human paleopathology. Cambridge: Cambridge University Press

Austin JJ, Ross AJ, SMITH AB, Fortey RA, Thomas RH. 1997. Problems of reproducibility – does geologically ancient DNA survive in amber–preserved insects? *Proc. R. Soc. London B Biol. Sci.* 264:467–474.

Avery OT, Macleod CM, McCarty M. 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation By a Desoxyribonucleic Acid Fraction Isolated From Pneumococcus Type Iii. *J. Exp. Med.* 79:137–158.

Ávila-Arcos MC, Sandoval-Velasco M, Schroeder H, Carpenter ML, Malaspinas A-S, et al. 2015. Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. *Methods Ecol. Evol.* 6:725–734.

Bailey JF, Richards MB, Macaulay VA, Colson IB, James IT, et al. 1996. Ancient DNA Suggests a Recent Expansion of European Cattle from a Diverse Wild Progenitor Species. *Proc. R. Soc. London B Biol. Sci.* 263:1467–1473.

Bailly M, Salanova L. 1999. Les dates radiocarbone du Campaniforme en Europe Occidentale : Analyse critique des principales séries de dates. *Mémoires la Société préhistorique française* 26:219–224.

Balaresque P, Bowden GR, Adams SM, Leung HY, King TE, et al. 2010. A predominantly neolithic origin for European paternal lineages. *PLoS Biol.* 8.

Balaresque P, Poulet N, Cussat-Blanc S, Gerard P, Quintana-Murci L, et al. 2015. Y-chromosome descent clusters and male

differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations. *Eur. J. Hum. Genet.*:1–10.

Bar-Yosef O. 2007. The Upper Paleolithic revolution. *Annu. Rev. Anthr.* 31:363–393.

Barbujani G, Di Benedetto G. 2001. Genetic variances within and between human groups. In: Donnelly P, Foley RA, editors. Genes, Fossils, and Behaviour: An Integrated Approach to Human Evolution. IOS Press. p. 63–77.

Barker G. 1985. Prehistoric Farming in Europe. Cambridge: Cambridge University Press

Barnett R, Larson G. 2012. A phenol–chloroform protocol for extract- ing DNA from ancient samples. In: Shapiro B, Hofreiter M, editors. Ancient DNA: Methods and Protocols. New York: Humana Press, Springer. p. 13–19.

Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB. 2004. Assessing the causes of late Pleistocene extinctions on the continents. *Science* 306:70–75.

Beleza S, Santos AM, McEvoy B, Alves I, Martinho C, et al. 2013. The timing of pigmentation lightening in Europeans. *Mol. Biol. Evol.* 30:24–35.

Belle EMS, Landry P-A, Barbujani G. 2006. Origins and evolution of the Europeans' genome: evidence from multiple microsatellite loci. *Proc. R. Soc. B-Biological Sci.* 273:1595–1602.

Benazzi S, Douka K, Fornai C, Bauer CC, Kullmer O, et al. 2011. Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* 479:525–528.

Bennett EA, Massilani D, Lizzo G, Daligault J, Geigl E-M, et al. 2014. Library construction for ancient genomics: single strand or double strand? *Biotechniques* 56:289–300.

Bentley D, Balasubramanian S, Swerdlow H, Smith G, Milton J, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, et al. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* 74:1111–1120.

Bogucki PI, Crabtree PJ. 2004. Ancient Europe, 8000 B.C. to A.D. 1000: An Encyclopedia of the Barbarian World. Charles Scribner's Sons

Bollongino R, Nehlich O, Richards MP, Orschiedt J, Thomas MG, et al. 2013. 2000 Years of Parallel Societies in Stone Age Central Europe. *Science* 342:479–481.

Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, et al. 2011. A draft genome of Yersinia pestis from victims of the Black Death. *Nature* 478:506–510.

Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, et al. 2009. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326:137–140.

Bramanti B. 2008. Ancient DNA: genetic analysis of aDNA from sixteen skeletons of the Vedrovice. *Anthropologie* 46:153–160.

Brandt G, Haak W, Adler CJ, Roth C, Szécsényi-Nagy A, et al. 2013. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* 342:257–261.

Brandt G, Szécsényi-Nagy A, Roth C, Alt KW, Haak W. 2015. Human paleogenetics of Europe - The known knowns and the known unknowns. *J. Hum. Evol.* 79:73–92.

Briggs AW, Good JM, Green RE, Krause J, Maricic T, et al. 2009. Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science* 325:318–321.

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, et al. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* 104:14616–14621.

Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, et al. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* 38:1–12.

Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, et al. 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res.* 35:5717–5728.

Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, et al. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* 4:1764.

Buermans HPJ, den Dunnen JT. 2014. Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta* 1842:1932–1941.

Burger J, Kirchner M, Bramanti B, Haak W, Thomas MG. 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl. Acad. Sci. U. S. A.* 104:3736–3741.

Busing FMTA, Meijer E, Van Der Leeden R. 1999. Delete- m Jackknife for Unequal m. *Stat. Comput.* 9:3–8.

Caramelli D, Lalueza-Fox C, Condemi S, Longo L, Milani L, et al. 2006. A highly divergent mtDNA sequence in a Neandertal individual from Italy. *Curr. Biol.* 16:630–632.

Cardoso JL. 2014. Absolute chronology of the Beaker phenomenon North of the Tagus estuary : demographic and social implications. *Trabajos de Prehistoria* 71:56–75.

Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, et al. 2013. Pulling out the 1%: Whole-Genome Capture

for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am. J. Hum. Genet.* 93:852–864.

Case H. 2004. Beakers and the Beaker culture. In: Czebreszuk J, editor. Similar but Different. Bell Beakers in Europe. p. 11–34.

Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, et al. 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc. Natl. Acad. Sci.* 113:1–6.

Castellano S, Parra G, Sánchez-Quinto F a, Racimo F, Kuhlwilm M, et al. 2014. Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci. U. S. A.* 111:6666–6671.

Cavalli-Sforza LL, Edwards a W. 1967. Phylogenetic analysis. Models and estimation procedures. *Am. J. Hum. Genet.* 19:233–257.

Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G. 1998. Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. *Proc. Natl. Acad. Sci. U. S. A.* 95:9053–9058.

Childe VG. 1925. The dawn of European civilization. London: Kegan Paul

Childe VG. 1950. The urban revolution. Liverpool: Liverpool University Press

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, et al. 2009. The Last Glacial Maximum. *Science* 325:710–714.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, et al. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* 409:704–707.

Cooper A, Poinar HN. 2000. Ancient DNA: Do It Right or Not at All. *Science* 289:1139.

Craig DW, Pearson J V, Szelinger S, Sekar A, Margot R, et al. 2008. Identification of Genetic Variants Using Barcoded Multiplexed Sequencing. *Nat. Methods* 5:887–893.

Czebreszuk J. 2004a. Bell Beakers from West to East.pdf. In: Bogucki PI, Crabtree PJ, editors. Ancient Europe, 8000 B.C. to A.D. 1000: An Encyclopedia of the Barbarian World. Charles Scribner's Sons. p. 476–485.

Czebreszuk J. 2004b. Similar But Different. Bell Beakers in Europe. Leiden: Sidestone Press

Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, et al. 2013a. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* 110:15758–15763.

Dabney J, Meyer M, Paabo S. 2013b. Ancient DNA Damage. *Cold Spring Harb. Perspect. Biol.* 5:a012567–a012567.

Dabney J, Meyer M. 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* 52:87–94.

Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, et al. 2015. Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* 5:11184.

de-la-Rua C, Izagirre N, Alonso S, Hervella M. 2015. Ancient DNA in the Cantabrian fringe populations: A mtDNA study from Prehistory to Late Antiquity. *Quat. Int.* 364:306–311.

Deguilloux M-F, Soler L, Pemonge M-H, Scarre C, Joussaume R, et al. 2011. News from the west: ancient DNA from a French megalithic burial chamber. *Am. J. Phys. Anthropol.* 144:108–118.

Diamond J. 1997. Guns, Germs, and Steel: The Fates of Human Societies. New York: W.W. Norton & Co

Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends Genet.* 30.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28:2239–2252.

Edwards AWF. 2003. Human genetic diversity: Lewontin's fallacy. *BioEssays* 25:798–801.

Engelhardt BE, Stephens M. 2010. Analysis of population structure: A unifying framework and novel methods based on sparse factor analysis. *PLoS Genet.* 6.

Enk JM, Devault AM, Kuch M, Murgha YE, Rouillard JM, et al. 2014. Ancient whole genome enrichment using baits built from modern dna. *Mol. Biol. Evol.* 31:1292–1294.

Falgueras J, Lara AJ, Fernández-Pozo N, Cantón FR, Pérez-Trabado G, et al. 2010. SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* 11:38.

Fortea J, de la Rasilla M, García-Tabernero A, Gigli E, Rosas A, et al. 2008. Excavation protocol of bone remains for Neandertal DNA analysis in El Sidrón Cave (Asturias, Spain). *J. Hum. Evol.* 55:353–357.

Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, et al. 2015. An early modern human from Romania with a recent Neanderthal ancestor. *Nature*.

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514:445–449.

Fu Q, Meyer M, Gao X, Stenzel U, Burbano H a, et al. 2013a. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U. S. A.* 110:2223–2227.

Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, et al. 2013b. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23:553–559.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, et al. 2016. The genetic history of Ice Age Europe. *Nature*.

Gallay A. 2001. L'énigme campaniforme. In: Nicolis F, editor. Bell Beakers Today. Pottery, people, culture, symbols in prehistoric Europe. Trento. p. 41–57.

Gamba C, Fernández E, Tirado M, Deguilloux MF, Pemonge MH, et al. 2012. Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Mol. Ecol.* 21:45–56.

Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, et al. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* 5:5257.

Gamble C, Davies W, Pettitt P, Richards M. 2004. Climate change and evolving human diversity in Europe during the last glacial. *Philos. Trans. R. Soc. London B Biol. Sci.* 359:243–254.

Gansauge M-T, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8:737–748.

Gansauge M-T, Meyer M. 2014. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.*

Gerbault P, Liebert A, Itan Y, Powell A, Currat M, et al. 2011. Evolution of lactase persistence: an example of human niche construction. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366:863–877.

Gilbert MTP, Binladen J, Miller W, Wiuf C, Willerslev E, et al. 2007. Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res.* 35:1–

10.

Gilbert MTP, Rudbeck L, Willerslev E, Hansen AJ, Smith C, et al. 2005. Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *J. Archaeol. Sci.* 32:785–793.

Gokhman D, Lavi E, Prüfer K, Fraga MF, Riancho JA, et al. 2014. Reconstructing the DNA Methylation Maps of the Neandertal and the Denisovan. *Science* 344:523–528.

Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, et al. 1990. Chloroplast DNA sequence from a Miocene Magnolia species. *Nature* 344:656–658.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:710–722.

Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, et al. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* 444:330–336.

Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, et al. 2013. Identifying recent adaptations in large-scale genomic data. *Cell* 152:703–713.

Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-Varela R, et al. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl. Acad. Sci.*:201509851.

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, et al. 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol.* 8:e1000536.

Haak W, Forster P, Bramanti B, Villems R, Renfrew C, et al. 2005. Ancient DNA from the First European Farmers in 7500-Year-Old Neolithic Sites. 310:1016–1019.

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522:207–211.

Hagelberg E, Sykes B, Hedges R. 1989. Ancient bone DNA amplified. *Nature* 342:485.

Härke H. 1998. Archaeologists and Migrations: A Problem of Attitude? *Curr. Anthropol.* 39:19–46.

Harrison RJ, Martin AM. 2001. Beakers and Social Complexity in Central Spain. In: Nicolis F, editor. Bell beakers today: pottery, people, culture, symbols in prehistoric Europe. proceedings of the International colloquium, Riva del Garda (Trento, Italy). Trento (Italy). p. 111–124.

Hebsgaard MB, Phillips MJ, Willerslev E. 2005. Geologically ancient DNA: Fact or artefact? *Trends Microbiol.* 13:212–220.

Heintzman PD, Chang D, Shapiro B, Biology E. 2015. Paleogenomics. *Rev. Cell Biol. Mol. Med.* 1:243–267.

Hermanussen M. 2003. Stature of early Europeans. *Hormones* 2:175–178.

Hervella M, Izagirre N, Alonso S, Fregel R, Alonso A, et al. 2012. Ancient DNA from hunter-gatherer and farmer groups from Northern Spain supports a random dispersion model for the Neolithic expansion into Europe. *PLoS One* 7:e34417.

Higham T, Compton T, Jacobi R, Shapiro B, Trinkaus E, et al. 2011. The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* 479:31–34.

Higham T, Douka K, Wood R, Ramsey CB, Brock F, et al. 2014. The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* 512:306–309.

Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the

horse family. *Nature* 312:282–284.

Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. 2001a. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* 29:4793–4799.

Hofreiter M, Paijmans JLA, Goodchild H, Speller CF, Barlow A, et al. 2014. The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays* 36:1–10.

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. 2001b. Ancient DNA. *Nat. Rev. Genet.* 2:353–359.

Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, Paabo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* 24:1304–1307.

Hublin JJ. 2009. The origin of Neandertals. *Proc. Natl. Acad. Sci.* 106.

Itan Y, Powell A, Beaumont M a, Burger J, Thomas MG. 2009. The origins of lactase persistence in Europe. *PLoS Comput. Biol.* 5:e1000491.

Jablonski NG, Chaplin G. 2000. The evolution of human skin coloration. *J. Hum. Evol.* 39:57–106.

Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, et al. 2015. Upper palaeolithic genomes reveal deep roots of modern eurasians. *Nat. Comm.*:1–8.

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1682–1684.

Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, et al. 2012. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* 3:698.

Kerpedjiev P, Frellsen J, Lindgreen S, Krogh A. 2014. Adaptable probabilistic mapping of short reads using position specific scoring matrices. *BMC Bioinformatics* 15:100.

Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40:1–8.

Kircher M. 2012. Analysis of High-Throughput Ancient DNA Sequencing Data. In: Shapiro B, Hofreiter M, editors. Ancient DNA: Methods and Protocols.

Kohl PL. 2007. The Making of Bronze Age Eurasia. Cambridge: Cambridge University Press

Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, et al. 2010a. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* 20:231–236.

Krause J, Fu Q, Good JM, Viola B, Shunkov M V, et al. 2010b. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464:894–897.

Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, et al. 2007. The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr. Biol.* 17:1908–1912.

Kremenetski K V. 2003. Steppe and forest steppe belt of Eurasia: Holocene environmental history. In: Levine M, Renfrew C, Boyle K, editors. Prehistoric Steppe Adaptations and the Horse. Cambridge: McDonald Institute for Archaeological Research. p. 11–29.

Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, et al. 2000. A view of Neandertal genetic diversity. *Nat. Genet.* 26:144–146.

Krings M, Geisert H, Schmitz RW, Krainitzki H, Pääbo S. 1999. DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen. *Proc. Natl. Acad. Sci. U. S. A.* 96:5581–5585.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, et al. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19–30.

Kristiansen K. 2011. The Bronze Age expansion of Indo-European languages: an archaeological model. In: Prescott C, Glorstad H, editors. Becoming European: The transformation of third millennium Northern and Western Europe. p. 165–181.

Kuhlwilm M, Gronau I, Hubisz MJ, de Filippo C, Prado-Martinez J, et al. 2016. Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature*.

Lacan M, Keyser C, Ricaut F-X, Brucato N, Duranthon F, et al. 2011a. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc. Natl. Acad. Sci. U. S. A.* 108:9788–9791.

Lacan M, Keyser C, Ricaut F-X, Brucato N, Tarrús J, et al. 2011b. Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc. Natl. Acad. Sci. U. S. A.* 108:18255–18259.

Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, Rosas A, et al. 2008. Genetic characterization of the ABO blood group in Neandertals. *BMC Evol. Biol.* 8:342.

Lalueza-Fox C, Krause J, Caramelli D, Catalano G, Milani L, et al. 2006. Mitochondrial DNA of an Iberian Neandertal suggests a population affinity with other European Neandertals. *Curr. Biol.* 16:629–630.

Lalueza-Fox C, Römpler H, Caramelli D, Stäubert C, Catalano G, et al. 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318:1453–1455.

Lalueza-Fox C, Sampietro ML, Caramelli D, Puder Y, Lari M, et al. 2005. Neandertal evolutionary genetics: Mitochondrial DNA data from the Iberian Peninsula. *Mol. Biol. Evol.* 22:1077–1081.

Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, et al. 2005. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 310:1782–1786.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.

Lanting JN, Van der Waals J. 1976. Beaker culture relations in the Lower Rhine Basin. In: Glockenbecher Symposion, Oberreid 1974. Haarlem. p. 1–80.

Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, et al. 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc. Natl. Acad. Sci. U. S. A.* 104:15276–15281.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513:409–413.

Lee EJ, Makarewicz C, Renneberg R, Harder M, Krause-Kyora B, et al. 2012. Emerging genetic patterns of the european neolithic: Perspectives from a late neolithic bell beaker burial site in Germany. *Am. J. Phys. Anthropol.* 148:571–579.

Lemercier O. 2004. Historical model of setting and spreading out of the Bell Beaker culture in Mediterranean France. In: Czebreszuk J, editor. Similar But Different: Bell Beakers in Europe. p. 193–205.

Lemercier O. 2012. Interpreting the Beaker phenomenon in Mediterranean France: an Iron Age analogy. *Antiquity* 86:131–143.

Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, et al. 2015. The fine-scale genetic structure of the British population. *Nature* 519:309–314.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009a. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Li R, Yu C, Li Y, Lam TW, Yiu SM, et al. 2009b. SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* 25:1966–1967.

Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362:709–715.

Lindgreen S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* 5:337.

Lipson M, Loh PR, Levin A, Reich D, Patterson N, et al. 2013. Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* 30:1788–1802.

Liu L, Li Y, Li S, Hu N, He Y, et al. 2012. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012.

Llorente MG, Jones ER, Eriksson A, Siska V, Arthur KW, et al. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* 350:820–822.

Ludwig A, Pruvost M, Reissmann M, Benecke N, Brockmann GA, et al. 2009. Coat color variation at the beginning of horse domestication. *Science* 324:485.

Lunter G, Goodson M. 2011. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 21:936–939.

Malmer MP. 2002. The Neolithic of South Sweden: TRB, GRK, and STR. Royal Academy of Letters

Malmström H, Gilbert MTP, Thomas MG, Brandström M, Storå J, et al. 2009. Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians. *Curr. Biol.* 19:1758–1762.

Malmström H, Linderholm A, Lidén K, Storå J, Molnar P, et al. 2010. High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evol. Biol.* 10:89.

Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, et al. 2010. The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS One* 5:16–20.

Marciniak S, Klunk J, Devault A, Enk J, Poinar HN. 2015. Ancient human genomics: the methodology behind reconstructing evolutionary pathways. *J. Hum. Evol.* 2014.

Maricic T, Whitten M, Pääbo S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5:e14004.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* 17:10.

Martin MD, Cappellini E, Samaniego JA, Zepeda ML, Campos PF, et al. 2013. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nat. Commun.* 4:2172.

Martiniano R, Caffell A, Holst M, Hunter-mann K, Montgomery J, et al. 2016. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nat. Commun.*:1–8.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528:499–503.

Maxam AM, Gilbert W. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* 74:560–564.

McVean G. 2009. A genealogical interpretation of principal components analysis. *PLoS Genet.* 5.

Melchior L, Lynnerup N, Siegismund HR, Kivisild T, Dissing J. 2010. Genetic diversity among ancient Nordic populations.

PLoS One 5:22–25.

Menozzi P, Piazza L, Cavalli-Sforza L. 1978. Synthetic Maps of Human Gene Frequencies in Europeans. *Science* 201:786–792.

Metzker ML. 2010. Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11:31–46.

Meyer M, Arsuaga J-L, de Filippo C, Nagel S, Aximu-Petri A, et al. 2016. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* 531:504–507.

Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, et al. 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505:403–406.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222–226.

Meyer M, Kircher M. 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc.*:http://dx.doi.org/10.1101/prot5448.

Miller W, Drautz DI, Ratan A, Pusey B, Qi J, et al. 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:387–390.

Moss WJ, Griffin DE. 2006. Global measles elimination. *Nat. Rev. Microbiol.* 4:900–908.

Müller J, van Willigen S. 2001. New radiocarbon evidence for European Bell Beakers and the consequences for the diffusion of the Bell Beaker phenomenon. In: Nicolis F, editor. Bell beakers today: pottery, people, culture, symbols in prehistoric Europe. proceedings of the International colloquium, Riva del Garda (Trento, Italy). Trento. p. 59–80.

Mullis KB, Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction.Enzymology BT-M in,

editor. *Methods Enzymol.* Volume 155:335–350.

Netea MG, Wijmenga C, O'Neill L a J. 2012. Genetic variation in Toll-like receptors and disease susceptibility. *Nat. Immunol.* 13:535–542.

Nielsen R, Mountain JL, Huelsenbeck JP, Slatkin M. 1998. Maximum-likelihood estimation of population divergence times and population phylogenies in models without mutation. *Evolution (N. Y).* 52:669–677.

Norton HL, Kittles R a, Parra E, McKeigue P, Mao X, et al. 2007. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* 24:710–722.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, et al. 2008. Genes mirror geography within Europe. *Nature* 456:98–101.

Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CWK, et al. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* 507:225–228.

Olalde I, Schroeder H, Sandoval-Velasco M, Vinner L, Lobón I, et al. 2015. A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Mol. Biol. Evol.* 32:3132–3142.

Orlando L, Calvignac S, Schnebelen C, Douady CJ, Godfrey LR, et al. 2008. DNA from extinct giant lemurs links archaeolemurids to extant indriids. *BMC Evol. Biol.* 8:121–129.

Orlando L, Darlu P, Toussaint M, Bonjean D, Otte M, et al. 2006. Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Curr. Biol.* 16:400–402.

Orlando L, Gilbert MTP, Willerslev E. 2015. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* 16:395–408.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, et al.

2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499:74–78.

Ovchinnikov I V, Götherström A, Romanova GP, Kharitonov VM, Lidén K, et al. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404:490–493.

Paabo S. 1985. Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314:644–645.

Pääbo S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. U. S. A.* 86:1939–1943.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, et al. 2012. Ancient admixture in human history. *Genetics* 192:1065–1093.

Pedersen JS, Valen E, Vargas Velazquez AM, Parker BJ, Rasmussen M, et al. 2014. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* 24:454–466.

Perlès C. 2001. The Early Neolithic in Greece. Cambridge: Cambridge University Press

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* 39:1256–1260.

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967.

Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, et al. 2015. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS One* 10:e0129102.

Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J. 2012. The genetic history of Europeans. *Trends Genet.* 28:496–505.

Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RDE, et al. 2006. Metagenomics to Paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311:392–394.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43–49.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505:87–91.

Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, et al. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*:1–20.

Ramírez O, Gigli E, Bover P, Alcover JA, Bertranpetit J, et al. 2009. Paleogenomics in a temperate environment: Shotgun sequencing from an extinct Mediterranean caprine. *PLoS One* 4:1–6.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, et al. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506:225–229.

Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, et al. 2011. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334:94–98.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757–762.

Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, et al. 2015a. The ancestry and affiliations of Kennewick Man. *Nature* 523:455–458.

Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, et al. 2015b. Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. *Cell* 163:571–582.

Reich D, Green RE, Kircher M, Krause J, Patterson N, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–1060.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* 461:489–494.

Renfrew C. 1996. Language families and the spread of farming. In: Harris DH, editor. The origins and spread of agriculture and pastoralism in Eurasia. London: University College of London Press. p. 70–92.

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, et al. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* 59:185–203.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67:1251–1276.

Richards MP, Schulting RJ, Hedges REM. 2003. Archaeology: sharp shift in diet at onset of Neolithic. *Nature* 425:366.

Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. 2015. Partial uracil – DNA – glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. London B.*

Rohland N, Hofreiter M. 2007. Ancient DNA extraction from bones and teeth. *Nat. Protoc.* 2:1756–1762.

Rohland N. 2012. DNA Extraction of Ancient Animal Hard Tissue Samples via Adsorption to Silica Particles. In: Shapiro B, Hofreiter M, editors. Ancient DNA: Methods and Protocols. Vol. 840. New York: Humana Press, Springer. p. 21–28.

Rougier H, Milota S, Rodrigo R, Gherase M, Sarcina L, et al. 2007. Peştera cu Oase 2 and the cranial morphology of early modern Europeans. *Proc. Natl. Acad. Sci. U. S. A.* 104:1165–1170.

Sampietro ML, Gilbert MTP, Lao O, Caramelli D, Lari M, et al. 2006. Tracking down human contamination in ancient human teeth. *Mol. Biol. Evol.* 23:1801–1807.

Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, et al. 2007. Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc. R. Soc. B-Biological Sci.* 274:2161–2167.

Sánchez-Quinto F, Schroeder H, Ramirez O, Avila-Arcos MC, Pybus M, et al. 2012. Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. *Curr. Biol.* 22:1494–1499.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74:5463–5467.

Der Sarkissian C, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, et al. 2013. Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. *PLoS Genet.* 9:e1003296.

Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* 7:e34131.

Sawyer S, Renaud G, Viola B, Hublin J, Gansauge M. 2015. Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl. Acad. Sci. U. S. A.* 112:15696–15700.

Schadt EE, Turner S, Kasarskis A. 2010. A window into third-generation sequencing. *Hum. Mol. Genet.* 19:227–240.

Schiffels S, Haak W, Paajanen P, Llamas B, Popescu E, et al. 2016. Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* 7:022723.

Schloss JA. 2008. How to get genomes at one ten-thousandth the cost. *Nat. Biotechnol.* 26:1113–1115.

Schroeder H, Ávila-Arcos MC, Malaspinas A-S, Poznik GD, Sandoval-Velasco M, et al. 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proc. Natl. Acad. Sci.* 112:3669–3673.

Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid K a S, et al. 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* 13:178.

Schwarz C, Debruyne R, Kuch M, McNally E, Schwarcz H, et al. 2009. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Res.* 37:3215–3229.

Scozzari R, Massaia A, D'Atanasio E, Myres NM, Perego U a, et al. 2012. Molecular dissection of the basal clades in the human Y chromosome phylogenetic tree. *PLoS One* 7:e49170.

Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas A, Manica A, et al. 2014. Genomic structure in Europeans dating back at least 36,200 years. *Science* 346:1113–1118.

Seguin-Orlando A, Schubert M, Clary J, Stagegaard J, Alberdi MT, et al. 2013. Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes. *PLoS One* 8:e78575.

Sherratt A. 1989. Sacred and Profane Substances: The Ritual Use of Narcotics in Later Neolithic Europe. In: Garwood P, editor. Sacred and Profane: Proceedings of a Conference on Archaeology, Ritual, and Religion, Oxford, 1989. Oxford: Oxford University Committee for Archaeology. p. 50–64.

Siegert MJ, Marsiat I. 2001. Numerical reconstructions of LGM climate across the Eurasian Arctic. *Quat. Sci. Rev.* 20:1595–1605.

Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, et al. 2014a. Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science* 201:786–792.

Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, et al. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336:466–469.

Skoglund P, Northoff BH, Shunkov M V., Derevianko a. P, Paabo S, et al. 2014b. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci.*:1–6.

Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ. 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *J. Hum. Evol.* 45:203–217.

Soejima M, Koda Y. 2007. Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. *Int. J. Legal Med.* 121:36–39.

Soltis PS, Soltis DE, Smiley CJ. 1992. An rbcL sequence from a Miocene Taxodium (bald cypress). *Proc. Natl. Acad. Sci. U. S. A.* 89:449–451.

Spencer Larsen C. 1995. Biological Changes in Human Populations with Agriculture. *Annu. Rev. Anthropol.* 24:185–213.

Stoneking M, Krause J. 2011. Learning about human population history from ancient and modern genomes. *Nat. Rev. Genet.* 12:603–614.

Sturm RA, Duffy DL, Zhao ZZ, Leite FPN, Stark MS, et al. 2008. A Single SNP in an Evolutionary Conserved Region within Intron 86 of the HERC2 Gene Determines Human Blue-Brown Eye Color. *Am. J. Hum. Genet.* 82:424–431.

Szécsényi-Nagy A, Brandt G, Haak W, Keerl V, Jakucs J, et al. 2015. Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proc. R. Soc. B-Biological Sci.* 282:1–9.

The 1000 Genomes Project Consortium 2015. A global reference for human genetic variation. *Nature* 526:68–74.

The International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945.

Trinkaus E. 2005. Early Modern Humans. *Annu. Rev. Anthropol.* 34:207–230.

Vandkilde H. 2006. Warriors and warrior institutions in Copper Age Europe. In: Otto T, Thrane H, Vandkilde H, editors. Warfare and Society: Archaeological and Social Anthropological Perspectives. Aarhus: Aarhus Univ. Press. p. 393–431.

Vilstrup JT, Seguin-Orlando A, Stiller M, Ginolhac A, Raghavan M, et al. 2013. Mitochondrial Phylogenomics of Modern and Ancient Equids. *PLoS One* 8.

Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, et al. 2014. Yersinia pestis and the Plague of Justinian 541-543 AD: A genomic analysis. *Lancet Infect. Dis.* 14:319–326.

Wall JD, Kim SK. 2007. Inconsistencies in neanderthal genomic DNA sequences. *PLoS Genet.* 3:1862–1866.

Watson JD, Crick FHC. 1953. Molecular structure of nucleic acids. *Nature* 171:737–738.

Whittle A, Cummings V. 2007. Going over: the Mesolithic-Neolithic transition in North-West Europe. Oxford University Press

Whittle A. 1996. Europe in the Neolithic: the Creation of New Worlds. Cambridge: Cambridge University Press

Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, et al. 2014. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. U. S. A.* 111:4832–4837.

Wolfe ND, Dunavan CP, Diamond J. 2007. Origins of major human infectious diseases. *Nature* 447:279–283.

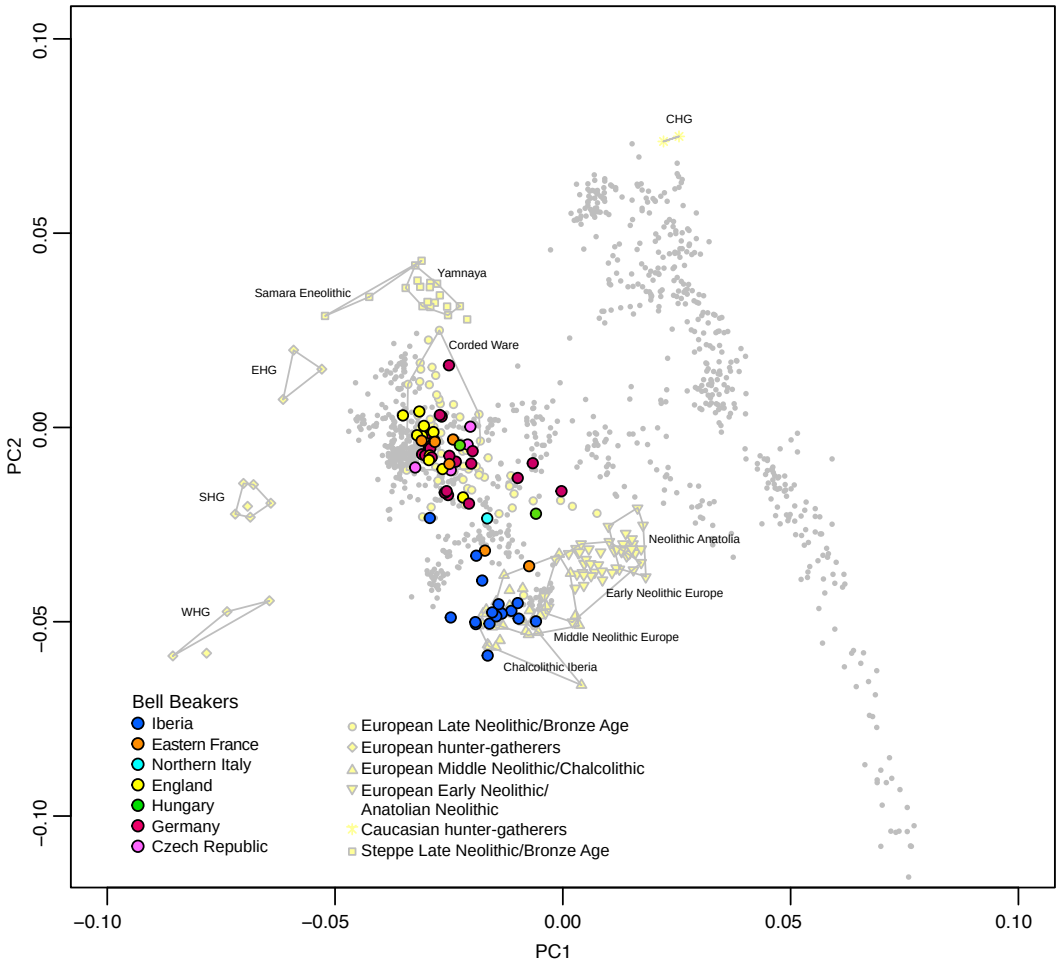
Woodward, Weyand NJ, Bunnell M. 1994. DNA sequence from Cretaceous period bone fragments. *Science* 266:1229–1232.

Zilhão J. 2000. From the Mesolithic to the Neolithic in the Iberian peninsula. In: Europe's First Farmers. Cambridge: Cambridge University Press. p. 144–182.

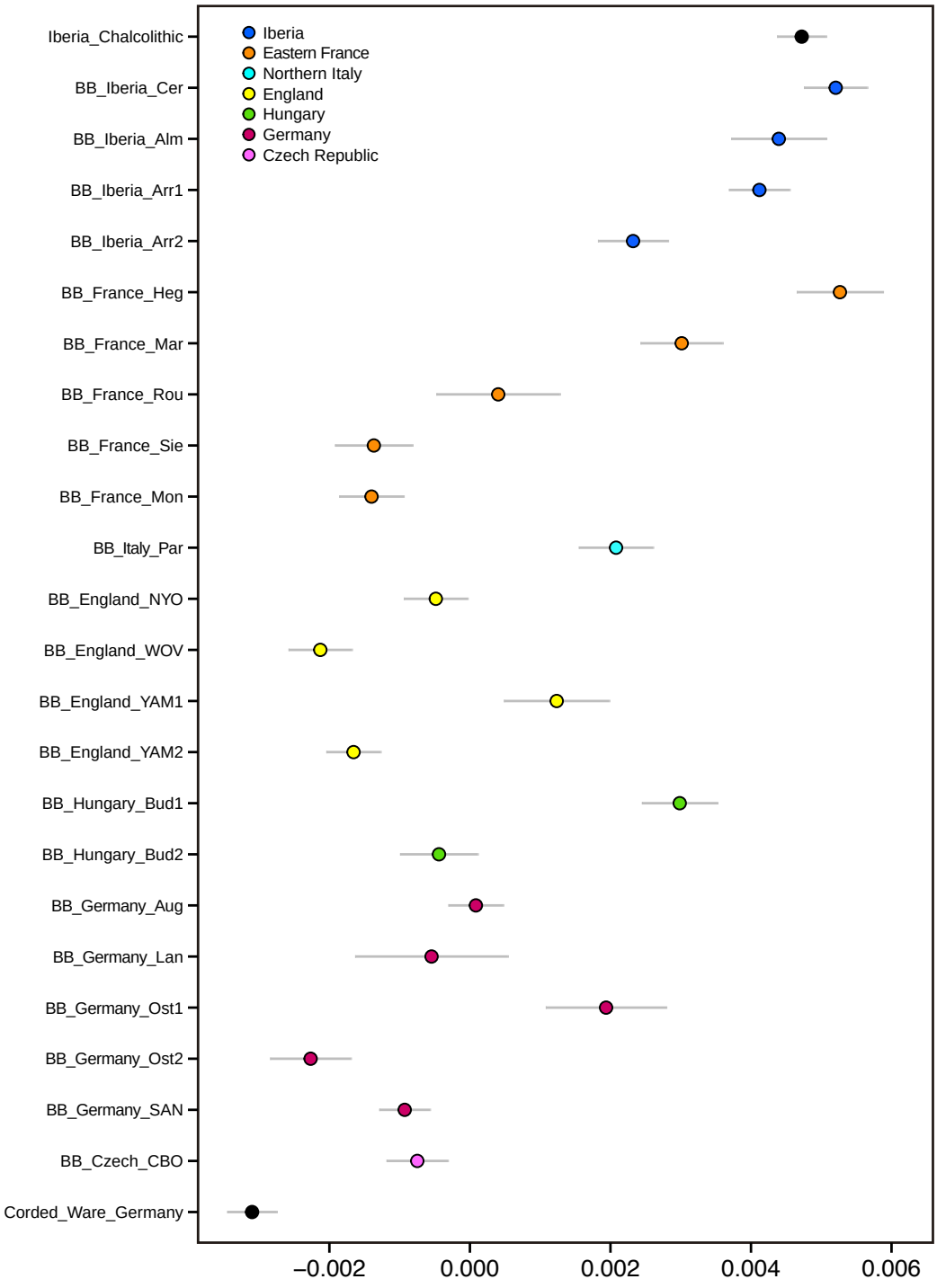
Electronic Appendix

Supplementary information for section 4.4

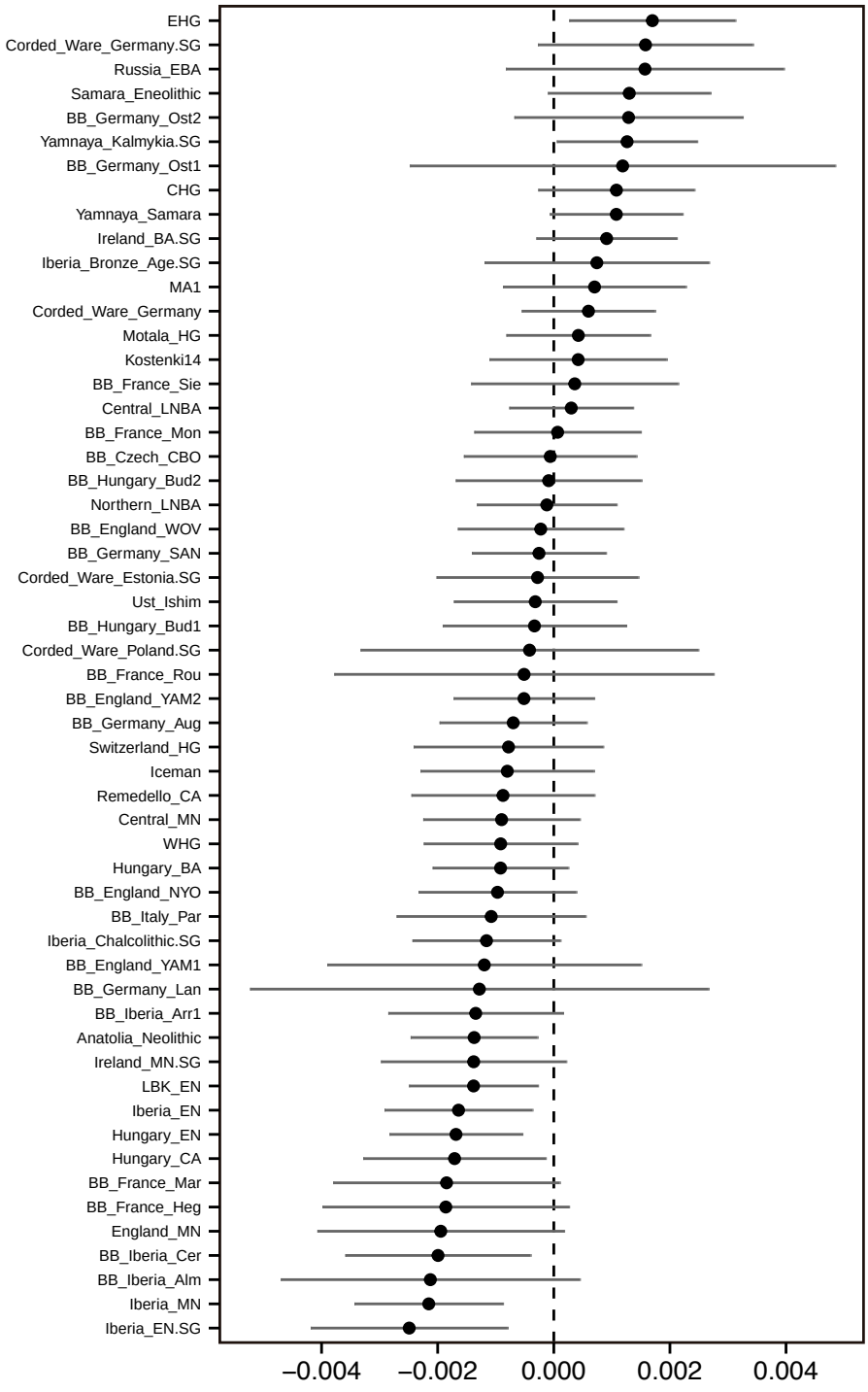
Supplementary Figures



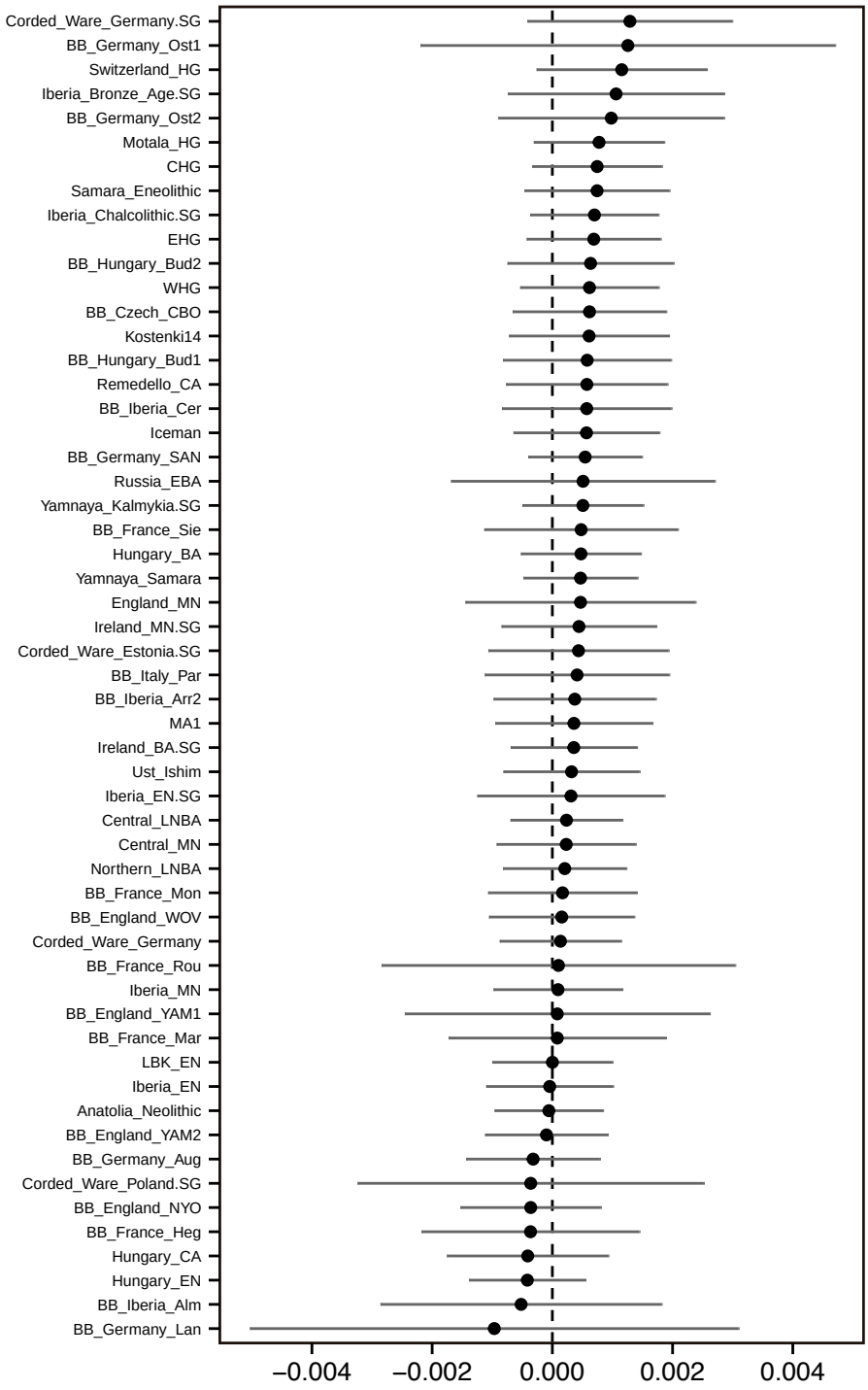
Supplementary Fig. 1. Principal component analysis of BBC individuals, other ancient individuals and present-day West Eurasian individuals from the Human Origins array (grey dots). Both transversions and transitions were used, with ancient individuals projected onto modern diversity.



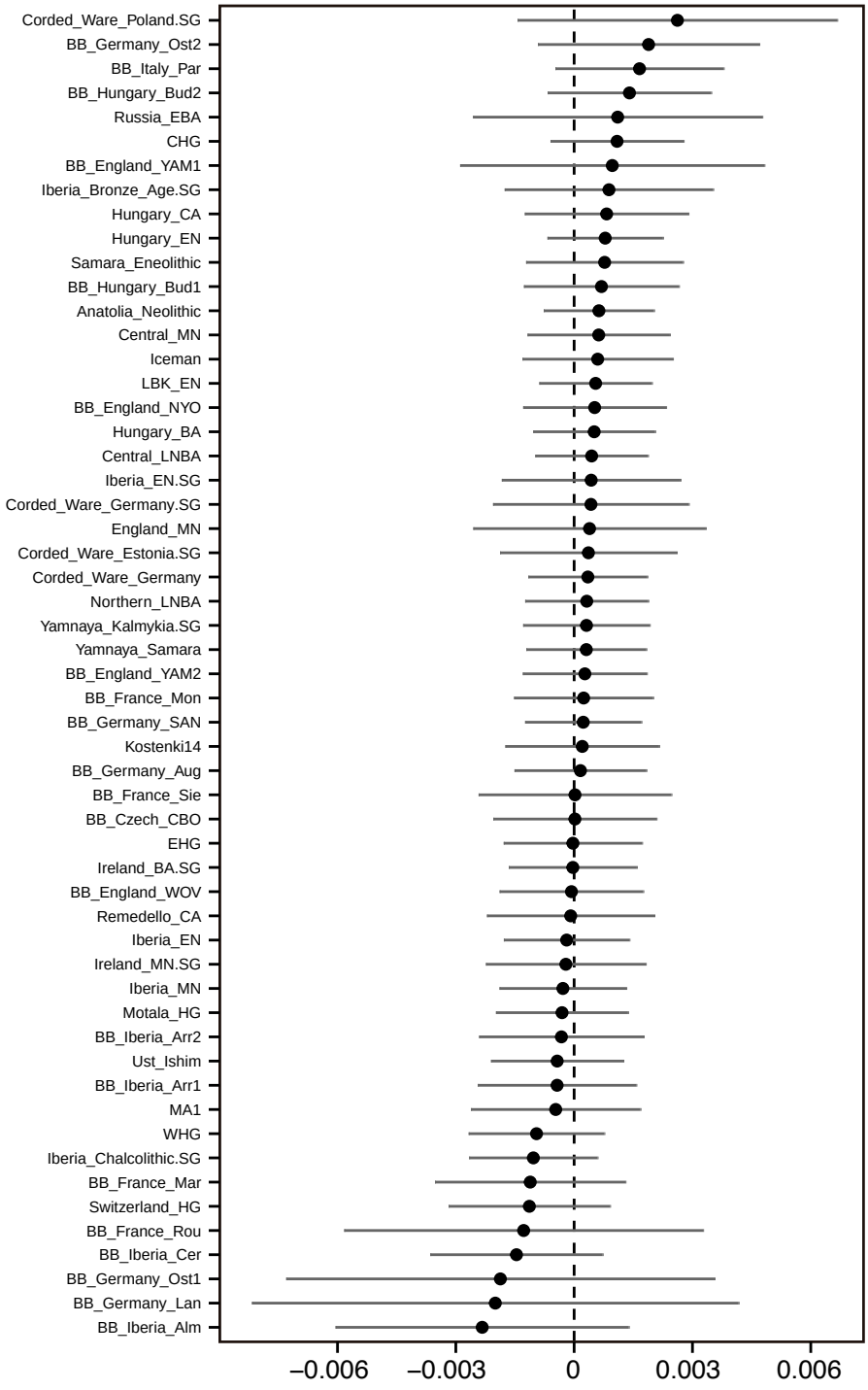
Supplementary Fig. 2. *f*-statistics of the form f_4(*Chimp*, *Test*; *Yamnaya_Samara*, *Anatolia_Neolithic*) using the grouping scheme in Supplementary Table 4. Error bars represent 1.5 standard errors.



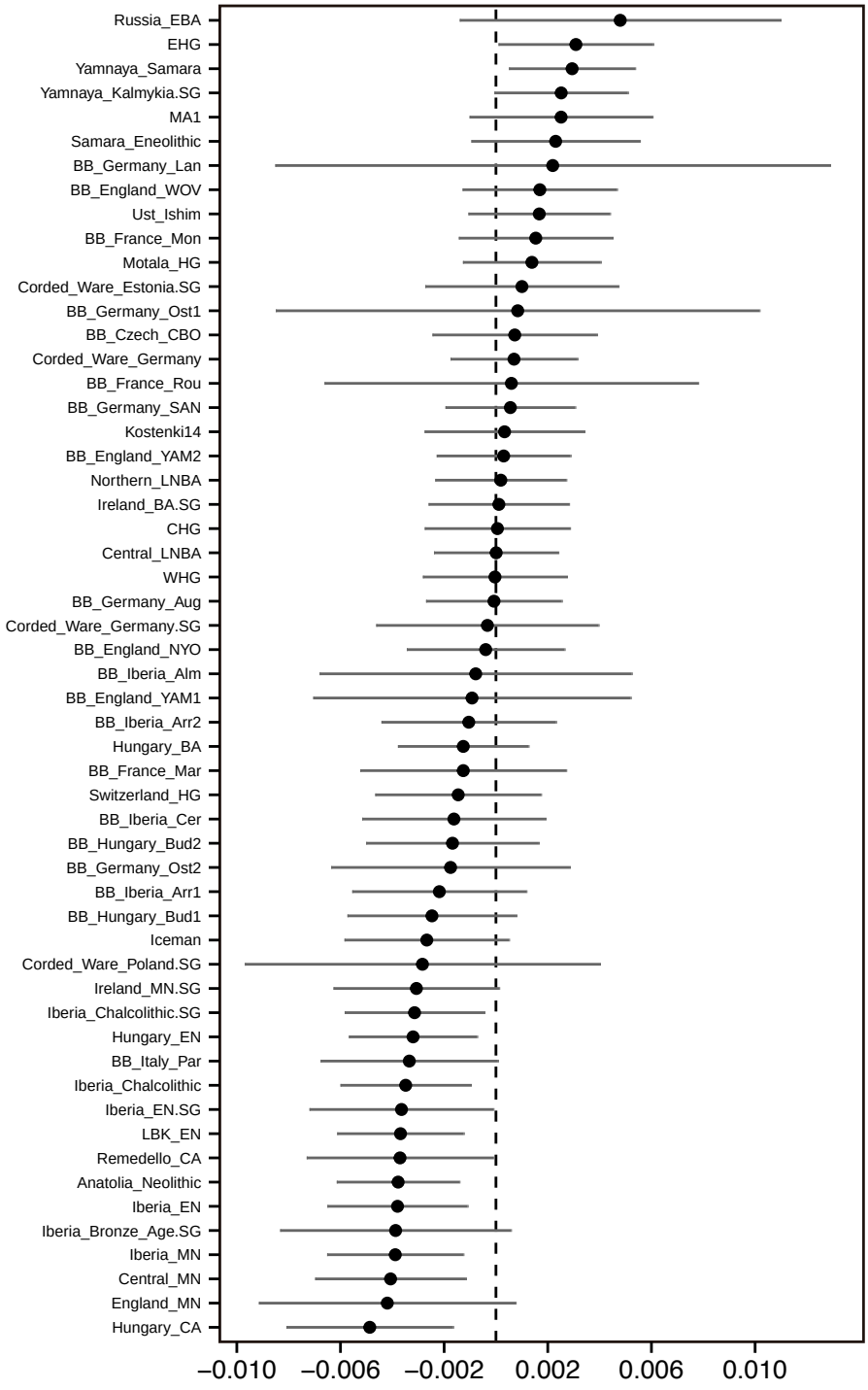
Supplementary Fig. 3. *f*-statistics of the form f_4(*Chimp*, *Test*; *Iberia_Chalcolithic*, *BB_Iberia_Arr2*). Error bars represent 3 standard errors.



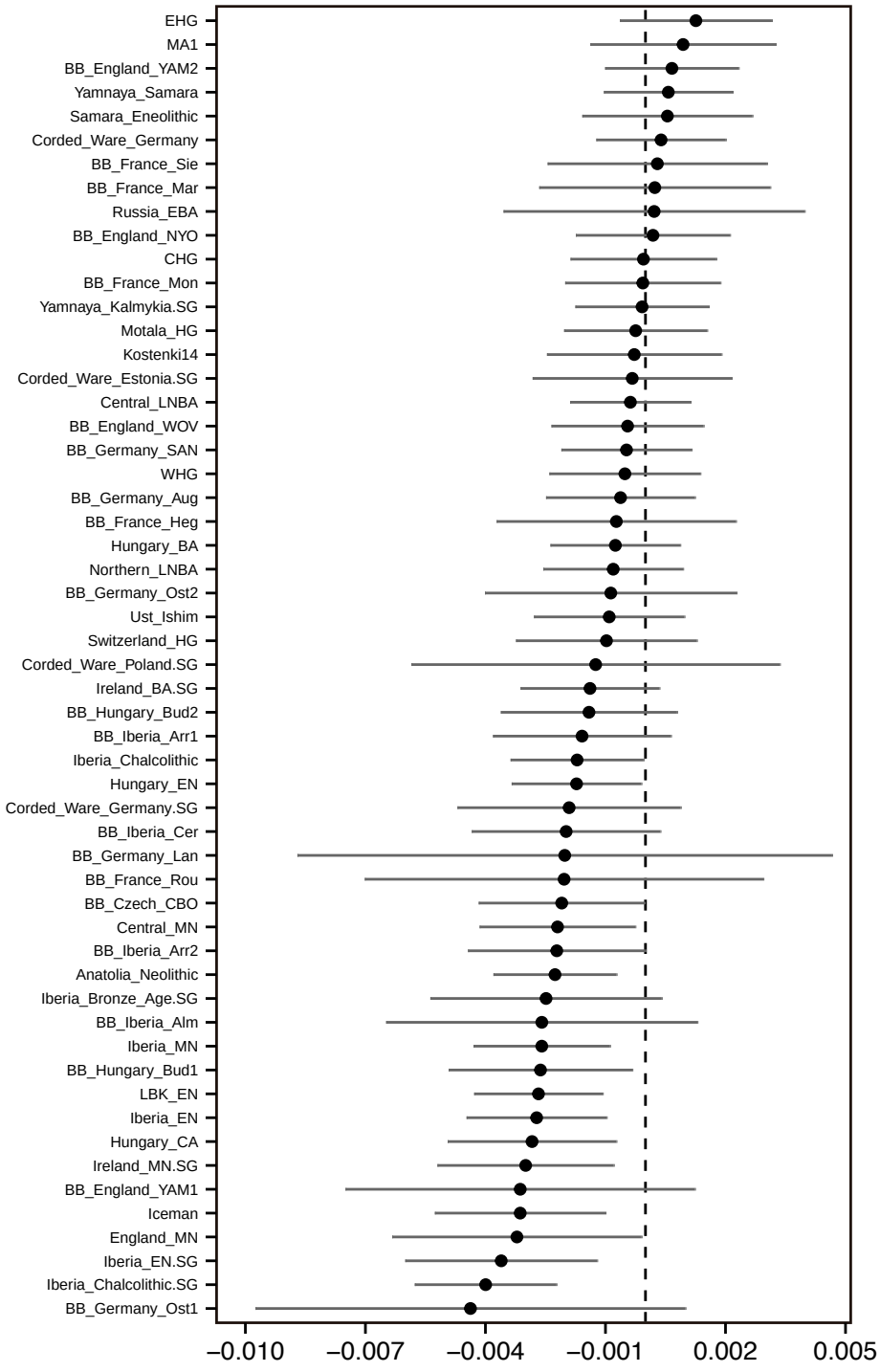
Supplementary Fig. 4. *f*-statistics of the form f_4(*Chimp, Test*; *Iberia_Chalcolithic*, *BB_Iberia_Arr1*). Error bars represent 3 standard errors.



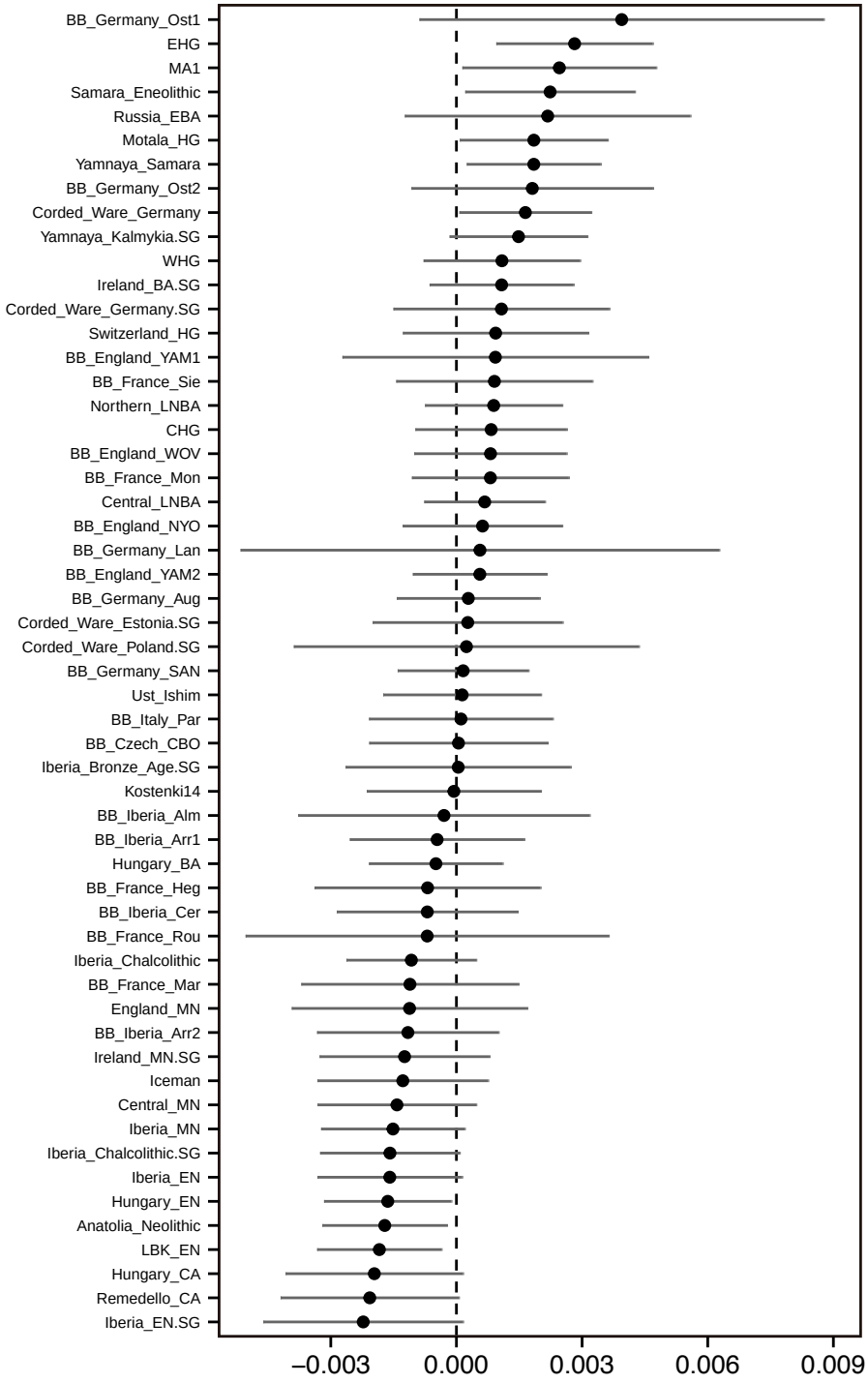
Supplementary Fig. 5. *f*-statistics of the form f_4(*Chimp*, *Test*; *Iberia_Chalcolithic*, *BB_France_Heg*). Error bars represent 3 standard errors.



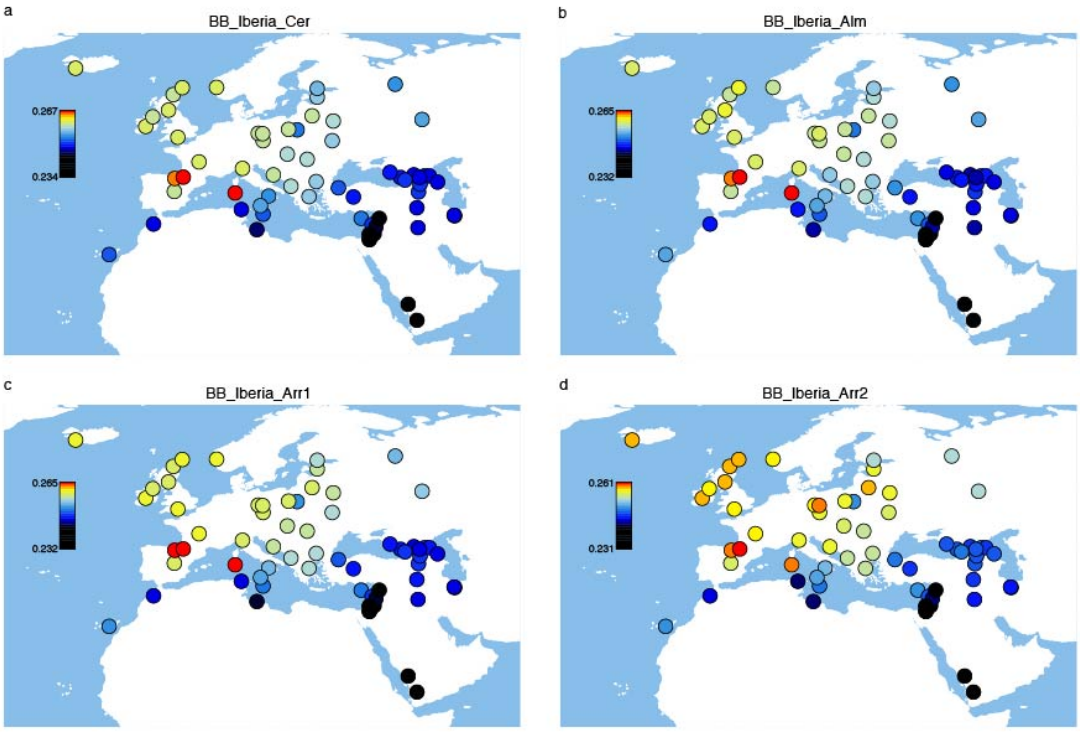
Supplementary Fig. 6. *f*-statistics of the form f_4(*Chimp*, *Test*; *BB_France_Heg*, *BB_France_Sie*). Error bars represent 3 standard errors.



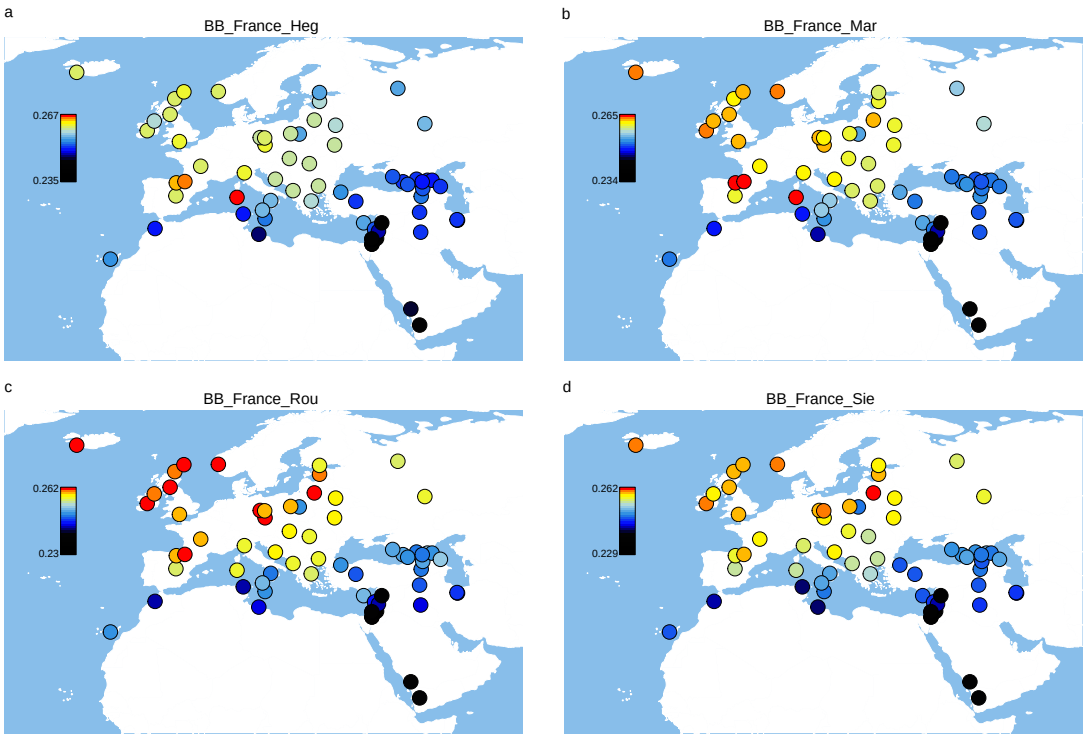
Supplementary Fig. 7. *f*-statistics of the form f_4(*Chimp, Test*; *Remedello_CA, BB_Italy_Par*). Error bars represent 3 standard errors.



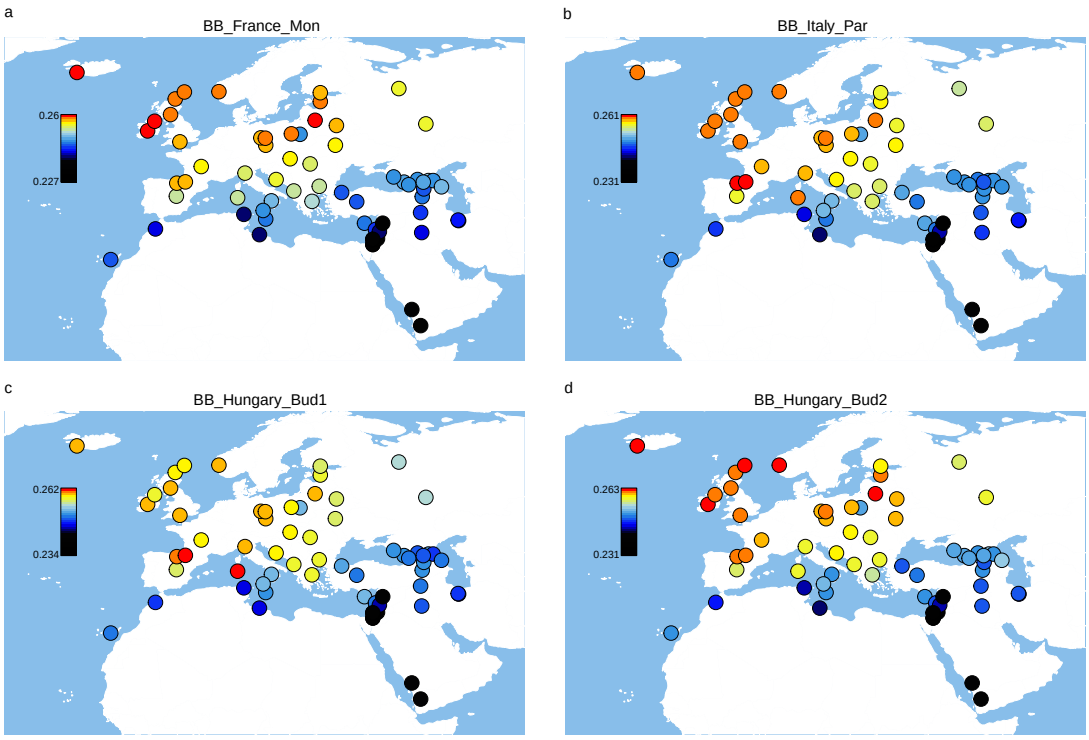
Supplementary Fig. 8. *f*-statistics of the form f_4(*Chimp*, *Test*; *BB_Hungary_Bud1*, *BB_Hungary_Bud2*). Error bars represent 3 standard errors.



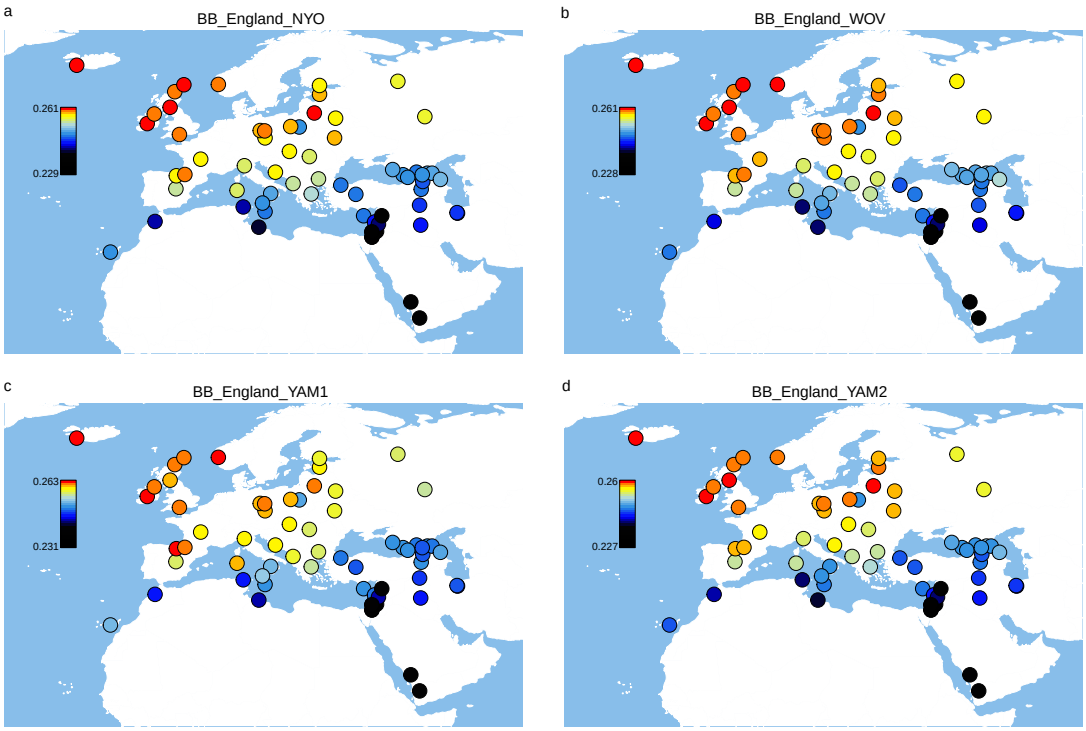
Supplementary Fig. 9. Shared genetic drift between Iberian Bell Beaker groups and present-day West Eurasians.



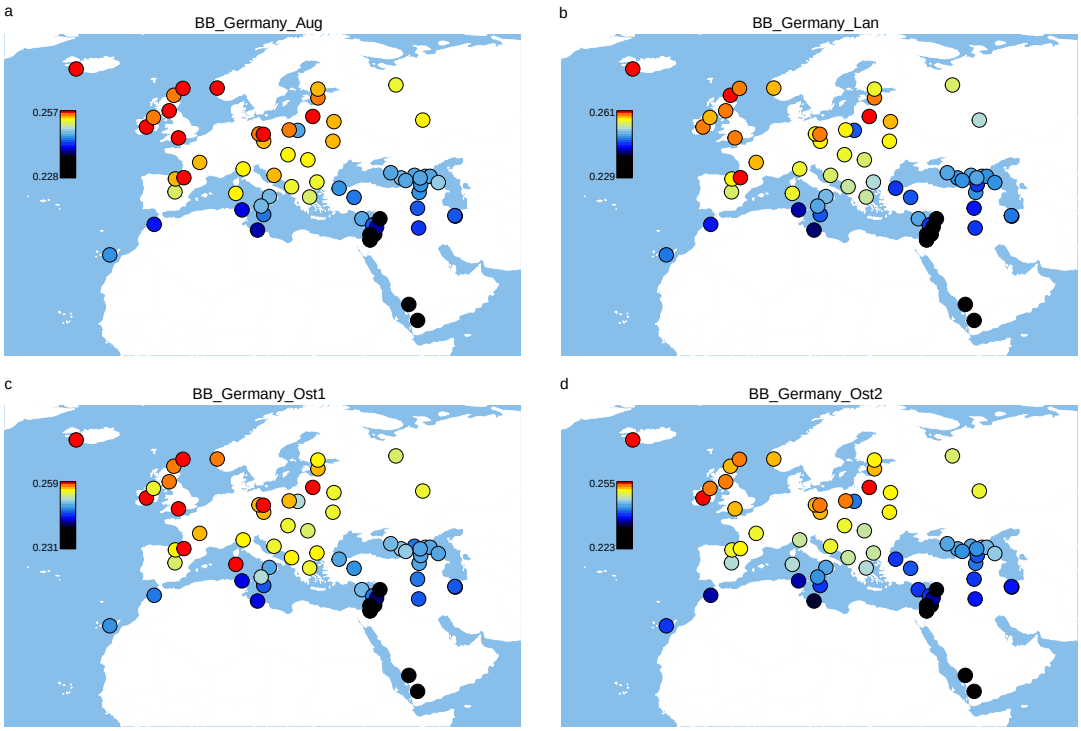
Supplementary Fig. 10. Shared genetic drift between Bell Beaker groups from Eastern France and present-day West Eurasians.



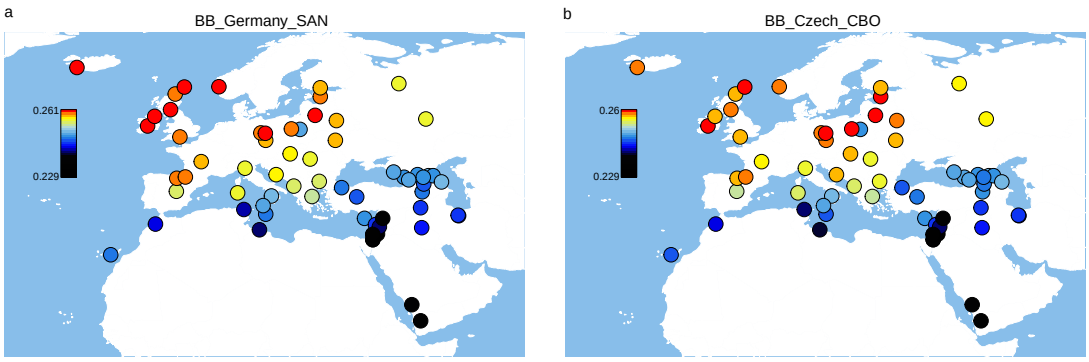
Supplementary Fig. 11. Shared genetic drift between present-day West Eurasians and Bell Beaker groups from a) Eastern France, b) Italy, c and d) Hungary.



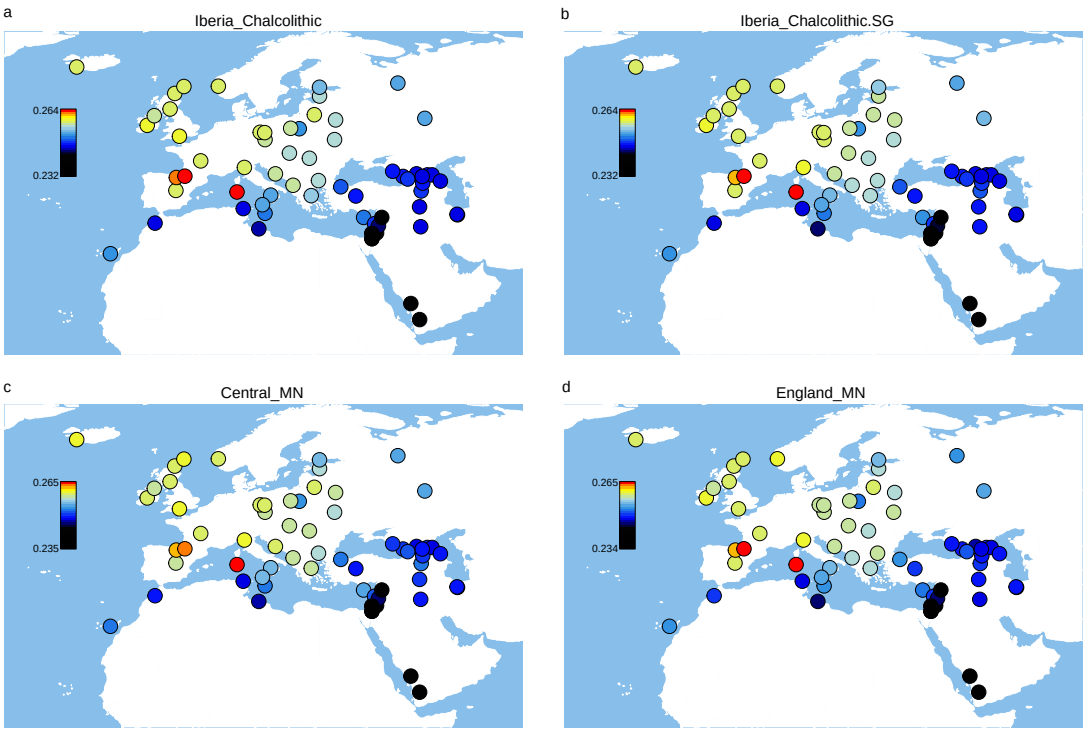
Supplementary Fig. 12. Shared genetic drift between Bell Beaker groups from England and present-day West Eurasians.



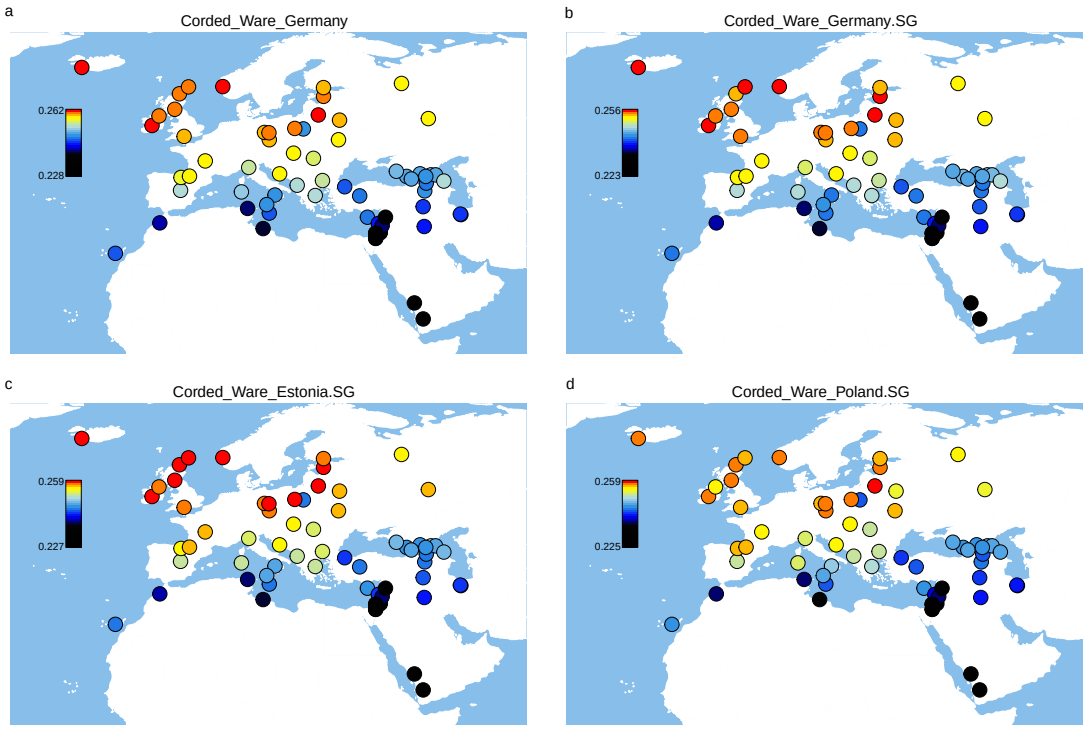
Supplementary Fig. 13. Shared genetic drift between Bell Beaker groups from Germany and present-day West Eurasians.



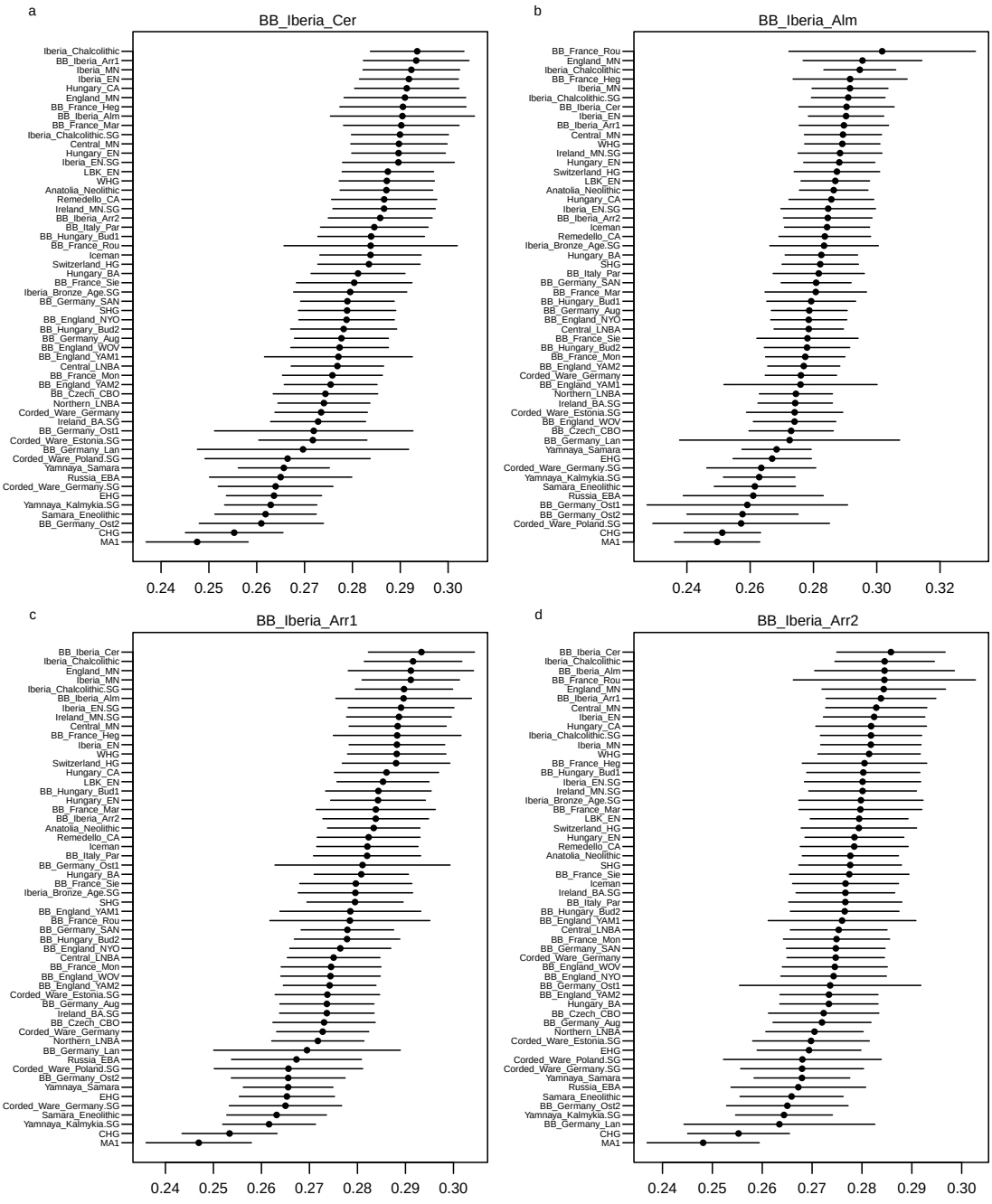
Supplementary Fig. 14. Shared genetic drift between present-day West Eurasians and Bell Beaker groups from a) Germany and b) Czech Republic.



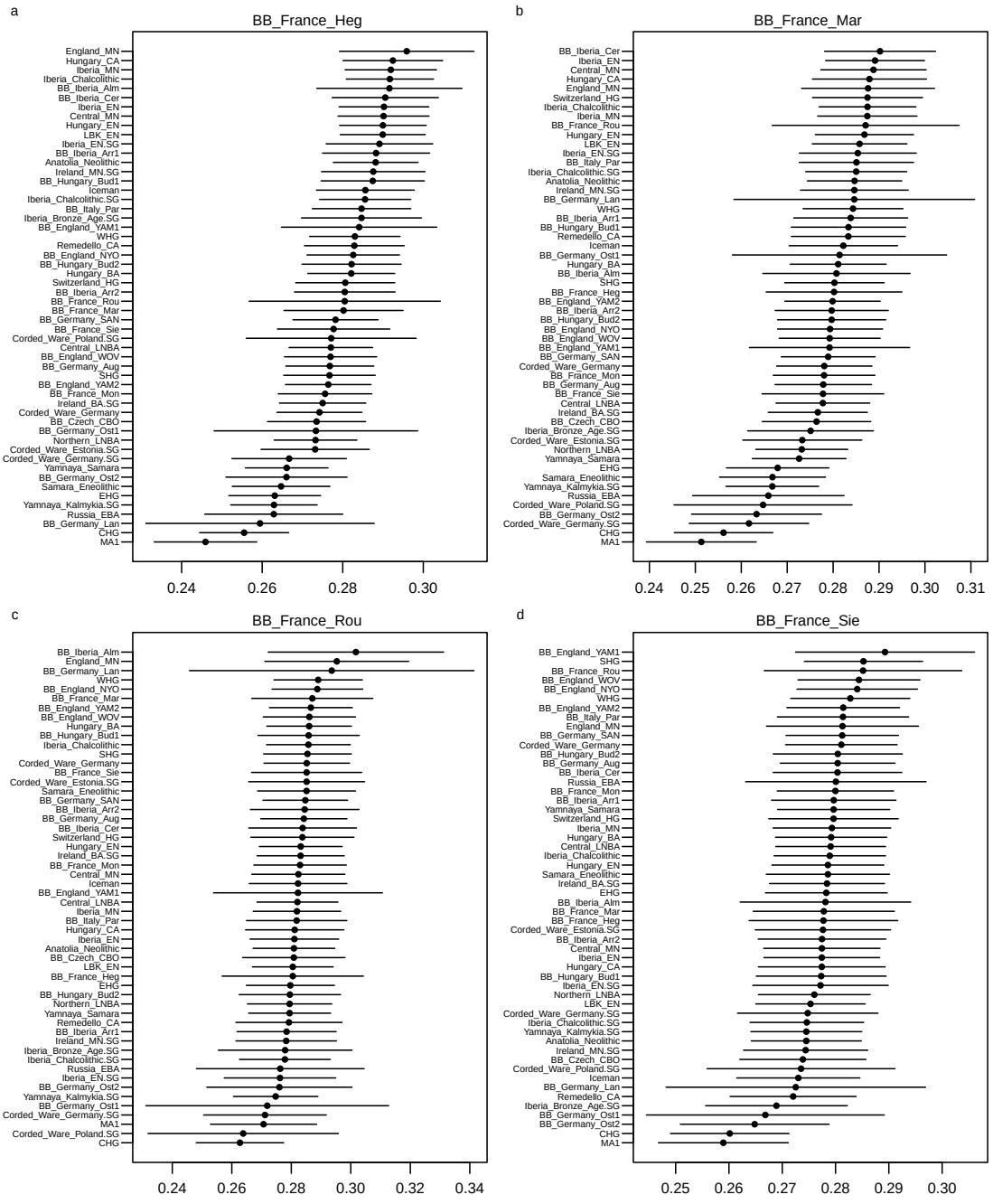
Supplementary Fig. 15. Shared genetic drift between Neolithic/Chalcolithic populations and present-day West Eurasians.



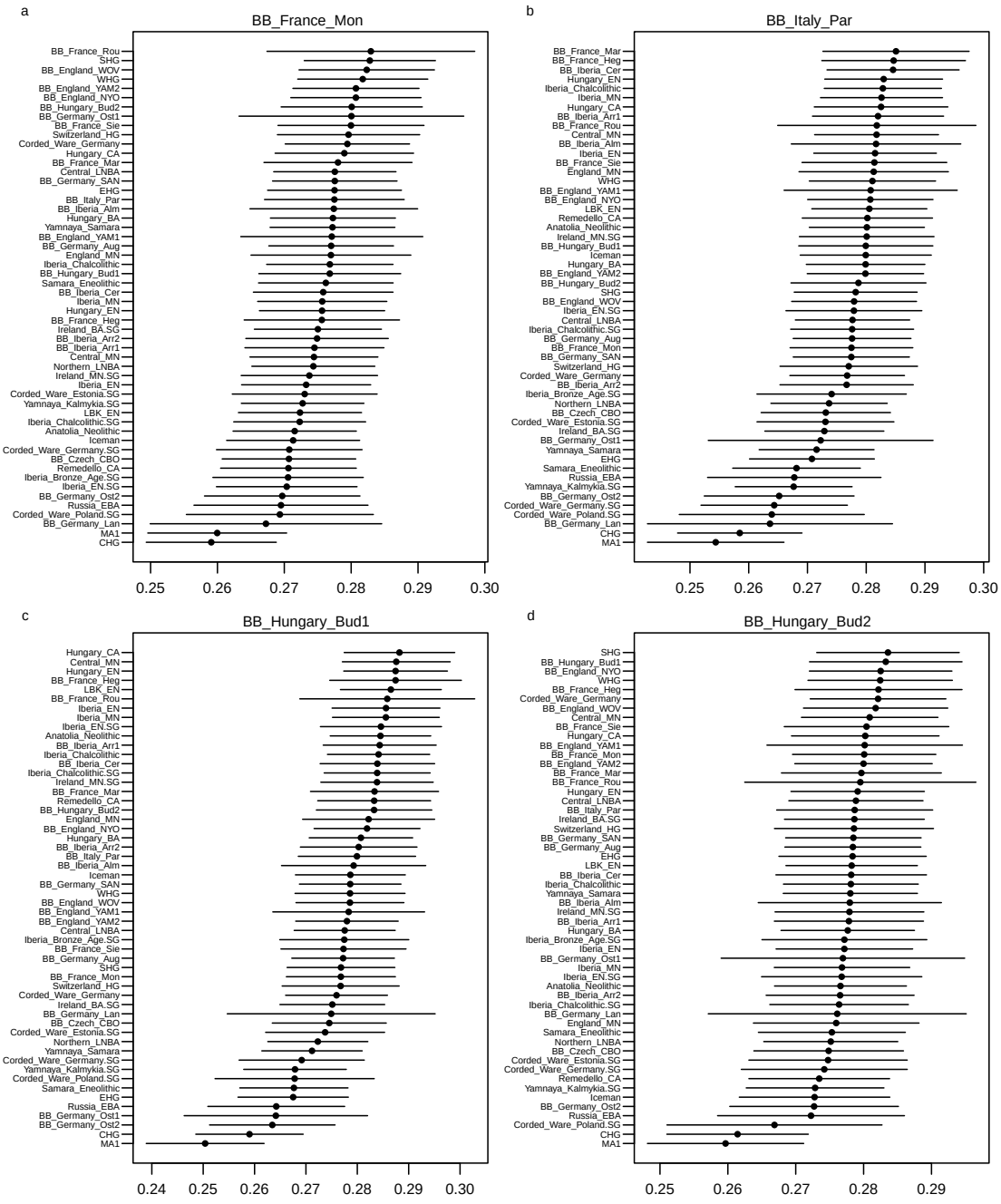
Supplementary Fig. 16. Shared genetic drift between Corded Ware groups and present-day West Eurasians.



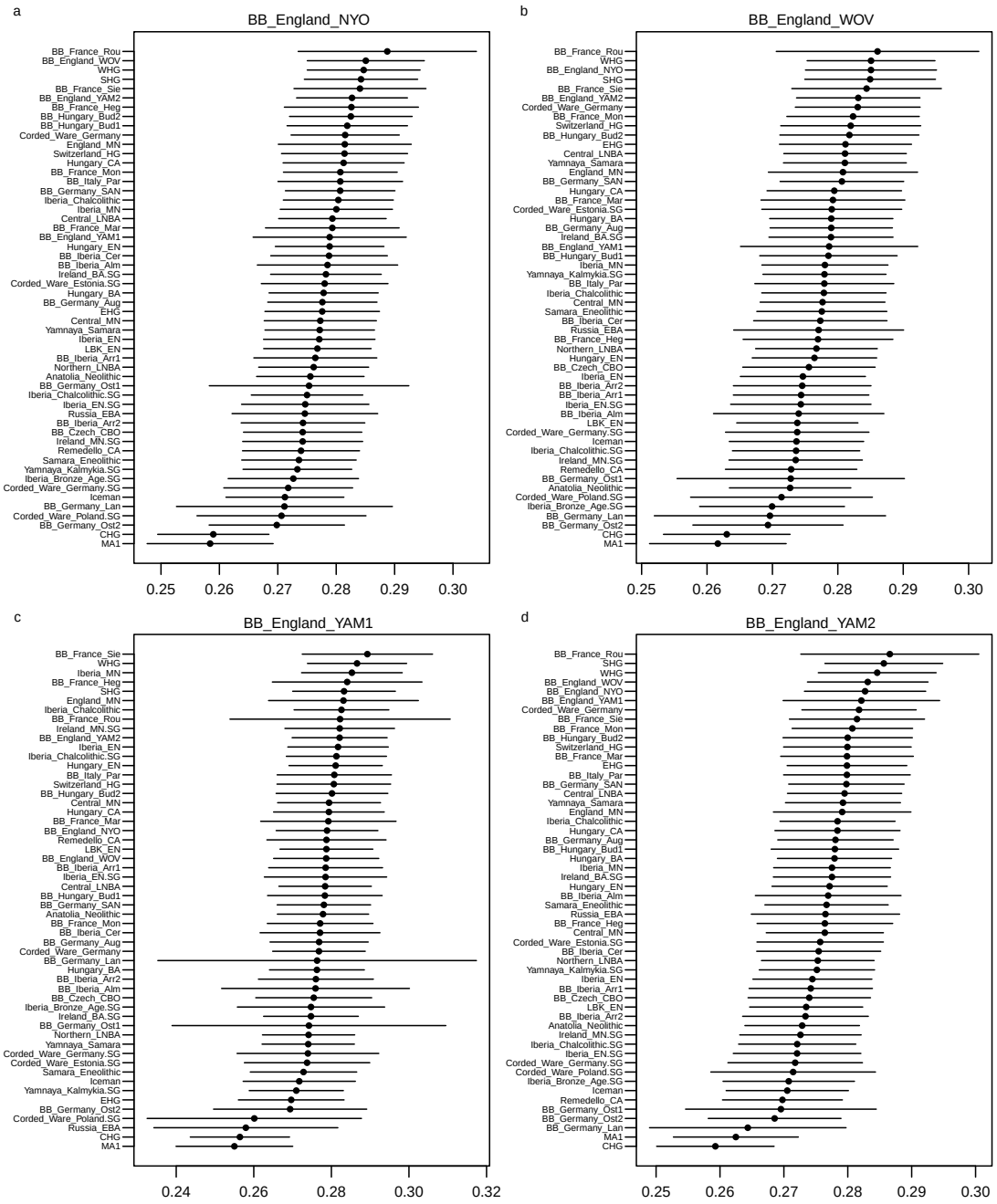
Supplementary Fig. 17. Shared genetic drift between Iberian Bell Beaker groups and ancient populations.



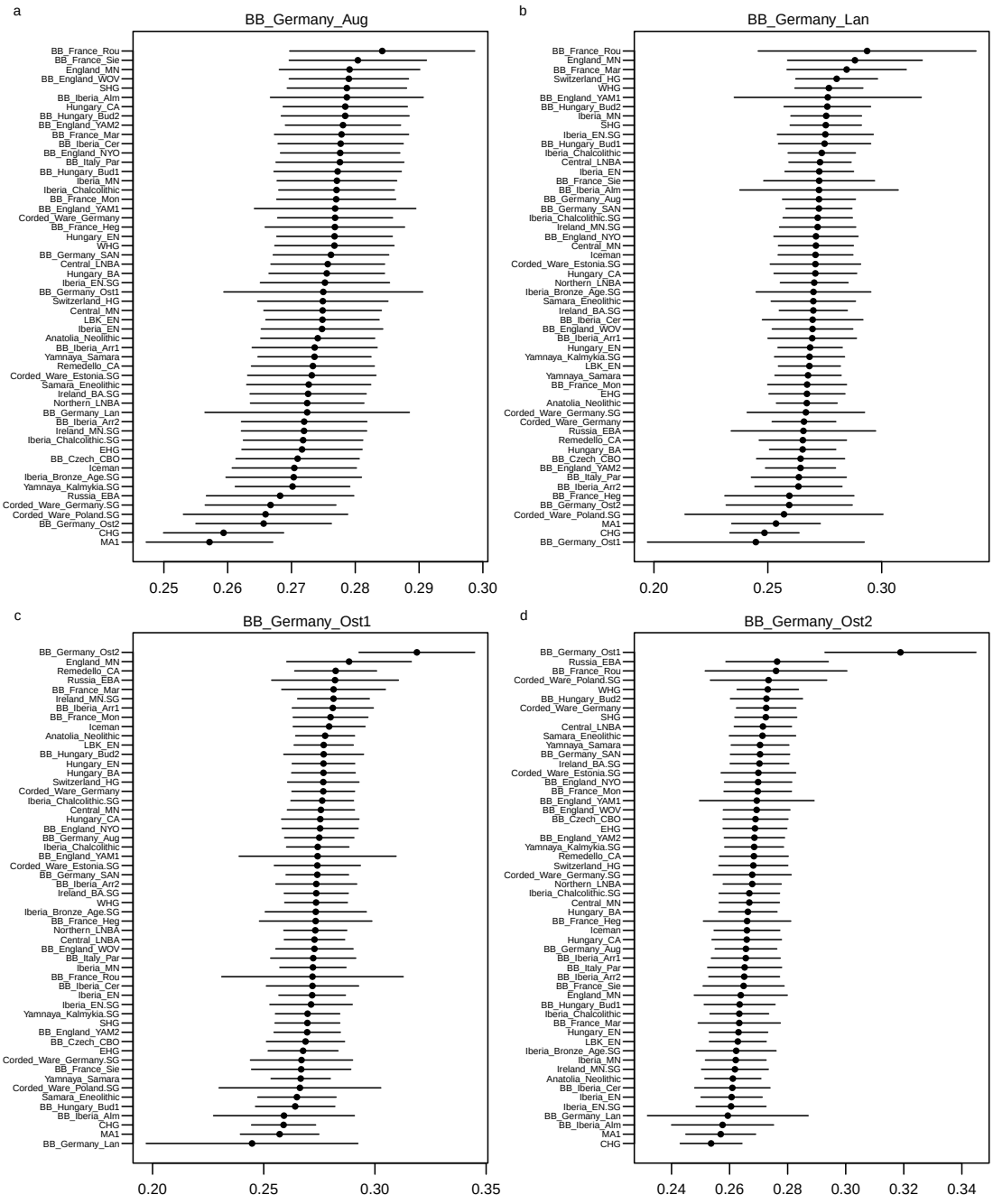
Supplementary Fig. 18. Shared genetic drift between Bell Beaker groups from Eastern France and ancient populations.



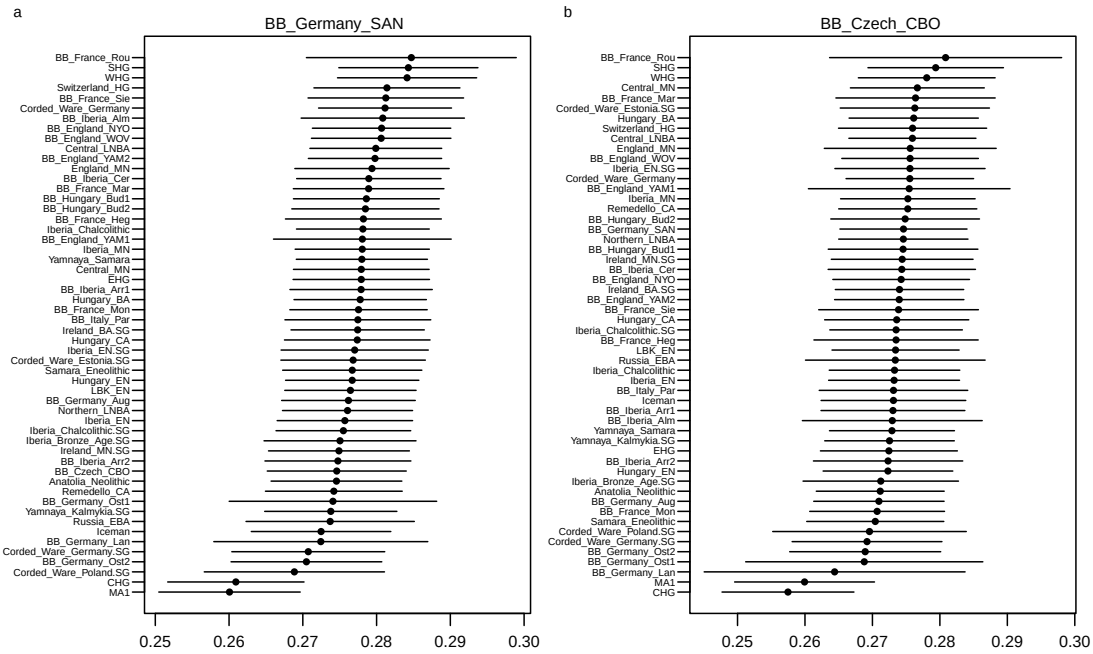
Supplementary Fig. 19. Shared genetic drift between ancient populations and Bell Beaker groups from a) Eastern France, b) Italy, c and d) Hungary.



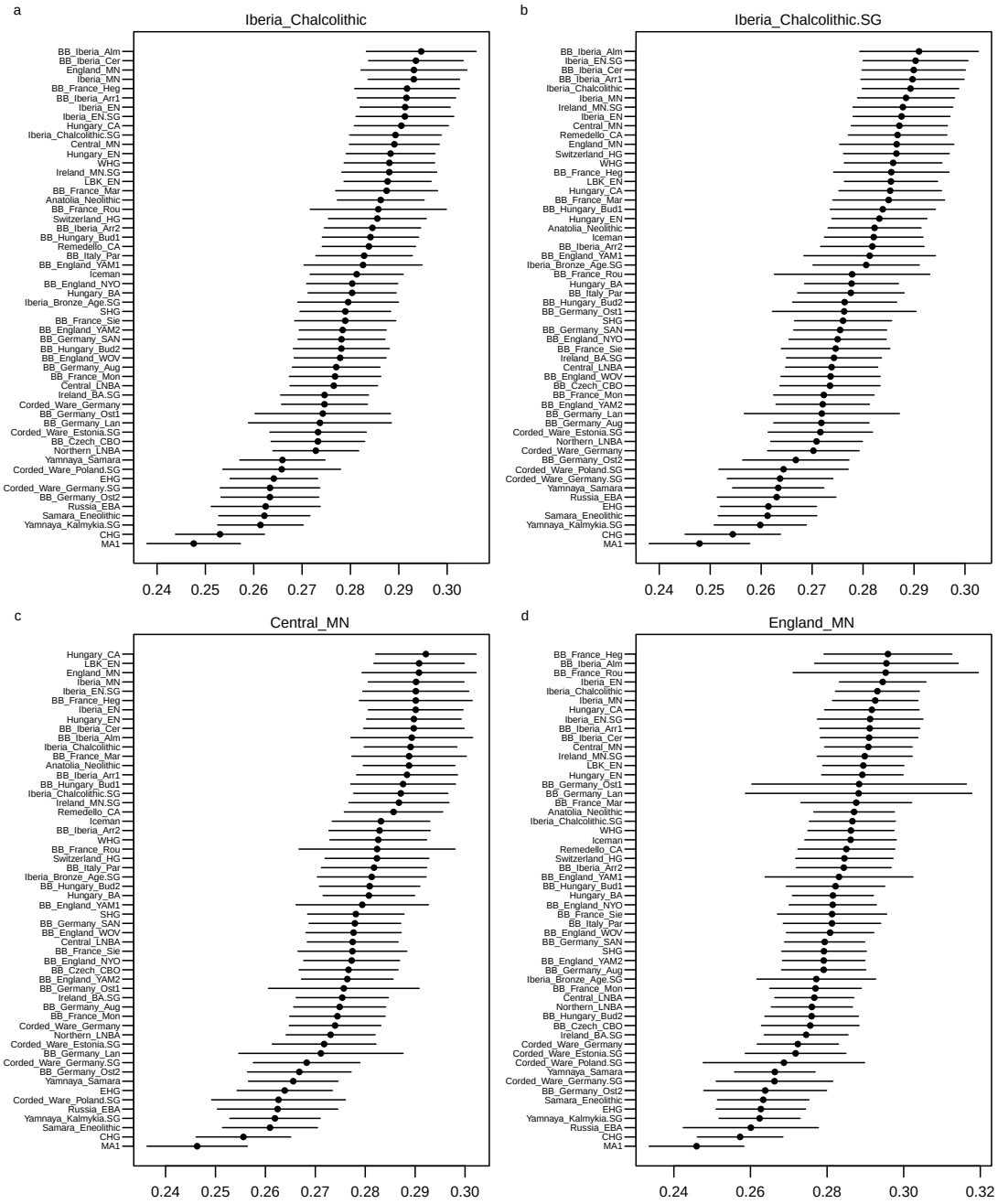
Supplementary Fig. 20. Shared genetic drift between Bell Beaker groups from England and ancient populations.



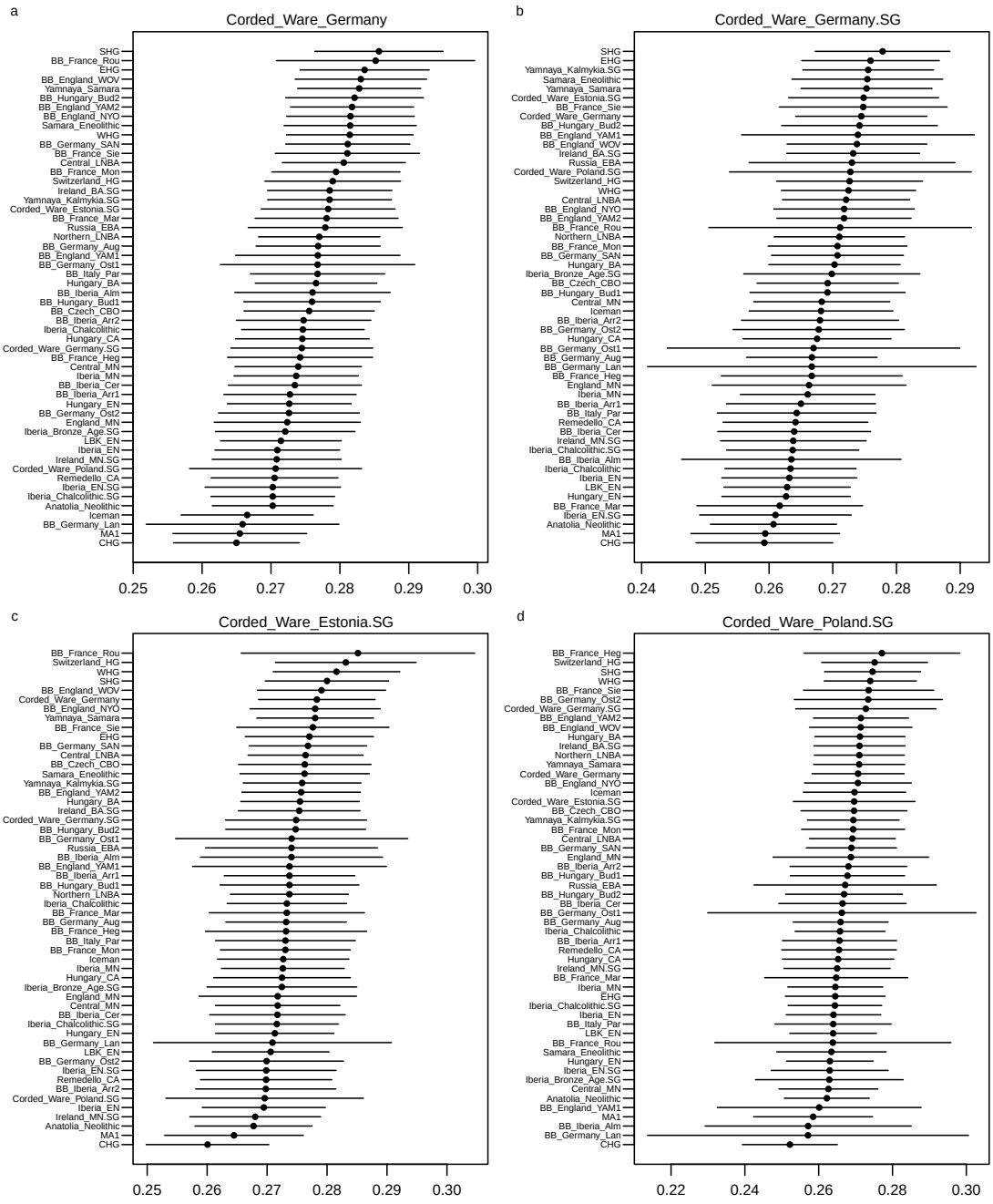
Supplementary Fig. 21. Shared genetic drift between Bell Beaker groups from Germany and ancient populations.



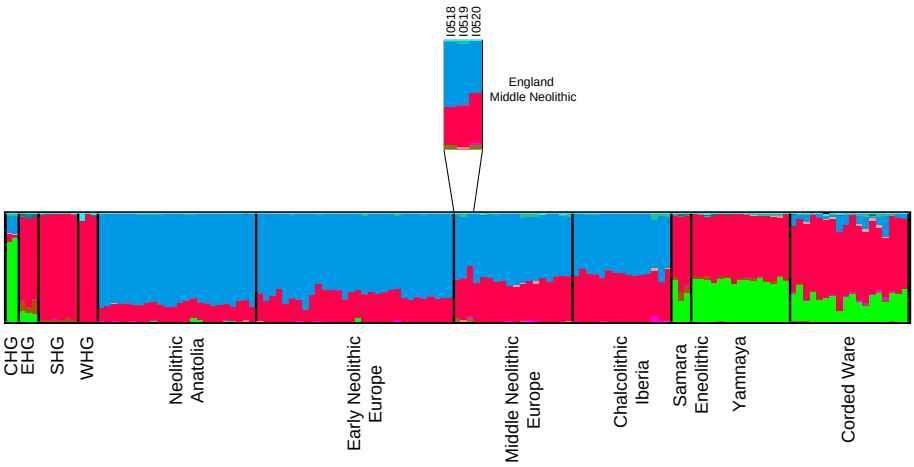
Supplementary Fig. 22 Shared genetic drift between ancient populations and Bell Beaker groups from a) Germany and b) Czech Republic.



Supplementary Fig. 23. Shared genetic drift between Neolithic/Chalcolithic populations and ancient populations.



Supplementary Fig. 24. Shared genetic drift between Corded Ware groups and ancient populations.



Supplementary Fig. 25. ADMIXTURE analysis. Equivalent to Fig. 2b but highlighting England_MN individuals.

Supplementary Tables

Supplementary Table 1. Mitochondrial haplogroup determination of newly reported individuals.

Sample ID	Mean Coverage	Haplogroup	Missing mutations	Private mutations	SNPs against rCRS
I0263	122.0	X2b+226			73G, 153G, 195C, 225A, 226C, 263G, 750G, 1438G, 1719A, 2706G, 4769G, 6221C, 6371T, 7028T, 8393T, 8860G, 11719A, 12705T, 13708A, 13966G, 14470C, 14766T, 15326G, 15927A, 16189C, 16223T, 16278T, 16519C
I1553	10.6	pre-H103	2352C, 3834A, 4243G, 16295T		152C, 263G, 750G, 1438G, 4769G, 7299G, 8860G, 15326G, 16519C
I0825	114.3	K1a4a1	16093C	15924G, heteroplasmy at pos 11671 (76% A, 24% G), heteroplasmy at pos 16154 (75% T, 25% C)	73G, 263G, 497T, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 6260A, 7028T, 8860G, 9055A, 9698C, 10398G, 10550G, 11299C, 11467G, 11485C, 11719A, 11840T, 12308G, 12372A, 13740C, 14167T, 14766T, 14798C, 15326G, 15924G, 16224C, 16311C, 16519C
I0260	329.6	K1a2a	16093C	9843G, 13145A	73G, 263G, 497T, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 5773A, 7028T, 8860G, 9055A, 9698C, 9843G, 10398G, 10550G, 11025C, 11299C, 11467G, 11719A, 12308G, 12372A, 13145A, 14167T, 14766T, 14798C, 15326G, 16224C, 16311C, 16519C
I0261	273.5	U5b1i			73G, 150T, 263G, 750G, 1438G, 2706G, 3105G, 3197C, 3498T, 4769G, 5656G, 6674C, 7028T, 7768G, 8860G, 9477A, 11467G, 11719A, 12308G, 12372A, 13617C, 14182C, 14766T, 15326G, 15777A, 16167T, 16192T, 16270T, 16311C, 16356C

I0823	88.7	H1		292C	263G, 292C, 750G, 1438G, 3010A, 4769G, 8860G, 15326G, 16519C
I0258	110.4	H1q			263G, 750G, 1438G, 3010A, 4769G, 4859C, 8860G, 15326G, 16519C
I0262	275.0	U5b3		16192C!	73G, 150T, 228A, 263G, 750G, 1438G, 2706G, 3197C, 4769G, 7028T, 7226A, 7768G, 8860G, 9477A, 11467G, 11719A, 12308G, 12372A, 13617C, 14182C, 14766T, 15326G, 16270T, 16304C
I0257	319.8	H1ax		Heteroplasmy at pos 146 (49% T, 51% C)	263G, 750G, 1438G, 3010A, 4769G, 5054C, 7471T, 8429T, 8860G, 15326G, 16519C
I0826	107.6	H1t			263G, 750G, 1438G, 3010A, 4769G, 8860G, 9986A, 15326G, 16519C
I0839	1.2	-	-	-	-
I0840	7.0	U5b1i	73G, 150T, 750G, 12372A, 16167T, 16192	12017G	263G, 1438G, 2706G, 3105G, 3197C, 3498T, 4769G, 5656G, 6674C, 7028T, 7768G, 8860G, 9477A, 11467G, 11719A, 12017G, 12308G, 13617C, 14182C, 14766T, 15326G, 15777A, 16270T, 16311C, 16356C
I0460	240.6	H45		93G, 16188T	93G, 263G, 750G, 1438G, 4769G, 8843C, 8860G, 15326G, 16188T, 16519C
I0458	320.2	K1a1b1		456T	73G, 114T, 263G, 456T, 497T, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 7028T, 8860G, 9055A, 9698C, 10398G, 10550G, 11299C, 11467G, 11470G, 11719A, 11914A, 12308G, 12372A, 14167T, 14766T, 14798C, 15326G, 15924G, 16093C, 16224C, 16311C, 16519C
I0459	427.0	U5b2b		10908C	73G, 150T, 263G, 750G, 1438G, 1721T, 2706G, 3197C, 4769G, 7028T, 7768G, 8860G, 9477A, 10908C, 11467G, 11653G, 11719A, 12308G, 12372A, 12634G, 13617C, 13630G, 13637G, 14182C, 14766T, 15326G, 16270T

I0461	306.3	K1a1b1	114T	6261A, 12599C, heteroplasmy at pos 16093 (74% T, 26% C)	73G, 263G, 497T, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 6261A, 7028T, 8860G, 9055A, 9698C, 10398G, 10550G, 11299C, 11467G, 11470G, 11719A, 11914A, 12308G, 12372A, 12599C, 14167T, 14766T, 14798C, 15326G, 15924G, 16093C (26%), 16224C, 16311C, 16519C
I0462	240.5	K1a+195		1811A!	73G, 195C, 263G, 497T, 750G, 1189C, 1438G, 2706G, 3480G, 4769G, 7028T, 8860G, 9055A, 9698C, 10398G, 10550G, 11299C, 11467G, 11719A, 12308G, 12372A, 14167T, 14766T, 14798C, 15326G, 16093C, 16224C, 16311C, 16519C
I1392	221.7	H1+152		5372C, 12630A	152C, 263G, 750G, 1438G, 3010A, 4769G, 5372C, 8860G, 12630A, 15326G, 16519C
I1388	625.1	H		16230G	263G, 750G, 1438G, 4769G, 8860G, 15326G, 16230G, 16519C
I1391	295.0	J1c4	228A	7271G, 13641C	73G, 185A, 263G, 295T, 462T, 489C, 750G, 1438G, 2706G, 3010A, 4216C, 4769G, 7028T, 7271G, 8860G, 9632G, 10398G, 11251G, 11719A, 12083G, 12612G, 13641C, 13708A, 14766T, 14798C, 15326G, 15452A, 16069T, 16126C
I1389	133.4	X2b4a			73G, 153G, 195C, 225A, 226C, 263G, 750G, 1438G, 1719A, 2706G, 3705A, 4769G, 6221C, 6371T, 7028T, 8393T, 8592A, 8860G, 11719A, 12705T, 13708A, 13966G, 14470C, 14766T, 15326G, 15927A, 16189C, 16223T, 16278T, 16519C
I1390	177.9	X2b4a			73G, 153G, 195C, 225A, 226C, 263G, 750G, 1438G, 1719A, 2706G, 3705A, 4769G, 6221C, 6371T, 7028T, 8393T, 8592A, 8860G, 11719A, 12705T, 13708A, 13966G, 14470C, 14766T, 15326G, 15927A, 16189C, 16223T, 16278T, 16519C
I1381	599.2	H		Heteroplasmy at pos 14060 (80% T, 20% C)	263G, 750G, 1438G, 4769G, 8860G, 15326G, 16519C
I1382	126.3	U5a2c3a			73G, 263G, 750G, 1438G, 2706G, 3197C, 4769G, 7028T, 7960C, 8860G, 9477A, 10619T, 10709C, 11465C, 11467G, 11719A, 12308G, 12372A, 13617C, 14766T, 14793G, 15326G, 16256T, 16270T, 16526A

I2478	251.1	K1a2a	16093C	5492C	73G, 263G, 497T, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 5492C, 5773A, 7028T, 8860G, 9055A, 9698C, 10398G, 10550G, 11025C, 11299C, 11467G, 11719A, 12308G, 12372A, 14167T, 14766T, 14798C, 15326G, 16224C, 16311C, 16519C
I1767	472.0	U5a1a1			73G, 263G, 750G, 1438G, 1700C, 2706G, 3197C, 4769G, 5495C, 7028T, 8860G, 9477A, 11467G, 11719A, 12308G, 12372A, 13617C, 14766T, 14793G, 15218G, 15326G, 15924G, 16256T, 16270T, 16399G
I1770	70.7	U5b1d2		15851G, 16192C!	73G, 150T, 263G, 750G, 1438G, 2706G, 3197C, 4769G, 5437T, 5656G, 7028T, 7085C, 7768G, 8860G, 9477A, 11467G, 11719A, 12308G, 12372A, 13617C, 14182C, 14766T, 15326G, 15851G, 16239T, 16270T
I2453	906.1	K2a			73G, 146C, 152C, 263G, 709A, 750G, 1438G, 1811G, 2706G, 3480G, 4561C, 4769G, 7028T, 8860G, 9055A, 9698C, 9716C, 10550G, 11299C, 11467G, 11719A, 12308G, 12372A, 14167T, 14766T, 14798C, 15326G, 16224C, 16311C, 16519C
I2454	171.8	R1a1a			73G, 263G, 295A, 750G, 1391C, 1438G, 2706G, 3360G, 4026G, 4769G, 4917G, 5378G, 5586T, 5823G, 6557T, 6671C, 7028T, 7424G, 7547C, 8388C, 8860G, 8887G, 10658G, 10825G, 11719A, 13948T, 14632T, 14766T, 15326G, 15721C, 16311C, 16519C
I2455	1064.7	U5a2c3a		Heteroplasmy at pos 16270 (40% C, 60% T)	73G, 263G, 750G, 1438G, 2706G, 3197C, 4769G, 7028T, 7960C, 8860G, 9477A, 10619T, 10709C, 11465C, 11467G, 11719A, 12308G, 12372A, 13617C, 14766T, 14793G, 15326G, 16256T, 16270 (60%), 16526A
I2444	189.5	T2c1d+152	16296T	8711G	73G, 146C, 152C, 263G, 279C, 709A, 750G, 1438G, 1888A, 2706G, 4216C, 4769G, 4917G, 5187T, 6261A, 7028T, 7873T, 8697A, 8711G, 8860G, 10463C, 10822T, 11251G, 11719A, 11812G, 13368A, 14233G, 14766T, 14905A, 15326G, 15452A, 15607G, 15928A, 16126C, 16292T, 16294T, 16519C

I2445	1070.1	X2b6		73G, 153G, 195C, 225A, 226C, 263G, 750G, 1438G, 1719A, 2706G, 4577T, 4769G, 6221C, 6371T, 7028T, 7570G, 8393T, 8860G, 11719A, 11914A, 12705T, 13708A, 13966G, 14016A, 14470C, 14766T, 15326G, 15927A, 16183C, 16189C, 16223T, 16278T, 16519C
I2416	421.2	K1b1a1		73G, 152C, 263G, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 5913A, 7028T, 8860G, 9055A, 9698C, 9962A, 10289G, 10398G, 10550G, 11299C, 11467G, 11719A, 11923G, 12308G, 12372A, 13967T, 14167T, 14766T, 14798C, 15257A, 15326G, 15946T, 16093C, 16224C, 16311C, 16319A, 16463G, 16519C
I2418	752.0	K1a4a1	16093C	73G, 263G, 497T, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 6260A, 7028T, 8860G, 9055A, 9698C, 10398G, 10550G, 11299C, 11467G, 11485C, 11719A, 11840T, 12308G, 12372A, 13740C, 14167T, 14766T, 14798C, 15326G, 16224C, 16311C, 16519C
I2457	542.2	I4a		73G, 199C, 204C, 250C, 263G, 573.XC, 750G, 1438G, 1719A, 2706G, 4529T, 4769G, 7028T, 8251A, 8519A, 8860G, 10034C, 10238C, 10398G, 10819G, 11719A, 12501A, 12705T, 13780G, 14766T, 15043A, 15326G, 15924G, 16129A, 16223T, 16391A, 16519C
I2459	340.1	T2a1a		73G, 263G, 709A, 750G, 1438G, 1888A, 2706G, 2850C, 4216C, 4769G, 4917G, 7022C, 7028T, 8697A, 8860G, 10463C, 11251G, 11719A, 11812G, 13368A, 13965C, 14233G, 14687G, 14766T, 14905A, 15326G, 15452A, 15607G, 15928A, 16126C, 16294T, 16296T, 16519C
I2364	119.7	U5a2b		73G, 263G, 750G, 1438G, 2706G, 3197C, 4769G, 7028T, 8860G, 9477A, 9548A, 11467G, 11719A, 12308G, 12372A, 13617C, 14766T, 14793G, 15326G, 16192T, 16256T, 16270T, 16526A
I2365	115.5	V3		72C, 263G, 750G, 1438G, 2706G, 4580A, 4769G, 7028T, 8860G, 12810G, 15326G, 15904T, 16298C

I0518	25.8	K1a3a1	16093C		73G, 263G, 497T, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 7028T, 7559G, 8440G, 8860G, 9055A, 9698C, 10398G, 10550G, 11299C, 11467G, 11719A, 12308G, 12372A, 13117G, 14167T, 14766T, 14798C, 15326G, 16224C, 16311C, 16519C
I0519	441.6	X2b+226		13225A	73G, 153G, 195C, 225A, 226C, 263G, 750G, 1438G, 1719A, 2706G, 4769G, 6221C, 6371T, 7028T, 8393T, 8860G, 11719A, 12705T, 13225A, 13708A, 13966G, 14470C, 14766T, 15326G, 15927A, 16189C, 16223T, 16278T, 16519C
I0520	44.3	U5a2c	143A, 152C, 4721G, 5354T, 6340T, 10055G, 10082T, 13521T, 16311C		73G, 143A, 152C, 263G, 750G, 1438G, 2706G, 3197C, 4721G, 4769G, 5354T, 6340T, 7028T, 8860G, 9477A, 10055G, 10082T, 10619T, 11467G, 11719A, 12308G, 12372A, 13521T, 13617C, 14766T, 14793G, 15326G, 16192T, 16256T, 16270T, 16311C, 16526A

Positions in red were represented by less than 2 reads and thus were not included in the analysis

Supplementary Table 2. Y-chromosome haplogroups and derived positions of Bell Beaker individuals.

ID	Site ID	Y haplogroup	Derived SNPs
I0257	Cer	R?	R:P224:17285993C->T;
I0261	Cer	P?	P:F91:7628900G->A;
I0825	Cer	G2	G2:F1189:8427005A->G; G:CTS9894:19124322A->T; G:L522:17533325A->C; G:M3470:7830068T->C; G:M3487:8602816G->C;
I0826	Cer	I2a2	I2a2:P218:17493630T->G; I:CTS11540:23156725C->T; I:PF3817:21939618G->A;
I0458	Arr	I2a2a	I2a2a:P220:24475669G->T; I2a2:M436:18747493G->C; I:CTS6497:16939794A->T; I:CTS8333:17940414G->A; I:CTS8742:18172947A->G; I:FGC2412:21689728A->G; I:PF3836:22525421T->G;
I0460	Arr	I2a2a1	I2a2a1:CTS616:6906332C->G; I2a2a:L36:17570599C->T; I2a2a:P219:15517851T->G; I2a2:L181:19077754G->T; I2a2:M436:18747493G->C; I2a:L460:7879415A->C; I:CTS1800:14073053G->A; I:CTS2536:14352669G->A; I:CTS4088:15389836T->C; I:CTS5946:16567253A->G; I:CTS6231:16751000C->T; I:CTS6497:16939794A->T; I:CTS7329:17424807C->T; I:CTS7593:17548890G->A; I:CTS7831:17692855T->A; I:CTS8420:18018313C->A; I:CTS8876:18257568G->A; I:CTS8963:18582617C->T; I:CTS9618:18992894T->C; I:FI2:8382265C->G; I:FI3:8485677C->A; I:L755:8465165C->T; I:L772:15615533C->A; I:PF3803:21452125A->G; I:PF3836:22525421T->G; I:Z16985:13804066G->C;
I1381	Mon	R1b1a2a1a	R1b1a2a1a:L151:16492547C->T; R1b1a2a1a:P311:18248698A->G; R1b1a2a1:L51:8502236G->A; R1b1a2:CTS623:6912992T->G; R1b1a2:CTS3575:15037433C->G; R1b1a2:CTS11468:23124367G->T; R1b1a2:L150.1:10008791C->T; R1b1a2:L265:8149348A->G; R1b1a2:L773:7220727A->G; R1b1a2:PF6430:8070532T->A; R1b1a2:PF6432:8194310C->A; R1b1a2:PF6434:8411202A->G; R1b1a2:PF6438:9464078C->T; R1b1a2:PF6482:18381735A->G; R1b1a2:PF6494:20811307G->A; R1b1a2:PF6495:20828795G->A; R1b1:M415:9170545C->A; R1b:M343:2887824C->A; R1:P225:15590342G->T; R1:P231:9989615A->G; R1:P238:7771131G->A; R1:P286:17716251C->T; R:M734:18066156C->T; R:P224:17285993C->T; R:P227:21409706G->C; R:P232:23035132G->A; R:P280:21843090C->G;

I1382	Mon	R1b1a2a1a2	R1b1a2a1a2:P312:22157311C->A; R1b1a2a1a:L151:16492547C->T; R1b1a2a1a:P310:18907236A->C; R1b1a2a1a:P311:18248698A->G; R1b1a2a1:L51:8502236G->A; R1b1a2a:L23:6753511G->A; R1b1a2:CTS623:6912992T->G; R1b1a2:CTS3575:15037433C->G; R1b1a2:CTS11468:23124367G->T; R1b1a2:L265:8149348A->G; R1b1a2:PF6430:8070532T->A; R1b1a2:PF6432:8194310C->A; R1b1a2:PF6434:8411202A->G; R1b1a2:PF6438:9464078C->T; R1b1a2:PF6475:17986687C->A; R1b1a2:PF6482:18381735A->G; R1b1a2:PF6495:20828795G->A; R1b:M343:2887824C->A; R1:M306:22750583C->A; R1:P234:21117888T->C; R1:P286:17716251C->T; R1:P294:7570822G->C; R:M734:18066156C->T; R:P224:17285993C->T; R:P280:21843090C->G; R:P285:19267344C->A;
I1389	Sie	R1b1a2a1a2	R1b1a2a1a2:P312:22157311C->A; R1b1a2:PF6425:7762947T->C;
I1390	Sie	R1b1a2a1a2	R1b1a2a1a2:P312:22157311C->A; R1b1a2a1a:P310:18907236A->C; R1b1a2a1:L51:8502236G->A; R1b1a2:CTS8665:18137831T->C; R1b1a2:L265:8149348A->G; R1b1a2:L773:7220727A->G; R1b1a2:PF6438:9464078C->T; R1b1a2:PF6475:17986687C->A; R1b1a2:PF6494:20811307G->A; R1b1a2:PF6495:20828795G->A; R1b1a2:PF6505:21784286G->A; R1b:M343:2887824C->A; R1:P238:7771131G->A; R:P280:21843090C->G;
I1388	Mar	R1b1a2a1a	R1b1a2a1a:L151:16492547C->T; R1b1a2:PF6475:17986687C->A; R1b1a2:PF6482:18381735A->G; R1b1a2:PF6494:20811307G->A; R1b1a2:PF6495:20828795G->A; R1:P225:15590342G->T; R1:P286:17716251C->T; R:M734:18066156C->T;
I2478	Par	R1b1a2a1a2c1b2b2a1	R1b1a2a1a2c1b2b2a1:F2338:16954365G->A; R1b1a2a1a2:P312:22157311C->A; R1b1a2a1a:L151:16492547C->T; R1b1a2a1a:P311:18248698A->G; R1b1a2a1:L51:8502236G->A; R1b1a2:CTS623:6912992T->G; R1b1a2:CTS11468:23124367G->T; R1b1a2:L265:8149348A->G; R1b1a2:PF6425:7762947T->C; R1b1a2:PF6434:8411202A->G; R1b1a2:PF6475:17986687C->A; R1b1a2:PF6495:20828795G->A; R1b1a2:PF6509:22190371A->G; R1b1:L278:18914441C->T; R1b:M343:2887824C->A; R1:P233:21166358T->G; R1:P286:17716251C->T; R:M734:18066156C->T;

I1767	Ing	I2a2a1a1a	I2a2a1a1a:L1195:18865320G->A; I2a2a1:CTS9183:18732197A->G; I2a2a:L368:6931594C->T; I2a2a:P221:8353707C->A; I2a2:L35:22725379C->A; I2a2:L37:17516123T->C; I2a2:P217:7628484C->T; I2a2:P218:17493630T->G; I2:L68:18700150C->T; I:CTS88:2723755G->A; I:CTS646:6926038T->A; I:CTS2193:14214481G->T; I:CTS2387:14286853T->C; I:CTS2514:14337364T->C; I:CTS2536:14352669G->A; I:CTS3384:14884659A->C; I:CTS4088:15389836T->C; I:CTS4209:15479899T->A; I:CTS4848:15862842C->T; I:CTS5946:16567253A->G; I:CTS6231:16751000C->T; I:CTS6265:16780748C->G; I:CTS7502:17511797A->G; I:CTS7593:17548890G->A; I:CTS7831:17692855T->A; I:CTS8345:17949402C->G; I:CTS8876:18257568G->A; I:CTS8963:18582617C->T; I:CTS9860:19104986G->A; I:CTS10058:19233673A->G; I:CTS10941:22845794A->G; I:CTS11540:23156725C->T; I:FGC2412:21689728A->G; I:FGC2413:8262092C->T; I:FGC2415:13835003T->C; I:FGC2416:7642823G->T; I:FI2:8382265C->G; I:FI3:8485677C->A; I:L41:19048602G->A; I:L503:21359407C->G; I:L578:8267857G->A; I:L758:8536868C->G; I:L846:7856500C->T; I:L1197:14974451C->T; I:PF3627.2:6662712C->T; I:PF3640:7681156T->A; I:PF3641:7688470T->C; I:PF3660:8466652G->A; I:PF3661:8484606C->A; I:PF3797:21130059A->G; I:PF3800:21402723A->G; I:PF3803:21452125A->G; I:PF3814:21839183A->G; I:PF3817:21939618G->A; I:PF3836:22525421T->G;
I2445	Yar	R1b1a2a1a2c1	R1b1a2a1a2c1:CTS8221:17885577C->T; R1b1a2a1a2:P312:22157311C->A; R1b1a2a1a:L151:16492547C->T; R1b1a2a1a:P310:18907236A->C; R1b1a2a1a:P311:18248698A->G; R1b1a2a1:L51:8502236G->A; R1b1a2:CTS623:6912992T->G; R1b1a2:CTS3575:15037433C->G; R1b1a2:CTS8665:18137831T->C; R1b1a2:L150.1:10008791C->T; R1b1a2:L265:8149348A->G; R1b1a2:L773:7220727A->G; R1b1a2:PF6430:8070532T->A; R1b1a2:PF6432:8194310C->A; R1b1a2:PF6434:8411202A->G; R1b1a2:PF6438:9464078C->T; R1b1a2:PF6475:17986687C->A; R1b1a2:PF6482:18381735A->G; R1b1a2:PF6495:20828795G->A; R1b1a2:PF6497:21222868C->G; R1b:M343:2887824C->A; R1:P225:15590342G->T; R1:P231:9989615A->G; R1:P233:21166358T->G; R1:P238:7771131G->A; R1:P245:8633545T->C; R1:P286:17716251C->T; R:M734:18066156C->T; R:P227:21409706G->C; R:P280:21843090C->G; R:P285:19267344C->A;
I2453	Wes	R1b1a2a1a2c1	R1b1a2a1a2c1:CTS8221:17885577C->T; R1b1a2a1a:L52:14641193C->T; R1b1a2a1a:P310:18907236A->C; R1b1a2a1a:P311:18248698A->G; R1b1a2a1:L51:8502236G->A; R1b1a2:L265:8149348A->G; R1b1a2:PF6430:8070532T->A; R1b1a2:PF6434:8411202A->G; R1b1a2:PF6497:21222868C->G; R1b1:M415:9170545C->A; R1b:M343:2887824C->A; R1:M306:22750583C->A; R1:P238:7771131G->A; R:M734:18066156C->T; R:P227:21409706G->C; R:P285:19267344C->A;

I2416	Ame	R1b1a2a1a	R1b1a2a1a:P310:18907236A->C; R1b1a2:CTS623:6912992T->G; R1b1a2:CTS11468:23124367G->T; R1b1a2:PF6430:8070532T->A; R1b1a2:PF6482:18381735A->G; R1b:M343:2887824C->A; R1:P286:17716251C->T; R:P280:21843090C->G;
I2457	Ame	R1b1a2a1a2c	R1b1a2a1a2c:L21:15654428C->G; R1b1a2a1a2:P312:22157311C->A; R1b1a2a1a:L52:14641193C->T; R1b1a2a1a:L151:16492547C->T; R1b1a2a1a:P310:18907236A->C; R1b1a2a1a:P311:18248698A->G; R1b1a2a1:L51:8502236G->A; R1b1a2:CTS623:6912992T->G; R1b1a2:CTS8665:18137831T->C; R1b1a2:CTS8728:18167403C->T; R1b1a2:CTS11468:23124367G->T; R1b1a2:L150.1:10008791C->T; R1b1a2:L265:8149348A->G; R1b1a2:L773:7220727A->G; R1b1a2:PF6430:8070532T->A; R1b1a2:PF6432:8194310C->A; R1b1a2:PF6475:17986687C->A; R1b1a2:PF6482:18381735A->G; R1b1a2:PF6494:20811307G->A; R1b1a2:PF6495:20828795G->A; R1b1a2:PF6497:21222868C->G; R1b1a2:PF6500:21410840G->T; R1b:M343:2887824C->A; R1:M173:15026424A->C; R1:P231:9989615A->G; R1:P238:7771131G->A; R:M734:18066156C->T; R:P224:17285993C->T; R:P227:21409706G->C; R:P280:21843090C->G;
I2364	Bud	H2	H2:L279:6932824G->T; H2:L285:21869856C->T; H2:P96:14869743C->A; H:M2896:16919982C->T; H:M2920:17476211G->T; H:M2936:17800761T->C; H:M2942:17887908A->G; H:M2955:18182848G->T; H:M2992:19535440A->T; H:M3035:21940316C->T; H:M3052:22701876C->G; H:M3058:22951796G->A;
I2365	Bud	R1b1a2a1a2	R1b1a2a1a2:P312:22157311C->A; R1b1a2a1a:L11:17844018T->C; R1b1a2a1a:L52:14641193C->T; R1b1a2a1a:L151:16492547C->T; R1b1a2a1a:P310:18907236A->C; R1b1a2a1a:P311:18248698A->G; R1b1a2a:L23:6753511G->A; R1b1a2:CTS623:6912992T->G; R1b1a2:CTS3575:15037433C->G; R1b1a2:CTS8728:18167403C->T; R1b1a2:CTS11468:23124367G->T; R1b1a2:CTS12478:28590278G->A; R1b1a2:L265:8149348A->G; R1b1a2:L773:7220727A->G; R1b1a2:L1353:19179540G->A; R1b1a2:PF6409:4352151G->A; R1b1a2:PF6434:8411202A->G; R1b1a2:PF6438:9464078C->T; R1b1a2:PF6475:17986687C->A; R1b1a2:PF6482:18381735A->G; R1b1a2:PF6495:20828795G->A; R1b1a2:PF6497:21222868C->G; R1b1:L278:18914441C->T; R1b1:M415:9170545C->A; R1b1:P25_3:27341589G->T; R1b:M343:2887824C->A; R1:P231:9989615A->G; R1:P238:7771131G->A; R1:P294:7570822G->C; R:M734:18066156C->T; R:P224:17285993C->T; R:P227:21409706G->C; R:P280:21843090C->G;

E09538	AugU	G2a2a1a2a1	G2a2a1a2a1:FGC5671:7652689G->A; G2a2a1a2a1:Z1370.2:7811508A->G; G2a2a1a2a1:Z6219:15894131C->T; G2a2a1a2a1:Z12228:7811520T->C; G2a2a1a2a:PF3239:17317628C->T; G2a2a1a2a:PF3240:17589665C->T; G2a2a1a2a:Z6130:6785680A->G; G2a2a1a2:L91:21645555G->C; G2a2a:PF3147:7738069G->A; G2a2a:PF3150:8476569T->C; G2a2a:PF3159:14815695C->G; G2a2a:PF3165:16582411C->A; G2a2a:PF3166:16735582T->G; G2a2a:PF3185:22894488C->T; G2a:CTS6753:17090976C->T; G2a:F2274:16802506G->T; G2a:L31:14028148C->A; G2a:M3408:22109159G->C; G2a:P15:23244026C->T; G2a:PF3141:23973594T->G; G2:CTS10089:19248446G->A; G2:F1189:8427005A->G; G2:F1239:8482393C->T; G2:F1294:8545324T->A; G2:F1393:8719593G->A; G2:F3198:21401188G->T; G2:F3220:21637589G->C; G2:F3344:22697266G->A; G2:F3536:23768744C->T; G2:M3480:8327892T->A; G2:M3488:8687693T->A; G2:M3579:21147058A->G; G:CTS282:2871867A->G; G:CTS692:6955839A->G; G:CTS827:7038432C->G; G:CTS2125:14190447A->G; G:CTS2251:14235140C->T; G:CTS2357:14273557C->T; G:CTS2517:14338503C->T; G:CTS4238:15504804C->T; G:CTS4761:15802681C->T; G:CTS5317:16203361G->C; G:CTS5658:16419934T->C; G:CTS8531:18070349G->C; G:CTS9707:19030998C->A; G:CTS10026:19215139A->T; G:CTS10706:22714204G->T; G:CTS10945:22848965A->G; G:CTS11228:23023554C->A; G:CTS11331:23074190A->G; G:CTS11529:23151673T->C; G:CTS11911:23346582A->C; G:L116:14989721C->G; G:L402:15204708T->G; G:L605:18393536G->C; G:M3450:6931141C->G; G:M3464:7537950C->T; G:M3468:7744050T->C; G:M3470:7830068T->C; G:M3473:7927218C->T; G:M3479:8231862G->C; G:M3485:8563874C->T; G:M3486:8600158A->T; G:M3582:21334507G->T; G:M3595:21671839C->T; G:P257:14432928G->A; G:PF3134:15275200C->G; G:S1435:13658486C->G; G:U21:15204710A->C;
E09569	AugU	R1b1a2a1a2b1	R1b1a2a1a2b1:L2:5755550C->T; R1b1a2a1a2:P312:22157311C->A; R1b1a2a1a:L151:16492547C->T; R1b1a2a1a:P310:18907236A->C; R1b1a2a1a:P311:18248698A->G; R1b1a2a1:L51:8502236G->A; R1b1a2:CTS623:6912992T->G; R1b1a2:CTS3575:15037433C->G; R1b1a2:CTS11468:23124367G->T; R1b1a2:L150.1:10008791C->T; R1b1a2:L265:8149348A->G; R1b1a2:L773:7220727A->G; R1b1a2:L1353:19179540G->A; R1b1a2:M520:4446430T->A; R1b1a2:PF6430:8070532T->A; R1b1a2:PF6438:9464078C->T; R1b1a2:PF6475:17986687C->A; R1b1a2:PF6482:18381735A->G; R1b1a2:PF6494:20811307G->A; R1b1a2:PF6495:20828795G->A; R1b1a2:PF6497:21222868C->G; R1b1a2:PF6509:22190371A->G; R1b1a:P297:18656508G->C; R1b1:L278:18914441C->T; R1b1:M415:9170545C->A; R1b:M343:2887824C->A; R1:M306:22750583C->A; R1:P225:15590342G->T; R1:P231:9989615A->G; R1:P238:7771131G->A; R1:P245:8633545T->C; R1:P286:17716251C->T; R1:P294:7570822G->C; R:M734:18066156C->T; R:P224:17285993C->T; R:P232:23035132G->A;

E09568_d	AugH	R1b1a2a1a	R1b1a2a1a:L52:14641193C->T; R1b1a2a1a:P311:18248698A->G; R1b1a2a1:L51:8502236G->A; R1b1a2:CTS623:6912992T->G; R1b1a2:CTS8728:18167403C->T; R1b1a2:CTS11468:23124367G->T; R1b1a2:F1794:14522828G->A; R1b1a2:L150.1:10008791C->T; R1b1a2:L265:8149348A->G; R1b1a2:L500:18180446C->A; R1b1a2:L773:7220727A->G; R1b1a2:M269:22739367T->C; R1b1a2:PF6430:8070532T->A; R1b1a2:PF6432:8194310C->A; R1b1a2:PF6434:8411202A->G; R1b1a2:PF6475:17986687C->A; R1b1a2:PF6482:18381735A->G; R1b1a2:PF6494:20811307G->A; R1b1a2:PF6495:20828795G->A; R1b1a2:PF6505:21784286G->A; R1b1a:P297:18656508G->C; R1b1:L278:18914441C->T; R1b:M343:2887824C->A; R1:M306:22750583C->A; R1:P231:9989615A->G; R1:P234:21117888T->C; R1:P294:7570822G->C; R:M734:18066156C->T; R:P224:17285993C->T; R:P227:21409706G->C;
RISE560	AugA	R1b1a2	R1b1a2:L150.1:10008791C->T; R1:P234:21117888T->C; R:P224:17285993C->T;
RISE563	Ost	R1b1a2a1a2b	R1b1a2a1a2b:PF6570:15333149C->T; R1b1a2a1a2:P312:22157311C->A; R1b1a2a1a:P310:18907236A->C; R1b1a2:CTS8665:18137831T->C; R1b1a2:PF6434:8411202A->G; R1b1a2:PF6494:20811307G->A; R1:P233:21166358T->G; R:M734:18066156C->T; R:P227:21409706G->C;
RISE564	Ost	R1b1a2a1	R1b1a2a1:L51:8502236G->A; R1b1a2:CTS11468:23124367G->T;
I0805	QVII	R1b1a2	R1b1a2:PF6430:8070532T->A; R1b1a2:PF6482:18381735A->G; R1b1a2:PF6500:21410840G->T; R1b1a2:PF6509:22190371A->G; R:M207:15581983A->G; R:M734:18066156C->T;
I0806	QVII	R1b1a2a1a2	R1b1a2a1a2:P312:22157311C->A; R1b1a2:L265:8149348A->G; R1b1a2:PF6495:20828795G->A; R1:P225:15590342G->T;
I1530	Rot	R1b1a2	R1b1a2:CTS623:6912992T->G; R1:P233:21166358T->G;
RISE566	Kne	R1b1a2a1a	R1b1a2a1a:P310:18907236A->C; R1b1a2:L1353:19179540G->A; R1b1a2:PF6430:8070532T->A; R1b1a2:PF6485:18719565T->C; R1b1a2:PF6497:21222868C->G; R:P229:8050994G->C;

Supplementary Table 3. Y-chromosome haplogroups and derived positions of England_MN individuals.

ID	Y haplogroup	Derived SNPs
I0518	I2a2	I2a2:P218:17493630T->G; I2:M438:16638804A->G; I:CTS4209:15479899T->A; I:CTS7329:17424807C->T;
I0519	I2a1b	I2a1b:CTS1293:7317227G->A; I:CTS8420:18018313C->A; I:CTS10941:22845794A->G; I:PF3641:7688470T->C;
I0520	I2c or I2a2a1	I2c:CTS4039:15349855G->A; I2c:PF3801:21414141T->A; I2a2a1:CTS9183:18732197A->G; I:L755:8465165C->T;

Supplementary Table 4. Grouping scheme based on steppe ancestry and geographic location.

Sample ID	Site ID	Grouping
I0263	Cer	BB_Iberia_Cer
I1553	Cer	BB_Iberia_Cer
I0825	Cer	BB_Iberia_Cer
I0260	Cer	BB_Iberia_Cer
I0261	Cer	BB_Iberia_Cer
I0823	Cer	BB_Iberia_Cer
I0258	Cer	BB_Iberia_Cer
I0262	Cer	BB_Iberia_Cer
I0257	Cer	BB_Iberia_Cer
I0826	Cer	BB_Iberia_Cer
I0839	Alm	BB_Iberia_Alm
I0840	Alm	BB_Iberia_Alm
I0460	Arr	BB_Iberia_Arr1
I0459	Arr	BB_Iberia_Arr1
I0461	Arr	BB_Iberia_Arr2
I0462	Arr	BB_Iberia_Arr2
I1392	Heg	BB_France_Heg
I1388	Mar	BB_France_Mar
I1391	Rou	BB_France_Rou
I1390	Sie	BB_France_Sie
I1381	Mon	BB_France_Mon
I1382	Mon	BB_France_Mon
I2478	Par	BB_Italy_Par
I1767	Ing	BB_England_NYO
I1770	Sta	BB_England_NYO
I2453	Wes	BB_England_WOV
I2454	Ove	BB_England_WOV
I2455	Ove	BB_England_WOV
I2444	Yar	BB_England_YAM2
I2445	Yar	BB_England_YAM2
I2416	Ame	BB_England_YAM1
I2418	Ame	BB_England_YAM2
I2457	Ame	BB_England_YAM2
I2459	Ame	BB_England_YAM2
I2364	Bud	BB_Hungary_Bud1
I2365	Bud	BB_Hungary_Bud2
E09537_d	AugU	BB_Germany_Aug
E09538	AugU	BB_Germany_Aug

E09569	AugU	BB_Germany_Aug
E09568_d	AugH	BB_Germany_Aug
E09613_d	AugH	BB_Germany_Aug
E09614_d	AugH	BB_Germany_Aug
RISE559	AugA	BB_Germany_Aug
RISE560	AugA	BB_Germany_Aug
RISE562	Lan	BB_Germany_Lan
RISE563	Ost	BB_Germany_Ost2
RISE564	Ost	BB_Germany_Ost1
I1549	Ben	BB_Germany_SAN
I1546	Ben	BB_Germany_SAN
I0806	QVII	BB_Germany_SAN
I0805	QVII	BB_Germany_SAN
I0113	QXII	BB_Germany_SAN
I0112	QXII	BB_Germany_SAN
I0060	Rot	BB_Germany_SAN
I0111	Rot	BB_Germany_SAN
I0108	Rot	BB_Germany_SAN
RISE568	Bra	BB_Czech_CBO
RISE569	Bra	BB_Czech_CBO
RISE566	Kne	BB_Czech_CBO
RISE567	Kne	BB_Czech_CBO

Supplementary Table 5. *f*-statistics of the form f_4(*Chimp, Yamnaya_Samara*; *Corded_Ware_Germany, Bell Beaker group*).

Bell Beaker group	f_4	Z	SNP number
BB_Iberia_Arr2	-0.00298	-7.583	643322
BB_France_Mar	-0.00215	-4.477	328052
BB_France_Rou	-0.00277	-3.295	77115
BB_France_Sie	-0.00100	-2.211	391993
BB_France_Mon	-0.00147	-4.484	695542
BB_Italy_Par	-0.00288	-7.146	541681
BB_England_NYO	-0.00123	-3.718	695801
BB_England_WOV	-0.00103	-3.277	673259
BB_England_YAM1	-0.00134	-1.906	119961
BB_England_YAM2	-0.00068	-2.618	804002
BB_Hungary_Bud1	-0.00298	-7.641	664971
BB_Hungary_Bud2	-0.00085	-2.178	681679
BB_Germany_Aug	-0.00186	-6.842	797110
BB_Germany_Lan	-0.00103	-1.052	56457
BB_Germany_Ost1	-0.00405	-4.945	76912
BB_Germany_Ost2	-0.00244	-4.677	282422
BB_Germany_SAN	-0.00079	-3.676	945754
BB_Czech_CBO	-0.00226	-6.178	669946

Supplementary Table 6. Modelling Bell Beaker groups with steppe ancestry as a mixture of *Central_MN* and *Corded_Ware_Germany*.

		Mixture proportions		
	P for rank=1	**Central_MN**	**Corded_Ware Germany**	**Std. Error**
BB_Iberia_Arr2	4.72E-01	0.817	0.183	0.088
BB_France_Mar	9.44E-01	0.752	0.248	0.082
BB_France_Rou	2.31E-01	0.265	0.735	0.189
BB_France_Sie	5.18E-01	0.443	0.557	0.110
BB_France_Mon	1.27E-01	0.219	0.781	0.089
BB_Italy_Par	5.36E-02	0.632	0.368	0.099
BB_England_NYO	8.26E-01	0.444	0.556	0.067
BB_England_WOV	8.47E-01	0.209	0.791	0.082
BB_England_YAM1	9.42E-01	0.583	0.417	0.131
BB_England_YAM2	1.17E-02	0.238	0.762	0.064
BB_Hungary_Bud1	5.85E-01	0.776	0.224	0.078
BB_Hungary_Bud2	6.97E-01	0.418	0.582	0.096
BB_Germany_Aug	1.08E-01	0.487	0.513	0.072
BB_Germany_Lan	2.79E-01	0.175	0.825	0.380
BB_Germany_Ost1	2.92E-01	0.618	0.382	0.201
BB_Germany_Ost2	2.29E-01	0.328	0.672	0.117
BB_Germany_SAN	6.28E-01	0.358	0.642	0.048
BB_Czech_CBO	4.78E-02	0.396	0.604	0.100

SI 1 - Samples and archaeological context

Galeria da Cisterna (Almonda karst system, Torres Novas, Portugal)

Galeria da Cisterna (39°30'17.32"N, 8°36'55.06"W) is a fossil karst spring of the River Almonda, which now flows ~5 m below, at the base of a ~75 m high cliff. The length of this narrow, meandering passage is approximately 100 m, and its cross-section is in general less than 2×2 m. The current entrance was exposed in the 1920s by a landslide, which allowed access and a first phase of limited archeological work, carried out between years 1937 and 1942[1–3]. The loci were identified in the ongoing excavations: AMD1, AMD2 and AMD3.

Its AMD2 locus, excavated in the years 1988-89 and more recently, featured a shallow Holocene deposit containing funerary contexts with grave goods spanning the interval between the Early Neolithic and the Iron Age. The lack of internal stratigraphic differentiation of this Holocene deposit is primarily due to the repeated prehistoric and early historic human frequentation of the site, compounded by the activity of burrowing animals. A set of typical Bell Beaker V-perforated ivory buttons and a small fragment of a gold spiral suggested that a Bell Beaker component ought to exist among the human bone remains. This was eventually corroborated by direct radiocarbon dating to this period of four right first pedal phalanges[4,5]. Two were successfully analyzed for aDNA: samples AMD2-F23-90 (I0839), dated to 3847±29 BP (OxA-28859; 2206-2457 cal BC, 2 σ) and AMD2-G18-187 (I0840), dated to 3836±29 BP (OxA-28857; 2201-2456 cal BC, 2 σ).

Paris Street (Cerdanyola del Vallès, Barcelona, Spain)

During urban development works at Paris Street in Cerdanyola del Vallès (Vallès Occidental, Barcelona province) in 2003? a large amount of skeletal material and associated pottery was unearthed. Posterior excavation work uncovered a Chalcolithic hypogeum with more than 9,000 human remains as well as lithic and ceramic material, the latter from the Bell Beaker tradition[6].

The hypogeum displays several occupational phases; the oldest one presented an ash layer underlying the first inhumations that could have a ritualistic significance; charcoal from that basal layer was dated to 4,110 ± 60 years BP (UBAR-817). The first funerary phase (UE-15) shows a large number of successive inhumations (minimal number of individuals 36) that are still in anatomical connection, placed in lateral decubitus and flexed knees. Seven arrow points were retrieved from this layer. A thin, upper layer (UE-5) probably represents a re-organization of the existing funerary space, previous to the second funerary moment (UE-2) at the hypogeum; at UE-5, two Bell Beaker vessels of maritime style were retrieved. The UE-2 layer comprises less inhumations, and all of them accompanied by typical Bell Beaker vessels: three on maritime style, two of epi-maritime style and numerous additional pieces of diverse typology. Over this layer, a final one, labelled UE-3, showed two more skeletons arranged over a river bed pebbles with a Bell Beaker vessel of a regional style known as "Pyrenaic". This last moment of use of the hypogeum is dated from a bone that yielded a date of 3,870 ± 45 years BP (UBAR-860). It therefore seems that the hypogeum was used by a local Chalcolithic community for a relatively short period of time. We recovered aDNA data from 10 individuals from this site.

Arroyal I (Arroyal, Burgos, Spain)

Arroyal I is a dolmen placed on an elevation dominating the river Ubierna in Burgos. It had a circular structure with a height of 1.8 meters and 12 meters in diameter[7]. This kind of megalithic funerary monuments in the Central Meseta of Spain have a long burial tradition, starting at the Neolithic and following a subsequent re-modelling and re-utilization during the Bell Beaker period. During this second phase several Bell Beaker items have been found associated to some skeletons, including two vessels of the so-called international maritime style. Subsequent burials provided some vessels of the so-called Ciempozuelos style. We successfully analyzed 5 individuals from this site: Roy1, Roy2. Roy3, Roy4 and Roy5 (dated to 2566-2346 calBCE (MAMS-25936)). Samples Roy1 and Roy3 were genetically first degree relatives and belonged to different mitochondrial haplogroups, which points to a father-son relationship.

Hégenheim (Haut-Rhin, France)

The Hégenheim site is located at the Rhin left bank, at few kilometers from the village of Bâle. It was subjected to an emergency excavation during summer 2004, when an individual Bell Beaker burial, next to a Merovingian necropolis was uncovered[8,9]. The burial consists of an oval pit, North-South oriented, of 1,80 meters long and 1,30 meters wide. The skeleton is placed at the bottom of the pit, in lateral decubitus over the right side, with flexed knees and elbows. The anatomical connection indicates that the body is in its primary arrangement. Some evidences suggest that the grave could have been covered by perishable material such as a wood structure.

The grave goods are limited to a decorated vessel, placed in a functional position, just behind the head. It is a big beaker of S profile and flat bottomed, with 24 cm of height and 20 cm of maximum diameter. The external colour goes from reddish to brown; the vessel has been polished and is decorated both externally and internally. The latter decoration is restricted to the first two centimetres at the edge and consists of four parallel lines impressed with an S-twisted cord. The external decoration covers all the beaker except for a short, one and half centimetre band below the edge. It consists of a series of ten strips alternated with oblique impressions done with a comb, limited above and below by a line made with a cord. The decoration can be attributed to a mixed maritime style, considered to be an early stage of the Bell Beaker tradition.

The Hégenheim individual (13-Grave9, I1392) is an adult mature, genetically female and dated to 2832-2476 calBCE (MAMS-25935). The spatial orientation and the grave goods are in accordance also with a female Bell Beaker burial.

Rouffach – Rue de Pfaffenheim (Haut-Rhin, France)

The burial of Rouffach « Rue de Pfaffenheim » was discovered in 2014 during a prospection that was not subsequently pursued. It consists of a grave of a female over 30 years of age, South-North oriented. It was deposited on her back with knees flexed right and hands turned to the face. The pit does not show any clear differentiation to the surrounding sediments.

The funerary goods included a small vessel of sinuous profile with a handle, placed in a functional position behind the body, at about ten centimetres from the left shoulder.

Seventeen V-perforated bone buttons, all placed around the right hemithorax were also found.

This kind of grave goods, specially the non-decorated vessel, suggests that this burial corresponds to a late, evolved Bell Beaker phase; the radiocarbon date undertaken on the skeleton confirms this attribution: 2346-2133 calBCE (Poz-68164).

Sierentz - Les Villas d'Aurèle (Haut-Rhin, France)

Villas d'Aurèle site is located in the municipality of Sierentz, at the Rhin left bank, 14 km away from the town of Mulhouse. The site is placed on the summit of the Rhin river upper terrace. It was subjected to an emergency excavation in 2010, when the remnants of numerous structures from the Neolithic to the early Iron Age were uncovered. Four Bell Beaker burials, composing a small funerary area of 55 meters of lenght in a North West- South East axis were excavated[10,11].

The burial 68 (I1390)

This very well preserved burial had a quadrangular shape with rounded corners, measuring 2,30 meters long by 1,80 meters wide. The walls were sub-verticals and the bottom was flat. Some evidences, such as traces of lines of dark material and fragments of wood stakes, indicate that it originally had a wooden layout, probably a structure around the body.

The individual is an adult male, aged to around 30-59 years. It was lying of his left side, in a hyper-flexed position following a North West-South West axis (the head facing the North West). The body was accompanied with two decorated vessels, eight flint elements (three of them arrow points of concave base), a grooved sandstone, a stone wristguard and a fragment of a wild board tusk.

The two vessels are beakers with an S-profile, of a beige colour and decorated with geometric, horizontal lines done with a comb and with a cord. One vessel alternates bands of short horizontal and vertical lines plus bands of incised diamonds while the other alternates oblique incised bands with herringbone patterns. The style of the pottery indicates a medium Bell Beaker phase, although the arrow points seems to suggest a Oriental tradition of the European Bell Beakers.

There are two radiocarbon dates from this skeleton:

- Poz-41226 : 2489-2299 calBCE

- Poz-41227 : 2566-2524 calBCE

The burial 69 (I1389)

This burial is also well preserved, similarly to burial 68. It is oriented in a North West-South East axis. The shape of the grave is quadrangular with rounded corners, and measures 2,25 meters long by 1,70 meters wide. There are remnants that indicate a now missing wooden structure around the body.

The individual is a male with an age around 17-19 years. It was left lying at the centre of the pit, in a flexed position over the left side of the body, in a North West-South East axis (the head facing the North West). Genetic data indicate that this individual is a first degree relative of individual I1390. They share both mitochondrial and Y-chromosome haplogroups, which points to a sibling relationship. The funerary goods accompanying the individual consist in two decorated vessels, thirteen flint elements (eight of them arrow points), a grooved sandstone, a fragment of marcasite and a pendant made of bone. The two beakers are very similar to those from the burial 68, although the decorations are different. The style of the pottery indicates a medium Bell Beaker phase, although the arrow with concave base and the pendant indicates an Oriental tradition of the European Bell Beakers.

There are two radiocarbon dates from this skeleton:

- Poz-41228 : 2468-2278 calBCE
- Poz-41229 : 2481-2289 calBCE

Mondelange - PAC de la Sente (Moselle, France)

The site is located at the Moselle valley, at about 20 kilometers North of the town of Metz. It was found during a preventive excavation in 2007 that uncovered 25 burials, nine of them dated to the later Bell Beaker tradition or to the transition to the Bronze Age[12,13].

The burial 487 (I1381)

The grave has a rectangular shape with rounded corners, measuring 2 meters long and 1.2 meters wide, with a preserved depth of 40 cm and a West-East orientation. The individual is a 10-11-year-old juvenile that lies on his left side, facing West and showing upper and lower limbs hyperflexed. There are numerous funerary elements within this grave,

including Bell Beaker vessels at the East corner and a lithic tool placed between the thorax and the right elbow.

The burial 515 (I1382)

The grave has a rectangular shape, with a flat bottom and a West-East orientation. Three of the corners show a semicircular digging that probably contained posts of 25 cm of diameter. The grave measures 2,4 meters long and 1,3 meters wide and is 0,8 meters deep. The filling is made by brownish sandy silt containing small pebbles. The individual was an adult male that was lying of his left side, with the face looking to the West. The upper limbs are flexed, with the right hand over the left humerus and the left hand placed in front of the face. The lower limbs are also flexed with the knees oriented to the North.

There are numerous funerary elements within this grave. A stone wristguard is placed next to the left shoulder. Two vessels are placed close to the feet, one near the axis of the body and the other one next to the South wall. One flint arrow point was found between both vessels, at 10 centimetres over the bottom of the pit. The two beakers, one decorated and the other not -with a peculiar morphology- suggest an evolved, late Bell Beaker phase, with oriental influences.

Datation : GrA-4468 : 2456-2418 calBCE

Marlens - Sur les Barmes (Haute-Savoie, France)

The Marlens - Sur les Barmes site is located in the French Alps, South-East of the lake Annecy. It consists more in an accidental discovery than an archaeological prospection. The site is a crevice that opens at the bottom of a rocky wall, near massive fallen rock debris; the skeleton has been placed inside[14]. The entrance is a very narrow gallery that has a height of only 60 centimetres. The interior space measures 2,50 meters by 1,50 meters with an irregular height that reaches 1,70 meters at the highest point. The walls consist of large, fractured blocks. It seems that the original cavity was enlarged by removing some blocks. The skull of the individual was lying on a small cavity formed by a natural layout of small stones.

The individual (I1388) is a young male of about 24 years of age. It was associated to a fragment of a Bell Beaker vessel. The decoration is made by the combination of horizontal bands and radial elements including ladder and lattice patterns. This type of

incised-printed decoration points to regional Bell Beakers, specifically to the group from the Rhone-Provence of a recent phase.
Datation : Lyon-3099 : 2435-2421 calBCE

Via Guidorossi (Parma, Italy)

The site of Via Guidorossi at Parma, in the Po plain, was excavated in 2009 and corresponds to an advanced Bell Beaker period, dated to 2,200-1,930 years cal BCE[15]. The tomb number 1 contained two skeletons, labelled individuals A (US-8) and B (US-9). Both were placed into an excavated structure of about 2.2x2.2 meters, with an opening at the North-East corner.
Individual A was a ~30-year-old woman placed in a South-North orientation, while individual B (I2478), the best preserved of both and the one successfully analyzed in this study, was a 30-40-year-old male dated to 2200-1930 BCE (LTL-5035A). This skeleton was placed with flexed legs on his left side, with his left arm also flexed and the right one extended. He was oriented North-South, with the head pointing North and the face looking at the East. Two Bell-Beaker vessels, one decorated with incised triangles in a central band and the other undecorated, were placed at his feet, while two additional vessels were located close to the opening of the funerary structure. The decorated pottery is similar to objects found in other Bell Beaker Italian sites such as Rubiera (Reggio Emilia).
Some lithic implements, including a remarkable knife, were found between the legs of individual B. The only similar lithic knife in a Bell Beaker context has been found at Fosso Conicchio near Viterbo.
A second tomb, excavated in a sub-quadrangular form, contained three more skeletons, labelled A (US-12), B (US-13) and C (U-14), all of them placed in flexed position. Individuals A and C were 60 and 50 year-old males, respectively, while individual B, who was lying between them and in inverted orientation (North-South), was a 15-18-year-old young female. Several decorated Bell Beaker vessels were located within this second grave. The bipolar orientation of the Guidorossi burials -South-North for females and North-South for males, all facing to the East- points to traditions found in central European Bell Beaker sites, such as those from lower Austria and Moravia.

SI 2 - Data generation

All ancient DNA extracts were prepared using a previously described method[16]. Libraries were prepared using double-stranded protocols and were UDG-treated in a way that retains some characteristic ancient DNA damage at the last nucleotide [17].

We hybridized the libraries to oligonucleotide probes overlapping the mitochondrial DNA genome (mtDNA) following the method in ref. [18]. We sequenced the enriched libraries on the Illumina MiSeq or HiSeq2500 platforms using a double index configuration. We demultiplexed the reads according to the expected index pairs, allowing one mismatch for each. We merged paired-end reads into a single fragment by requiring an overlap of at least 11 bp (with one mismatch allowed), using a modified version of SeqPrep in which the base and quality score is determined by the read that has higher quality. After stripping adapters, we mapped merged fragments which we required to be at least 30bp in length to the mtDNA revised Cambridge Reference Sequence (rCRS) with BWA (v0.6.1) using the *samse* command. We identified duplicated fragments based on having the same orientation, start and end positions, and kept the highest quality fragment. We excluded fragments with mapping quality <30.

We performed in-solution enrichment for a targeted set of 1,237,207 SNPs using previously reported protocols[18–20]. The targeted SNP set merges 394,577 SNPs first reported in ref. [19] (390k capture), and 842,630 SNPs first reported in ref. [21](840k capture). We generated 2×75bp reads on Illumina HiSeq2500 or NextSeq500 instruments. We processed the fragments as for mtDNA, except we aligned to the human reference genome, *hg19,* and required an overlap of at least 15bp. We mapped with the command bwa –n 0.01 and –l 16500.

SI 3 - Data curation

We required all analysed libraries to have a damage profile consistent with ancient DNA Specifically, we restricted analysis to libraries that had a rate of ≥3% cytosine-to-thymine substitutions in the terminal nucleotide[17].

We used mtDNA data to flag samples as possibly contaminated. We declared a library possibly contaminated if, after running the contamination estimator ContamMix[18], either of two criteria were met: (i) The fraction of fragments matching the reconstructed

consensus better than any of 311 worldwide mtDNA sequences used for comparison is <95%. (ii) The ninety-five percent confident lower bound of the fraction of fragments matching the consensus better than any of 311 worldwide mtDNA sequences is <85%.
Males have only one X chromosome and thus are not expected to be polymorphic in this part of their genome. This can be used to obtain a conservative estimate of contamination in males given sufficient X chromosome coverage[22]. We used the ANGSD software to run this test on all males where it gave good resolution (we only used X chromosome estimates for males for whom at least 200 SNPs covered at least twice). We considered libraries as effectively uncontaminated if their X chromosome contamination estimates were less than 2.5%.
Libraries failing at one of these steps were either not considered for analysis or restricted to fragments carrying characteristic ancient DNA damage at their terminal ends, which has been shown to considerably reduce contamination[23].
Finally, only samples with more than 30,000 SNPs hit at least once were included in the final dataset (Supplementary Data Table 1).

SI 4 - Mitochondrial haplogroup determination

We used the mitochondrial capture bam files to determine the mitochondrial haplogroup of each newly reported sample, restricting to reads with MAPQ>=30 and base quality >=30. First, a consensus sequence was constructed with samtools and bcftools[24], using a majority rule and requiring a minimum coverage of 2. Haplogroups were called with HaploGrep tool[25] based on phylotree[26] (mtDNA tree Build 17 (18 Feb 2016)). Mutations against the rCRS and corresponding haplogroups can be viewed in Supplementary Table 1. We find heteroplasmy in five samples: I0825, I0257, I0461, I1381 and I2455. In all the cases, haplogroup calls were unambiguous, making contamination unlikely.

SI 5 - Y-chromosome analysis

In order to study the Y-chromosome haplogroup composition of our Bell Beaker male individuals (both published and new samples), we made use of the sequences mapping to 1240k Y-chromosome targets, restricting to reads with mapping quality >=30 and bases

with base quality >=30. We called haplogroups by determining the most derived mutation for each sample, using the nomenclature of the International Society of Genetic Genealogy (http://www.isogg.org) version 10.70. Haplogroups and corresponding derived mutations can be viewed in Supplementary Table 2. We further comment on specific samples below:

-Sample I0257 could be assigned to haplogroup R based on one derived allele at SNP P224:17285993C->T. However, we caution that the derived allele could be a consequence of post-mortem deamination. Haplogroups I, G and H can be excluded due to the presence of ancestral alleles for I (CTS2193:14214481G->T, PF3641:7688470 T->C, PF3660:8466652G->A), G (CTS1283:7309873T->G, CTS2016.1:14155765G->A, CTS2125:14190447A->G, CTS4761:15802681C->T, CTS9011:18615020A->T, M3474:7930724C->A, PF3134:15275200G->C) and H (M2942:17887908A->G).

-Sample I0261 could belong to haplogroup P* based on mutation F91:7628900G->A, with upstream haplogroup IJK also supported by mutation L15:6753519A->G. Again, derived allele at mutation F91 might represent ancient DNA damage. Haplogroups I, G, R1a and R1b1a2a can be excluded due to the presence of ancestral alleles for I (CTS11979:23401471C->T), G (M3600:21954611G->A, PF3134:15275200C->G), R1a (L145:14138745C->A) and R1b1a2a (L23:6753511G->A).

-Sample I2478 belonged to haplogroup R1b1a2a1a2c1b2b2a1 based on mutation F2338:16954365G->A, which could correspond to ancient DNA damage. Nevertheless, upstream haplogroup R1b1a2a1a2 (P312:22157311C->A) is also supported.

Summarizing the results from our BBC set, we find a clear pattern in the Y-chromosome distribution that is highly correlated with steppe ancestry in the nuclear genome. Iberian individuals with enough data to obtain a reliable call belonged to haplogroups I2a2 and G2, both present in high frequencies in European Neolithic farmers[19,27–29] and also in the Chalcolithic population from El Mirador site in Spain[30]. Haplogroup G2 probably entered Europe from the Middle East during the Neolithic expansion and haplogroup I2a2 was likely introduced into the Neolithic population through admixture with European hunter-gatherers. Outside the Iberian Peninsula we observe a striking uniformity in the paternal lineages, with 19 of the 22 individuals belonging to haplogroup R1b1a2. In the cases

where R1b1a2 downstream mutations could be determined, all of them were positive for R-M412 mutation (n=16) and for R-S116 mutation (n=10). Haplogroup R-S116 is common in Western Europe, with frequency peaks in Ireland and the Franco-Cantabrian region[31]. Its presence in Central and Northern European individuals suggests a role for Bell Beakers in the dissemination of this haplogroup throughout Western Europe. The three individuals outside Iberia without R1b1a2 Y-chromosomes belonged to haplogroups H2, G2a and I2a2 (Supplementary Table 2), which were frequent in Europe before the arrival of steppe migrants. Interestingly, those individuals present low amounts of steppe ancestry in the nuclear genome compare to other individuals from the same region (Fig. 3a).

Y-chromosome analysis of England_MN individuals

Using the same approach, we study the Y-chromosome of the three newly reported male individuals from Middle Neolithic England (Supplementary Table 3). Sample I0518 belongs to haplogroup I2a2 based on mutation P218:17493630T->G, with haplogroups I2 and I also supported. Sample I0519 could be assigned to haplogroup I2a1b based on mutation CTS1293:7317227G->A, with haplogroup I also supported by three mutations. However, mutation CTS1293 could be a consequence of post-mortem deamination. Sample I0520 could belong to haplogroup I2c based on mutations CTS4039:15349855G->A and PF3801:21414141T->A or to haplogroup I2a2a1 based on mutation CTS9183:18732197A->G. Haplogroup I is also supported by mutation L755:8465165C->T.

SI 6 - Reference datasets for population genetics analyses

We assembled two reference datasets for population genetics analyses:

- *HO* includes 2,572 present-day individuals from worldwide populations genotyped on the Human Origins Array[32], of which 2,334 were analysed by Lazaridis et al.[20] and 238 firstly described in Lazaridis et al.[33], and 246 ancient individuals. The ancient set includes 60 Bell Beakers (42 newly reported)(Supplementary Data Table 1) and 186 samples from relevant ancient populations (Supplementary Data Table 2), of which 173 have been

previously published[20,27,30,34–41], 10 previously published[30] whose coverage we increase and 3 newly reported in this study. We kept 591,642 autosomal SNPs which resulted from intersecting autosomal SNPs in the 1240k capture with the analysis set of 594,924 SNPs in Lazaridis et al.[20].

-*HOIll* includes the same 246 ancient samples and 31 present-day individuals from 6 populations (Han, Chukchi, Karitiana, Mbuti, Onge and Papuan) sequenced to high coverage as part of the Simons Genome Diversity Project. For this dataset, 1,054,671 autosomal SNPs were used, excluding SNPs of the 1240k array located on sex chromosomes or with functional effects.

For both datasets, ancient individuals were merged by randomly sampling one read at each SNP position.

SI 7 - Principal component analysis and ADMIXTURE analysis

Principal component analysis

Principal component analysis (PCA) was carried out on the *HO* dataset using the *smartpca* program in EIGENSOFT[42]. We included 990 present-day West Eurasians and 243 ancient individuals, of which 60 are Bell Beaker individuals. Many of the previously published ancient samples were not subjected to UDG treatment and consequently had high rates of C to T and G to A deamination at read ends, which could potentially affect their position in the PCA. Thus, in order to avoid the effect of different deamination rates in the ancient individuals, PCA was generated using 110,016 transversion positions (Fig. 2a). Fig. 2a reveals a marked genetic heterogeneity among Bell Beaker individuals, apparent at both continental and regional scales, and underlined by different amounts of steppe-related ancestry. Furthermore, even within the same site we find individuals with very different genomic profiles, e.g. Arroyal I and Budapest-Békásmegyer sites.

In Supplementary Fig. 1 we provide a PCA computed on all 591,642 SNPs with the ancient individuals projected using the lsqproject: YES option in *smartpca*.

Admixture

We performed model-based clustering analysis using ADMIXTURE[43] on the *HO* reference dataset, including 2,572 present-day individuals from worldwide populations, 60 Bell Beaker individuals and 137 relevant ancient individuals. First, we carried out LD-pruning on the dataset using PLINK[44] with the flag --indep-pairwise 200 25 0.4, keeping 298,594 SNPs that were used for analysis. ADMIXTURE was run with the cross validation (--cv) flag considering K=2 to K=20, with 20 replicates for each value of K and keeping for each value of K the replicate with highest log likelihood. In Fig. 2b we show the ancestry proportions for K=11 of Bell Beaker individuals and other relevant ancient samples for comparison. This value of K was the lowest for which components of CHG and European hunter gatherers are maximized.

Both PCA and ADMIXTURE provide a similar view of the genomic affinities of the Bell Beaker horizon. A first group of samples, including Iberian individuals and one individual from Eastern France, cluster with European Middle Neolithic and Chalcolithic populations lacking steppe ancestry (Fig. 2a). In ADMIXTURE, they harbor two ancestral components; a blue component associated with the Anatolian and European Neolithic, and a red component related to the European pre-Neolithic hunter-gatherers (Fig. 2b). A second group of Bell Beakers, composed by Central European and English individuals, plot near German Corded Ware samples in the PCA (Fig. 2a) and present a similar ancestry composition in ADMIXTURE (Fig. 2b), with a higher proportion of blue and lower proportion of the green component, which was introduced into Europe by steppe migrations (i.e. Yamnaya). A third group corresponds to samples plotting in intermediate positions in the PCA, reflecting variable amounts of steppe related ancestry.

SI 8 - *f*-statistics

We studied *f*-statistics using ADMIXTOOLS[32] with default parameters. Standard errors were computed using a weighted block jackknife approach[45] over 5 Mb blocks.

f_4-statistics

First, in order to study the degree of steppe ancestry in our dataset, we computed on the *HOIll* dataset statistics of the form f_4(*Chimp, Test*; *Yamnaya_Samara*,

Anatolia_Neolithic) for each Bell Beaker individual (Fig. 3a). This statistic takes minimum values with increasing affinity to Yamnaya steppe herders. Two ancient populations were included in the analysis; one representing a Late Neolithic population with no evidence of steppe ancestry, e.g. *Iberia_Chalcolithic*, and one representing the Late Neolithic population from Europe with highest affinity to Yamnaya, e.g *Corded_Ware_Germany*. In Fig. 3a we see individuals with variable affinity to Yamnaya population. Most of the samples show a clear signal of steppe ancestry, though not as strong as in the Corded Ware. Samples from the Iberian Peninsula and one individual from Hégenheim in Eastern France present values similar to that of *Iberia_Chalcolithic* population, indicating the lack of steppe ancestry.

In order to increase statistical power to detect allele frequency differences, we sought to create groups of individuals excavated in nearby sites presenting similar population affinities. Specifically, we grouped individuals (Supplementary Table 4) if they fulfilled the following conditions:

- They were excavated in sites separated by less than 100 kilometers.
- They did not significantly differ with regard to statistic f_4(*Chimp*, *Test*; *Yamnaya_Samara*, *Anatolia_Neolithic*), using 1.5 standard errors.

We then recomputed the statistic f_4(*Chimp*, *Test*; *Yamnaya_Samara*, *Anatolia_Neolithic*) with the new grouping scheme in Supplementary Table 4 (Supplementary Fig. 2). We used this steppe ancestry-based grouping scheme to test for symmetry between pairs of populations.

Presence of steppe ancestry in two samples from Arroyal I in Burgos (Northern Spain)

We show in Supplementary Fig. 3 f_4-statistics of the form f_4(*Chimp*, *Test*; *Iberia_Chalcolithic*, *BB_Iberia_Arr2*). Several populations are asymmetrically related to *Iberia_Chalcolithic* and *BB_Iberia_Arr2* (including samples I0461 and I0462). Interestingly, two populations with steppe ancestry, *EHG* (Z=3.6) and *Yamnaya_Kalmykia.SG* (Z=3.1) are significantly closer to *BB_Iberia_Arr2* than to *Iberia_Chalcolithic*. In Supplementary Fig. 4 we find *BB_Iberia_Arr1* (including samples I0459 and I0460) and *Iberia_Chalcolithic* to be symmetrically related to ancient West Eurasian populations. This confirms the visual impression from the PCA (Fig 2a and Supplementary Fig. 1) and supports the separation of Arroyal I samples in two groups, Arr2 with steppe ancestry and Arr1 without it. To our knowledge, Arr2 samples represent the earliest observation of steppe affinities in the Iberian Peninsula.

Ancestry heterogeneity in Haut-Rhin (France)

In Supplementary Fig. 5 we show that *Iberia_Chalcolithic* and *BB_France_Heg* are symmetrically related to ancient West Eurasians, indicating the lack of steppe ancestry in the sample from Hégenheim, and consistent with PCA and ADMIXTURE results. Next, we test symmetry between *BB_France_Heg* and *BB_France_Sie* (Supplementary Fig. 6), which consists of one individual excavated in Sierentz, 12 kilometers apart from Hégenheim. Both *EHG* (Z=3.1) and *Yamnaya_Samara* (Z=3.6) share more alleles with *BB_France_Sie* than with *BB_France_Heg*, documenting very different population affinities in samples excavated from nearby sites.

Population discontinuity in Northern Italy

Our Bell Beaker sample from Parma is slightly shifted towards populations with steppe ancestry in the PCA (Fig 2a and Supplementary Fig. 1), an observation further supported by statistic f_4(*Chimp*, *BB_Italy_Par*; *Yamnaya_Samara*, *Anatolia_Neolithic*) (Fig. 3a). We tested for symmetry between *BB_Italy_Par* and *Remedello_CA*[37], a culture preceding Bell Beakers in Northern Italy. Several populations from the Early Neolithic (*Iberia_EN.SG*, *Iberia_EN*, *LBK_EN*, *Anatolia_Neolithic* and *Hungary_EN*) and Middle/Late Neolithic (*Iberia_Chalcolithic.SG*, *Iberia_Chalcolithic*, *England_MN*, *Ireland_MN.SG*, *Hungary_CA*, *Iberia_MN* and *Central MN*) share more alleles with *Remedello_CA* than with *BB_Italy_Par* (Supplementary Fig. 7), indicating that our Italian Bell Beaker harbors an ancestry component not present in the previous Remedello culture.

Ancestry heterogeneity at Budapest-Békásmegyer (Hungary)

The Carpathian Basin represents the easternmost limit of Bell Beaker distribution. Our dataset includes two samples from Budapest-Békásmegyer site in Hungary that behave differently in PCA and ADMIXTURE (Fig 2a and Fig 2b). Testing for symmetry between *BB_Hungary_Bud1* and *BB_Hungary_Bud2* (Supplementary Fig. 8) supports a different genomic profile for these two individuals, with *EHG* (Z=4.5), *Samara_Eneolithic* (Z=3.3) and *Yamnaya_Samara* (Z=3.5) steppe populations sharing more alleles with *BB_Hungary_Bud2* than with *BB_Hungary_Bud1*. Similar to Arroyal I in Northern Spain, in Budapest-Békásmegyer site we find individuals with different degrees of steppe-related ancestry.

Outgroup f_3-statistics

We studied the shared genetic drift between Bell Beaker groups and other ancient and modern populations by computing statistics of the form f_3(Mbuti; Bell Beaker group, X). We used the *HO* dataset when analyzing shared genetic drift present-day West Eurasians (Supplementary Fig. 9-14), and the *HOIll* dataset when analyzing shared genetic drift with ancient populations (Supplementary Fig. 17-22). We also explored the shared genetic drift pattern of relevant ancient populations: Corded Ware, Chalcolithic Iberia and Central European Middle Neolithic (Supplementary Fig. 15-16 and Supplementary Fig. 23-24).

Several interesting patterns can be observed. Chalcolithic Iberians and Iberian Bell Beaker groups: *BB_Iberia_Cer*, *BB_Iberia_Alm* and *BB_Iberia_Arr1* present a very similar pattern of shared genetic drift with present-day populations (Supplementary Fig. 9a-c and Supplementary Fig. 15a-b). Furthermore, for those Bell Beaker groups *Iberia_Chalcolithic* always appears among the three ancient populations with highest shared genetic drift (Supplementary Fig. 17a-c), confirming the close affinity between these Iberian populations. The remaining Iberian Bell Beaker group, *BB_Iberia_Arr2*, shares more genetic drift with Central and Northern European populations (Supplementary Fig. 9d), consistent with the presence of steppe ancestry already detected in previous analyses (Fig. 2a and Fig. 3a). Regarding Eastern France groups, *BB_France_Heg* and *BB_France_Mar* show a high shared genetic drift with Sardinians and Basques (Supplementary Fig. 10a-b), similar to other Neolithic Europeans, with *BB_France_Mar* sharing more drift with Northern Europeans than *BB_France_Heg* does. The other French groups, *BB_France_Rou*, *BB_France_Sie* and *BB_France_Mon*, present strong affinity with Northern Europeans (Supplementary Fig. 10c-d and Supplementary Fig. 11a) and with Corded Ware populations (Supplementary Fig. 18c-d and Supplementary Fig. 19a), indicating a strong steppe influence. Such a strong steppe influence is also evident in English, German and Czech Bell Beaker groups (Supplementary Fig. 12-14 and Supplementary Fig. 20-22), with *BB_England_Yam1* and *BB_Germany_Ost1* sharing more drift with Sardinians and Basques (Supplementary Fig. 12c and Supplementary Fig. 13c). The two samples from Budapest-Békásmegyer site in Hungary (*BB_Hungary_Bud1* and *BB_Hungary_Bud2*) present different affinities with modern and ancient populations (Supplementary Fig. 11c-d and Supplementary Fig. 19c-d), replicating previous results (Fig. 2 and Fig. 3a). The pattern observed for

BB_Italy_Par (Supplementary Fig. 11b and Supplementary Fig. 19b) is consistent with predominantly farmer-related ancestry but with some degree of steppe influence.

SI 9 - qpWave a qpAdm analyses

In this section we aim to fit the ancestry of our Bell Beaker individuals into genetic models with estimated mixture coefficients. We use the knowledge gained in previous sections to investigate different models under the framework described in Haak et al[19] and implemented in the *qpWave* and *qpAdm* programs (https://github.com/DReichLab). These methods relate a *Test* population to a set of *Outgroup* populations via a set of *Reference* populations. If the *References* share different amounts of genetic drift with the *Outgroups*, mixture coefficients can be estimated without having to explicitly model the relationship between the *Outgroups* and the *References*.

We carried out the analysis on the HOIll dataset and used a set of 10 *Outgroups*: Han, Chukchi, Karitiana, Mbuti, Onge, Papuan, Ust_Ishim, Kostenki14, MA1, CHG. These populations are located in informative places of the phylogeny and are unlikely to have contributed directly to our *Test* populations, i.e. Bell Beakers. For the present-day populations included in the *Outgroup* set, we used individuals sequenced as part of the Simons Genome Diversity Project. For the Bell Beaker individuals, we used the steppe ancestry-based grouping scheme (Supplementary Table 4) described in SI8.

We first attempted to model the ancestry of Bell Beaker groups lacking steppe affinities in PCA (Fig. 2a), ADMIXTURE (Fig. 2b) and f_4-statistics (Fig. 3a), e.g. *BB_Iberia_Cer*, *BB_Iberia_Alm*, *BB_Iberia_Arr1* and *BB_France_Heg*. Previous work[30] has successfully modelled European Neolithic populations before the arrival of Eastern migrants as a mixture of two genetic components: *Anatolia_Neolithic* and *WHG*. Thus, using *qpwave* we tested if the triple (*Anatolia_Neolithic*, *WHG*, Bell Beaker group) could be related via 2 streams of ancestry to the *outgroup* set. Rank =1 is accepted for our four populations of interest (Table 1), which means they can be explained as a mixture between *Anatolia_Neolithic* and *WHG*. We then estimated mixture coefficients with *qpAdm* (Table 1). Consistent with previous results[30], Iberian populations present a higher proportion of the *WHG* component than our central European representative, e.g. *BB_France_Heg*.

Next, we studied the ancestry of the Bell Beaker groups with evidence of steppe affinities in previous analyses. Those analyses show that many Bell Beaker groups are genetically close to Corded Ware people (Fig. 2a and Fig. 3a), with less steppe ancestry as shown by negative f_4 (*Chimp*, *Yamnaya_Samara*; *Corded_Ware_Germany*, Bell Beaker group) (Supplementary Table 5) and an excess or farmer-related ancestry not present in the Corded Ware. In light of these observations, we tested the model *Corded_Ware_Germany+Central_MN*. We chose *Central_MN* as a representative of a European Neolithic population without steppe ancestry. Rank=1 is accepted for all the Bell Beaker groups but *BB_England_YAM2* and *BB_Czech_CBO*, which is marginally rejected (p= 0.0478) (Supplementary Table 6). In Fig. 3b we plot the geographic distribution of estimated Corded Ware-related ancestry in Bell Beaker groups, excluding *BB_Germany_Lan* and *BB_Germany_Ost1* due to large standard errors.

SI 10 - England_MN population affinities

In this note we describe the ancestry of the three Middle Neolithic samples from Northampton (England) presented in this study and labelled as *England_MN* (Supplementary Data Table 2).

These individuals cluster with other Middle Neolithic Europeans in the PCA (Figure 5c) and present similar ancestry components and proportions in ADMIXTURE analysis (Supplementary Fig. 25). The pattern of shared genetic drift with other ancient and modern populations (Supplementary Fig. 15d and Supplementary Fig. 23d) further supports a strong similarity with contemporaneous populations in continental Europe. Following the same methodology as in SI9, we modelled the ancestry of *England_MN* as a mixture of *Anatolia_Neolithic* and WHG and obtain an estimate of 83.4±6.5% *Anatolia_Neolithic* and 16.6 ±6.5% WHG (P value for Rank=1 is 0.33). These results indicate a lack of steppe ancestry in *England_MN*, which contrasts with the widespread steppe influence detected in subsequent populations from the British Isles.

References

1. Zilhão, J. *O Paleolítico Superior da Estremadura portuguesa*. (Colibri, 1997).
2. Zilhão, J. in *Europe's First Farmers* 144–182 (Cambridge University Press, 2000). doi:10.1017/CBO9780511607851.007
3. Zilhão, J. The Early Neolithic artifact assemblage from the Galeria da Cisterna (Almonda karstic system, Torres Novas, Portugal). in *De Méditerranée et d'ailleurs. Mélanges offerts à Jean Guilaine* 821–835 (Archives d'Écologie Préhistorique, 2009).
4. Martins, H. *et al.* Radiocarbon dating the beginning of the Neolithic in Iberia: new results, new problems. *J. Mediterr. Archaeol.* **28,** 105–131 (2015).
5. Zilhão, J. Beaker People without Beaker Pots: the Chalcolithic Funerary Context from the Galeria da Cisterna (Almonda Karst System, Torres Novas, Portugal). *Serie de Trabajos Varios del SIP, Valencia* **In press,** (2015).
6. Francés-Farré, J., Majó-Ortín, T., Sala-Navas, Ò., Guàrdia i LLorens, M. & Hernández-Díaz, J. L'hipogeu calcolític del carrer de París (Cerdanyola del Vallès). *Cota Zero* **19,** 7–9 (2004).
7. Carmona-Ballestero, E., Arnaiz-Alonso, M. Á. & Alameda, M. del C. El dolmen de Arroyal I : usos y modificaciones durante el III milenio cal AC . in *Actas de las segundas jornadas de jóvenes investigadores del valle del Duero 2012* 41–54 (2012).
8. Billoin, D. Le Néolithique final : la sépulture campaniforme d'Hégenheim. in *10000 ans d'histoire ! Dix ans de fouilles archéologiques en Alsace* 31–42 (Musées de la ville de Strasbourg, 2009).
9. Billoin, D., Jeunesse, C., Denaire, A. & Thiol, S. Une nouvelle

sépulture campaniforme à Hégenheim (F-Haut-Rhin). in *Du Néolithique final au Bronze ancien dans le Nord-Est de la France. Actualité de la recherche* 31–42 (APRAA, 2010).

10. Vergnaud, L. The Bell Beaker funeral group from Sierentz « les Villas d'Aurèle » (Haut-Rhin, France). in *Current researches on Bell Beakers. Proceedings of the 15th International Bell Beaker Conference: From Atlantic to Ural (5th-9th May 2011, Poio, Spain)* 51–59 (Galician ArcheoPots, 2013).
11. Vergnaud, L. Les sépultures campaniformes de Sierentz « Les Villas d'Aurèle » (Haut-Rhin). in *Données récentes sur les pratiques funéraires néolithiques de la plaine du Rhin supérieur* 173–210 (Archaeopress, 2014).
12. Lefebvre, A., Gazenbeek, M. & Pernot, P. Les sépultures campaniformes du site de Mondelange « La Sente » (Moselle). Résultats préliminaires. *Internéo* **7,** 187–201 (2008).
13. Lefebvre, A. Les sépultures du Néolithique final / bronze ancien en Lorraine: vers l'émergence de nouvelles problématiques. in *Du Néolithique final au Bronze ancien dans le Nord-Est de la France. Actualité de la recherche* 103–118 (APRAA, 2010).
14. Serralongue, J. & Rey, P.-J. La sépulture campaniforme de Marlens, Sur les Barmes. *La Rev. Savoisienne* **145e année,** 41–46 (2005).
15. Bernabò-Brea, M. & Mazzieri, P. Nuovi dati sul campaniforme in Emilia. in *L'Età del Rame. La pianura padana e le Alpi al tempo di Ötzi* 503–524 (Compagnia della Stampa, 2013).
16. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 15758–63 (2013).
17. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil – DNA – glycosylase treatment for screening of ancient

DNA. *Philos. Trans. R. Soc. London B* (2015). doi:10.1098/rstb.2013.0624

18. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 2223–7 (2013).
19. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).
20. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
21. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* (2015). doi:10.1038/nature14558
22. Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334,** 94–8 (2011).
23. Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci.* 1–6 (2014). doi:10.1073/pnas.1318934111
24. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–9 (2009).
25. Kloss-Brandstätter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32,** 25–32 (2011).
26. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30,** E386–94 (2009).
27. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3,** 698 (2012).
28. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5,** 5257 (2014).
29. Skoglund, P. *et al.* Genomic Diversity and Admixture Differs for

Stone-Age Scandinavian Foragers and Farmers. *Science* **201,** 786–792 (2014).

30. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528,** 499–503 (2015).
31. Valverde, L. *et al.* New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia. *Eur. J. Hum. Genet.* 1–5 (2015). doi:10.1038/ejhg.2015.114
32. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–93 (2012).
33. Lazaridis, I. The genetic structure of the world's first farmers. *Submitted* (2016).
34. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
35. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346,** 1113–1118 (2014).
36. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).
37. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522,** 167–172 (2015).
38. Olalde, I. *et al.* A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Mol. Biol. Evol.* **32,** 3132–3142 (2015).
39. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl. Acad. Sci.* 201509851 (2015). doi:10.1073/pnas.1509851112
40. Jones, E. R. *et al.* Upper palaeolithic genomes reveal deep roots of modern eurasians. *Nat. Comm.* 1–8 (2015). doi:10.1038/ncomms9912
41. Cassidy, L. M. *et al.* Neolithic and Bronze Age migration to Ireland

and establishment of the insular Atlantic genome. *Proc. Natl. Acad. Sci.* **113,** 1–6 (2016).

42. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
43. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
44. Purcell, S. *et al.* PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
45. Busing, F. M. T. A., Meijer, E. & Van Der Leeden, R. Delete- m Jackknife for Unequal m. *Stat. Comput.* **9,** 3–8 (1999).

SupplementaryDataTable1

ID	Colloquial ID	Sex	mtDNA haplogroup	Y haplogroup	SNPs in autosomes	Data Source	Publication	Date(95% CI)
I0263	10385A	F	X2b+226	-	31569	1240k capture	This study	2800-2000 BCE
I1553	10388A	F	pre-H103	-	37205	1240k capture	This study	2800-2000 BCE
I0825	10394A	M	K1a4a1	G2	42018	1240k capture	This study	2474-2300 calBCE (MAMS-25939)
I0260	10370A	F	K1a2a	-	51816	1240k capture	This study	2800-2000 BCE
I0261	10378A	M	U5b1i	P?	52264	1240k capture	This study	2800-2000 BCE
I0823	10360A	F	H1	-	60463	1240k capture	This study	2800-2000 BCE
I0258	10367A	F	H1q	-	62877	1240k capture	This study	2800-2000 BCE
I0262	10381A	F	U5b3	-	92174	1240k capture	This study	2800-2000 BCE
I0257	10362A	M	H1ax	R?	107954	1240k capture	This study	2571-2350 calBCE (MAMS-25937)
I0826	10400A	M	H1t	I2a2	112957	1240k capture	This study	2833-2480 calBCE (MAMS-25940)
I0839	AMD2-F23-90	F	-	-	37761	1240k capture	This study	2206-2457 calBCE (OxA-28859)
I0840	AMD2-G18-187	F	U5b1i	-	133290	1240k capture	This study	2201-2456 BCE (OxA-28857)
I0460	Roy3	M	H45	I2a2a1	344361	1240k capture	This study	3000-2000 BCE
I0458	Roy1	M	K1a1b1	I2a2a	131125	1240k capture	This study	3000-2000 BCE
I0459	Roy2	F	U5b2b	-	527866	1240k capture	This study	3000-2000 BCE
I0461	Roy4	F	K1a1b1	-	720612	1240k capture	This study	3000-2000 BCE
I0462	Roy5	F	K1a+195	-	32916	1240k capture	This study	2566-2346 calBCE (MAMS-25936)
I1392	13-Grave9	F	H1+152	-	283190	1240k capture	This study	2832-2476 calBCE (MAMS-25935)
I1388	9-Grave1	M	H	R1b1a2a1a	375748	1240k capture	This study	2435-2421 calBCE (Lyon-3099)
I1391	12-Grave	F	J1c4	-	87166	1240k capture	This study	2346-2133 calBCE (Poz-68164)
I1389	10-Grave69	M	X2b4a	R1b1a2a1a2	97913	1240k capture	This study	2481-2289 calBCE (Poz-41229)
I1390	11-Grave68	M	X2b4a	R1b1a2a1a2	449568	1240k capture	This study	2566-2524 calBCE (Poz-41227)
I1381	2-Grave487	M	H	R1b1a2a1a	690979	1240k capture	This study	2400-1900 BCE
I1382	3-Grave515	M	U5a2c3a	R1b1a2a1a2	607837	1240k capture	This study	2456-2418 calBCE (GrA-4468)
I2478	T1 ind B US-9	M	K1a2a	R1b1a2a1a2c1b2b2a1	622069	1240k capture	This study	2200-1930 BCE (LTL-5035A)
I1767	Sk2	M	U5a1a1	I2a2a1a1a	729987	1240k capture	This study	2200-1970 calBCE
I1770	BurialXI	F	U5b1d2	-	483474	1240k capture	This study	..
I2453	skeleton 1126	M	K2a	R1b1a2a1a2c1	481035	1240k capture	This study	2500-1900 BCE
I2454	skeleton 5487	F	R1a1a	-	661010	1240k capture	This study	2200-1980 calBCE (OxA-24595)
I2455	skeleton 5486	F	U5a2c3a	-	95027	1240k capture	This study	2130-1910 calBCE (OxA-24594)
I2444	SK 8772 (YCF 95), 2	F	T2c1d+152	-	372507	1240k capture	This study	2290-1980 BCE

SupplementaryDataTable1

I2445	SK 8633 (YCF 95)	M	X2b6	R1b1a2a1a2c1	719230	1240k capture	This study	2400-2040 BCE
I2416	53535_25004	M	K1b1a1	R1b1a2a1a	136956	1240k capture	This study	..
I2418	56240_6033	F	K1a4a1	-	656573	1240k capture	This study	3835¬±25
I2457	65530_13382	M	I4a	R1b1a2a1a2c	718630	1240k capture	This study	2480-2280 calBCE (SUERC-36210)
I2459	85684_62014	F	T2a1a	-	519046	1240k capture	This study	2460-2140 calBCE (SUERC-54823)
I2364	GEN_10a	M	U5a2b	H2	767419	1240k capture	This study	3779 ¬± 28BP and 3883 ¬± 29 BP
I2365	GEN_11a	M	V3	R1b1a2a1a2	787155	1240k capture	This study	3858 ¬± 32 BP and 3871 ¬± 29 BP
E09537_d	68Fznr33Sk1UNTA58	F	K1a	-	87970	1240k capture	This study	2462-2213 calBCE (MAMS 23730)
E09538	68Fznr34Sk2UNTA58	M	J1c	G2a2a1a2a1	545714	1240k capture	This study	2470-2310 calBCE (MAMS 18935)
E09569	1343UNTA85	M	J1c2	R1b1a2a1a2b1	810078	1240k capture	This study	2397-2149 calBCE (MAMS 18949)
E09568_d	180Sk1HUGO	M	U5b2b	R1b1a2a1a	32968	1240k capture	This study	2461-2210 calBCE (MAMS 18918)
E09613_d	168HUGO	F	H1	-	174567	1240k capture	This study	2289-2141 calBCE (MAMS 18913)
E09614_d	190HUGO	F	J1c3	-	46955	1240k capture	This study	2268-2046 calBCE (MAMS 18921)
RISE559	-	F	H46	-	140816	Shotgun	AllentoftNature2015	..
RISE560	-	M	U5a1a1	R1b1a2	56722	Shotgun	AllentoftNature2015	..
RISE562	-	F	H2a1e	-	65499	Shotgun	AllentoftNature2015	..
RISE563	-	M	K1c1	R1b1a2a1a2b	333228	Shotgun	AllentoftNature2015	..
RISE564	-	M	H+16311	R1b1a2a1	91035	Shotgun	AllentoftNature2015	..
I1549	BZH15	F	W1c1	-	780302	1240k capture	MathiesonNature2015	2500-2050 BCE
I1546	BZH2	F	U5a1b1	-	149628	1240k capture	MathiesonNature2015	2500-2050 BCE
I0806	QLB28	M	H1	R1b1a2a1a2	138919	1240k capture	MathiesonNature2015	2431-2150 calBCE (MAMS 22820)
I0805	QLB26	M	H1	R1b1a2	225361	1240k capture	MathiesonNature2015	2467-2142 calBCE (Er8558)
I0113	QUEXII4	F	J1c5	-	585967	1240k capture	MathiesonNature2015	2346-2033 calBCE (Er7283)
I0112	QUEXII6	F	H13a1a2	-	949947	1240k capture	MathiesonNature2015	2457-2142 calBCE (Er7038)
I0060	ROT3	F	K1a2c	-	114800	1240k capture	MathiesonNature2015	2428-2149 calBCE (MAMS 22819)
I1530	ROT1	M	H3ao	R1b1a2	32225	1240k capture	MathiesonNature2015	2458-2140 calBCE (Er8715)
I0111	ROT4	F	H3ao	-	540050	1240k capture	MathiesonNature2015	2475-2204 calBCE (Er8712)
I0108	ROT6	F	H5a3	-	735474	1240k capture	MathiesonNature2015	2575-2299 calBCE (Er8710)
RISE568	-	F	H44a	-	60816	Shotgun	AllentoftNature2015	..
RISE569	-	F	H1af	-	708529	Shotgun	AllentoftNature2015	..
RISE566	-	M	H	R1b1a2a1a	117200	Shotgun	AllentoftNature2015	..
RISE567	-	F	U5b2c	-	31668	Shotgun	AllentoftNature2015	..

UDG	Damage restriction	Kinship	Steppe ancestry-based grouping	Site ID	Location	Latitude	Longitude	Country
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Cer	Cer	Paris Street, Cerdanyola, Barcelona	41.49	2.14	Spain
half	No	-	BB_Iberia_Alm	Alm	Galeria da Cisterna, Almonda	39.50	-8.62	Portugal
half	No	-	BB_Iberia_Alm	Alm	Galeria da Cisterna, Almonda	39.50	-8.62	Portugal
half	No	-	BB_Iberia_Arr1	Arr	Arroyal I, Burgos	42.40	-3.75	Spain
half	No	1st_rel_I0460	-	Arr	Arroyal I, Burgos	42.40	-3.75	Spain
half	No	-	BB_Iberia_Arr1	Arr	Arroyal I, Burgos	42.40	-3.75	Spain
half	No	-	BB_Iberia_Arr2	Arr	Arroyal I, Burgos	42.40	-3.75	Spain
half	No	-	BB_Iberia_Arr2	Arr	Arroyal I, Burgos	42.40	-3.75	Spain
half	No	-	BB_France_Heg	Heg	Hegenheim, Necropole, Haut-Rhin	47.56	7.52	France
half	No	-	BB_France_Mar	Mar	Marlens, Sur les Barmes, Haute-Savoie	45.46	6.21	France
half	No	-	BB_France_Rou	Rou	Rouffach, Haut-Rhin	47.96	7.30	France
half	No	1st_rel_I1390	-	Sie	Sierentz, Les Villas d'Aurele, Haut-Rhin	47.65	7.45	France
half	No	-	BB_France_Sie	Sie	Sierentz, Les Villas d'Aurele, Haut-Rhin	47.65	7.45	France
half	No	-	BB_France_Mon	Mon	Mondelange, PAC de la Sente, Moselle	49.15	6.10	France
half	No	-	BB_France_Mon	Mon	Mondelange, PAC de la Sente, Moselle	49.15	6.10	France
half	No	-	BB_Italy_Par	Par	via Guidorossi, Parma	44.78	10.30	Italy
half	No	-	BB_England_NYO	Ing	Ingleby Barwick, Stockton-on-Tees, County Durham	54.52	-1.31	England
half	No	-	BB_England_NYO	Sta	Staxton Beacon, Staxton, North Yorkshire	54.20	-0.43	England
half	No	-	BB_England_WOV	Wes	West Deeping	52.67	-0.35	England
half	No	-	BB_England_WOV	Ove	Over Barrows	52.34	0.03	England
half	No	-	BB_England_WOV	Ove	Over Barrows	52.34	0.03	England
half	No	-	BB_England_YAM2	Yar	Yarnton	51.80	-1.32	England

half	No	-	BB_England_YAM2	Yar	Yarnton	51.80	-1.32	England
half	No	-	BB_England_YAM1	Ame	Amesbury Down, Wiltshire	51.16	-1.77	England
half	No	-	BB_England_YAM2	Ame	Amesbury Down, Wiltshire	51.16	-1.77	England
half	No	-	BB_England_YAM2	Ame	Amesbury Down, Wiltshire	51.16	-1.77	England
half	No	-	BB_England_YAM2	Ame	Amesbury Down, Wiltshire	51.16	-1.77	England
half	No	-	BB_Hungary_Bud1	Bud	Budapest-Békásmegyer	47.60	19.05	Hungary
half	No	-	BB_Hungary_Bud2	Bud	Budapest-Békásmegyer	47.60	19.05	Hungary
half	Yes	-	BB_Germany_Aug	AugU	Unterer Talweg 58-62 (Augsburg)	48.32	10.89	Germany
half	No	-	BB_Germany_Aug	AugU	Unterer Talweg 58-62 (Augsburg)	48.32	10.89	Germany
half	No	-	BB_Germany_Aug	AugU	Unterer Talweg 58-62 (Augsburg)	48.32	10.89	Germany
half	Yes	-	BB_Germany_Aug	AugH	Hugo-Eckener-Straße (Augsburg)	48.33	10.90	Germany
half	Yes	-	BB_Germany_Aug	AugH	Hugo-Eckener-Straße (Augsburg)	48.33	10.90	Germany
half	Yes	-	BB_Germany_Aug	AugH	Hugo-Eckener-Straße (Augsburg)	48.33	10.90	Germany
minus	No	-	BB_Germany_Aug	AugA	Augsburg	48.33	10.90	Germany
minus	No	-	BB_Germany_Aug	AugA	Augsburg	48.33	10.90	Germany
minus	No	-	BB_Germany_Lan	Lan	Landau an der Isar	48.66	12.71	Germany
minus	No	-	BB_Germany_Ost2	Ost	Osterhofen-Altenmarkt	48.69	13.02	Germany
minus	No	-	BB_Germany_Ost1	Ost	Osterhofen-Altenmarkt	48.69	13.02	Germany
half	No	-	BB_Germany_SAN	Ben	Benzingerode-Heimburg	51.82	10.91	Germany
half	No	-	BB_Germany_SAN	Ben	Benzingerode-Heimburg	51.82	10.91	Germany
half	No	-	BB_Germany_SAN	QVII	Quedlinburg VII	51.79	11.14	Germany
half	No	-	BB_Germany_SAN	QVII	Quedlinburg VII	51.79	11.15	Germany
Yes	No	-	BB_Germany_SAN	QXII	Quedlinburg XII	51.79	11.14	Germany
plus	No	-	BB_Germany_SAN	QXII	Quedlinburg XII	51.79	11.14	Germany
plus	No	-	BB_Germany_SAN	Rot	Rothenschirmbach	51.45	11.54	Germany
half	No	1st_rel_I0111	-	Rot	Rothenschirmbach	51.45	11.54	Germany
Yes	No	-	BB_Germany_SAN	Rot	Rothenschirmbach	51.45	11.54	Germany
Yes	No	-	BB_Germany_SAN	Rot	Rothenschirmbach	51.45	11.54	Germany
minus	No	-	BB_Czech_CBO	Bra	Brandysek	50.19	14.16	Czech Republic
minus	No	-	BB_Czech_CBO	Bra	Brandysek	50.19	14.16	Czech Republic
minus	No	-	BB_Czech_CBO	Kne	Knezeves	50.12	14.26	Czech Republic
minus	No	-	BB_Czech_CBO	Kne	Knezeves	50.12	14.26	Czech Republic

SupplementaryDataTable2

ID	Sex	mtDNA haplogroup	SNPs in autosomes	Data type	Publication	Date(95% CI)	UDG
I0434	M	U4a2 or U4d	61851	1240k capture	MathiesonNature2015	5200-4000 BCE	half
I0433	M	U5a1i	431545	1240k capture	MathiesonNature2015	5200-4000 BCE	half
I0122	M	H2a1	569453	1240k capture	MathiesonNature2015	5200-4000 BCE	Yes
I0124	M	U5a1d	477876	1240k capture	MathiesonNature2015	5650-5555 calBCE (Beta 392490)	half
I0231	M	U4a1a or U4a1d	1065807	1240k capture	MathiesonNature2015	2910-2875 calBCE (Beta 392487)	Yes
I0370	M	H13a1a1	695086	1240k capture	MathiesonNature2015	3300-2700 BCE	half
I0441	F	H2b	101678	1240k capture	MathiesonNature2015	3010-2622 calBCE (AA47805)	half
I0444	M	H6a1b	539947	1240k capture	MathiesonNature2015	3300-2700 BCE	half
I0439	M	U5a1a1	268426	1240k capture	MathiesonNature2015	3305-2925 calBCE (Beta 392491)	half
I0357	F	W6c	523435	1240k capture	MathiesonNature2015	3090-2910 calBCE	Yes
I0429	M	T2c1a2	603212	1240k capture	MathiesonNature2015	3339-2917 calBCE (AA47804)	half
I0438	M	U5a1a1	567478	1240k capture	MathiesonNature2015	3021-2635 calBCE (AA47807)	half
I1282	M	H3	68114	1240k capture	MathiesonNature2015	2880-2630 BCE (Beta-296225)	half
I1276	F	H3c3	113480	1240k capture	MathiesonNature2015	2880-2630 BCE (Beta-296225)	half
I1284	M	H3	123157	1240k capture	MathiesonNature2015	2880-2630 BCE (Beta-296225)	half
I1280	F	J1c1	202133	1240k capture	MathiesonNature2015	2880-2630 BCE (Beta-296225)	half
I1314	M	J2a1a1	220329	1240k capture	MathiesonNature2015	2880-2630 BCE (Beta-296225)	half
I1277	M	H3	553616	1240k capture	MathiesonNature2015	2830-2820 cal BCE (Beta-416457)	half
I1272	F	K1b1a1	578928	1240k capture	MathiesonNature2015	2880-2830 cal-BCE (Beta 416456)	half
I1281	F	H1t	623901	1240k capture	MathiesonNature2015	2895-2855 cal BCE (Beta-416458)	half
I1300	F	K1a2a	779093	1240k capture	MathiesonNature2015	2880-2630 BCE (Beta-296225)	half
I0585	M	U5b2c	976082	1240k capture	MathiesonNature2015	5990-5740 calBCE (Beta-226472)	half
I0807	M	H1e1a	91637	1240k capture	MathiesonNature2015	3970-3710 calBCE (Er7784)	half
I0559	M	HV	175263	1240k capture	MathiesonNature2015	3652-3527 calBCE (MAMS 22818)	half
I0560	F	T2e1	375978	1240k capture	MathiesonNature2015	3640-3376 calBCE (Er7856)	half
I0171	F	U5a1a2a	140559	1240k capture	MathiesonNature2015	2287-2041 calBCE (KIA27952)	plus
I0059	F	H1	725307	1240k capture	MathiesonNature2015	2337-2138 calBCE (MAMS 21486)	Yes
I1542	M	J2a2a	76340	1240k capture	MathiesonNature2015	2500-2050 BCE	half
I1536	M	U5a1g	114655	1240k capture	MathiesonNature2015	2500-2050 BCE	half
I1544	M	K2b2	114655	1240k capture	MathiesonNature2015	2500-2050 BCE	half

SupplementaryDataTable2

I1538	M	J1c5	135269	1240k capture	MathiesonNature2015	2500-2050 BCE	half
I1539	F	J1c1b1a	138220	1240k capture	MathiesonNature2015	2625-2291 calBCE (Er7779)	half
I1534	M	K1a1b2a	164095	1240k capture	MathiesonNature2015	2500-2050 BCE	half
I0106	F	T2a1b1	222103	1240k capture	MathiesonNature2015	2464-2210 calBCE (MAMS 21490)	plus
I1540	M	J1c5	285866	1240k capture	MathiesonNature2015	2500-2050 BCE	half
I1532	M	J1c2e	464060	1240k capture	MathiesonNature2015	2500-2050 BCE	half
I0049	F	X2b4	514322	1240k capture	MathiesonNature2015	2464-2210 calBCE (MAMS 21489)	Yes
I0099	M	H23	962265	1240k capture	New libraries added to previously	1193-979 calBCE (MAMS 21484)	half
I0211	M	U4a	146885	1240k capture	MathiesonNature2015	5500-5000 BCE	half
I0061	M	C1	1025148	1240k capture	MathiesonNature2015	8800-7950 BP	Yes
I0550	F	T1a1	161187	1240k capture	MathiesonNature2015	2570-2471 calBCE (MAMS 23344)	half
I0821	M	X2d1	393410	1240k capture	New libraries added to previously	5201-4850 calBCE (KIA40348)	half
I0057	F	N1a1a1	243485	1240k capture	New libraries added to previously	5218-5019 calBCE (MAMS 21483)	half
I0048	M	K1a	437950	1240k capture	New libraries added to previously	5211-5009 calBCE (MAMS 21482)	half
I0659	M	N1a1a1a2	597818	1240k capture	New libraries added to previously	5211-4963 calBCE (KIA40350)	half
I0046	F	T2c1	915831	1240k capture	New libraries added to previously	5212-4989 calBCE (MAMS 21479)	half
I0797	M	H46b	94125	1240k capture	MathiesonNature2015	5500-4775 BCE	half
I0795	M	H1 or H1au1b	104752	1240k capture	MathiesonNature2015	5216-5036 calBCE (MAMS 22823)	half
I0176	F	N1a1a1a3	69444	1240k capture	MathiesonNature2015	5207-4944 calBCE (Beta-310038)	plus
I0551	M	U3a1	102382	1240k capture	MathiesonNature2015	3400-3025 BCE	half
I0409	F	J1c3	511453	1240k capture	MathiesonNature2015	5310-5218 calBCE (MAMS 16159)	Yes
I0405	M	K1a1b1	321750	1240k capture	MathiesonNature2015	3900-3600 BCE	plus
I0174	M	N1a1a1	227966	1240k capture	MathiesonNature2015	5702-5536 calBCE (MAMS 11939)	plus
I0115	F	U5a1i	352282	1240k capture	MathiesonNature2015	1954-1760 calBCE (MAMS 21494)	plus
I0117	F	I3a	843093	1240k capture	MathiesonNature2015	2272-2039 calBCE (MAMS 21496)	Yes
I0804	M	H3	61524	1240k capture	MathiesonNature2015	2137-1965 calBCE (MAMS 22821)	half
I0803	F	H4a1a1a	213212	1240k capture	MathiesonNature2015	2132-1942 calBCE (MAMS 22822)	half
I0047	F	V9	840659	1240k capture	New libraries added to previously	2111-1891 calBCE (MAMS 21481)	half
I0164	F	U5b2a1b	834837	1240k capture	MathiesonNature2015	2023-1894 calBCE (MAMS 21497)	plus
I0022	F	T2e	368546	1240k capture	MathiesonNature2015	5500-4800 BCE	half
I0026	F	T2b	893563	1240k capture	MathiesonNature2015	5500-4800 BCE	half
I0013	M	U5a1	318346	1240k capture	MathiesonNature2015	5898-5531 calBCE	plus

SupplementaryDataTable2

I0011	F	U5a1	621320	1240k capture	MathiesonNature2015	5898-5531 calBCE	plus
I0015	M	U5a2d	600144	1240k capture	MathiesonNature2015	5898-5531 calBCE	plus
I0012	M	U2e1	787312	1240k capture	MathiesonNature2015	5898-5531 calBCE	plus
I0014	F	U5a2d	739313	1240k capture	MathiesonNature2015	5898-5531 calBCE	plus
I0017	M	U2e1	920441	1240k capture	MathiesonNature2015	5898-5531 calBCE	plus
I1508	F	K1a	565253	1240k capture	MathiesonNature2015	5710-5570 calBCE (OxA-28101)	half
I1500	M	J1c1	830691	1240k capture	MathiesonNature2015	5210-4990 calBCE (OxA-23763)	half
I1100	F	K1a or K1a6	294382	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1102	M	K1a3a	414820	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1099	M	T2b	544031	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1103	M	K1b1b1	607546	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1101	M	T2b	663312	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1097	M	W1-T119C	755758	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I0744	M	J1c11	884258	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1579	F	K1a-C150T	935726	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1581	F	U3	892187	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1096	M	N1a1a1	759115	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1580	F	H5	974979	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1098	F	X2d2	780873	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1585	F	J1 or J1c	953325	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I0708	M	N1b1a	980130	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I0745	M	U8b1b1	996769	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I0746	M	K1a or K1a1	1003994	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1583	M	K1a2	1023789	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I0707	F	K1a4	990132	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I0709	M	U3	987745	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I0736	F	N1a1a1a	799433	1240k capture	MathiesonNature2015	6500-6200 BCE	half
I1499	F	X2b-T226C	817728	1240k capture	MathiesonNature2015	5210-5010 calBCE (OxA-27732)	half
I1502	F	K1c1	795741	1240k capture	MathiesonNature2015	2190-1980 calBCE (OxA-23799)	half
I1497	F	H	813947	1240k capture	MathiesonNature2015	2900-2700 calBCE (MAMS-14825)	half
I1495	M	N1a1a1a	829966	1240k capture	MathiesonNature2015	4490-4360 calBCE (MAMS-14819)	half
I1498	F	H	810430	1240k capture	MathiesonNature2015	5290-5060 calBCE (OxA-27858)	half

SupplementaryDataTable2

I1507	M	R1b1	800965	1240k capture	MathiesonNature2015	5780-5650 calBCE (OxA-23757)	half
I1504	M	K1a1a	687165	1240k capture	MathiesonNature2015	1270-1110 calBCE (OxA-27859)	half
I1506	F	U5b2c	704606	1240k capture	MathiesonNature2015	5310-5070 calBCE (OxA-27861)	half
I1496	M	K1a3a3	812201	1240k capture	MathiesonNature2015	5300-4950 calBCE (MAMS-14821)	half
I1505	F	J1c5	769987	1240k capture	MathiesonNature2015	5290-5050 calBCE (OxA-28020)	half
MA1	M	U	820035	Shotgun	RaghavanNature2014	24520-24090 calBP (UCIAMS-7966614)	minus
Kostenki14.SG	M	U2	1070534	Shotgun	Seguin-OrlandoScience2014	37845-38650 calBCE (OxA-X-2395-15)	Yes
Iceman	M	K1	1162609	Shotgun	KellerNatureCommunications2012	3359-3105 BCE	minus
RISE00	F	H5a1	631510	Shotgun	AllentoftNature2015	..	minus
RISE47	M	I	102112	Shotgun	AllentoftNature2015	1499-1324 calBCE (OxA-28258)	minus
RISE61	M	J1c4	287949	Shotgun	AllentoftNature2015	2851-2492 calBCE (OxA-28296)	minus
RISE71	F	H3b	198510	Shotgun	AllentoftNature2015	2196-2023 calBCE (OxA-28269)	minus
RISE94	M	K1a2a	607380	Shotgun	AllentoftNature2015	2621-2472 calBCE (OxA-29033)	minus
RISE97	F	K2a5	561162	Shotgun	AllentoftNature2015	2025-1885 calBCE (OxA-28986)	minus
RISE98	M	K1b1a1	1089349	Shotgun	AllentoftNature2015	2275-2032 calBCE (OxA-28987)	minus
RISE109	F	U4	205999	Shotgun	AllentoftNature2015	1954-1772 calBCE (UB-16557)	minus
RISE150	F	U5a1b1	635649	Shotgun	AllentoftNature2015	1885-1693 calBCE (Ua-42401)	minus
RISE154	F	K1a4a1	165439	Shotgun	AllentoftNature2015	1925-1765 calBCE (Uba-16555)	minus
RISE175	M	T1a1	111000	Shotgun	AllentoftNature2015	1395-1132 calBCE (OxA-28998)	minus
RISE179	M	K1a3	50109	Shotgun	AllentoftNature2015	2010-1776 calBCE (OxA-29193)	minus
RISE210	M	T2a1a	50468	Shotgun	AllentoftNature2015	1432-1292 calBCE (OxA-29654)	minus
RISE240	F	U5a1d1	206474	Shotgun	AllentoftNature2015	2880-2632 calBCE (GrA-45038)	minus
RISE247	M	H11a	233887	Shotgun	AllentoftNature2015	1746-1611 calBCE (OxA-29769)	minus
RISE254	M	J1c9	42379	Shotgun	AllentoftNature2015	2128-1909 calBCE (OxA-29842)	minus
RISE276	M	T2b	91597	Shotgun	AllentoftNature2015	794-547 calBCE (OxA-30485)	minus
RISE349	F	T2b3	92396	Shotgun	AllentoftNature2015	2034-1784 calBCE (OxA-30987)	minus
RISE371	F	U5a2b	51763	Shotgun	AllentoftNature2015	2136-1941 calBCE (OxA-30988)	minus
RISE373	F	K1a2a	250888	Shotgun	AllentoftNature2015	1886-1696 calBCE (OxA-31104)	minus
RISE374	M	T2b	143488	Shotgun	AllentoftNature2015	1866-1619 calBCE (OxA-30989)	minus
RISE431	M	T2e	122135	Shotgun	AllentoftNature2015	2286-2048 calBCE (OxA-27967)	minus
RISE434	M	U4	109337	Shotgun	AllentoftNature2015	2880-2630 calBCE (UBA-27946)	minus
RISE435	F	J1b1a1	49066	Shotgun	AllentoftNature2015	2863-2498 calBCE (UBA-27947)	minus

SupplementaryDataTable2

RISE436	M	U5b1c2	62088	Shotgun	AllentoftNature2015	2868-2580 calBCE (UBA-27948)	minus
RISE446	M	U5b1c2	184657	Shotgun	AllentoftNature2015	2829-2465 calBCE (UBA-27950)	minus
RISE471	M	J1c1b	114375	Shotgun	AllentoftNature2015	..	minus
RISE479	M	T2b	861756	Shotgun	AllentoftNature2015	2000-1500 BCE	minus
RISE480	F	U5a2a	153887	Shotgun	AllentoftNature2015	1700-1500 calBCE	minus
RISE483	F	H2a1	130881	Shotgun	AllentoftNature2015	2000-1500 BCE	minus
RISE484	F	T1a1	163477	Shotgun	AllentoftNature2015	2000-1500 BCE	minus
RISE486	M	J1c1b	199113	Shotgun	AllentoftNature2015	2134-1773 calBCE (ETH-12913)	minus
RISE487	M	H2a	233950	Shotgun	AllentoftNature2015	3483-3107 calBCE (OxA-X-2621)	minus
RISE489	M	X2c1	498464	Shotgun	AllentoftNature2015	2908-2578 calBCE (ETH-12188)	minus
RISE546	M	U5a1d2b	164552	Shotgun	AllentoftNature2015	..	minus
RISE547	M	T2a1a	613723	Shotgun	AllentoftNature2015	2887-2634 calBCE (GrA-58960)	minus
RISE548	M	U4	717431	Shotgun	AllentoftNature2015	..	minus
RISE550	M	U5a1i	467373	Shotgun	AllentoftNature2015	3334-2635 calBCE (IGAN-2880)	minus
RISE552	M	T2a1a	945230	Shotgun	AllentoftNature2015	2849-2143 calBCE (IGAN-4079)	minus
RISE555	M	N1a1a-T152C	226813	Shotgun	AllentoftNature2015	2857-2497 calBCE (AAR-20358)	minus
RISE577	F	T2b	711716	Shotgun	AllentoftNature2015	..	minus
RISE586	F	K1b1a	188038	Shotgun	AllentoftNature2015	..	minus
Loschbour	M	U5b1a	1139327	Shotgun	LazaridisNature2014	8160-7940 calBP (OxA-773815)	minus
Ust_Ishim	M	R	1179991	Shotgun	FuNature2014	47480-42560 calBP (OxA-25516)	Yes
Stuttgart	F	T2c1b	1130723	Shotgun	LazaridisNature2014	7260-7020 calBP (MAMS-24635)	minus
I1271	F	K1a	231359	1240k capture	MathiesonNature2015	2880-2630 BCE (Beta-296225)	half
I1303	M	U3a1	457683	1240k capture	MathiesonNature2015	2880-2630 BCE (Beta-296225)	half
I1550	F	K1a2	545732	1240k capture	New libraries added to previously	5500-4775 BCE	half
I0100	F	N1a1a1a	982315	1240k capture	New libraries added to previously	5202-4852 calBCE (KIA40341)	half
I0407	F	K1b1a1	552221	1240k capture	MathiesonNature2015	3900-3600 BCE	plus
I0025	F	T2b	836923	1240k capture	MathiesonNature2015	5500-4800 BCE	half
I0443	M	W3a1a	995838	1240k capture	MathiesonNature2015	3300-2700 BCE	Yes
I0581	M	X2b	1027108	1240k capture	MathiesonNature2015	3010-2975 cal BCE (Beta-416455)	Yes
I0726	F	H or H5-C16192T	228543	1240k capture	MathiesonNature2015	6400-5600 BCE	Yes
I0103	F	W6a	1049840	1240k capture	MathiesonNature2015	2578-2468 calBCE (MAMS 21488)	Yes
I0172	M	T2b	1035229	1240k capture	MathiesonNature2015	3360-3086 calBCE (Er8699)	Yes

SupplementaryDataTable2

I0408	F	U5b1	907690	1240k capture	MathiesonNature2015	3900-3600 BCE	Yes
I0412	M	N1a1a1	983710	1240k capture	MathiesonNature2015	5308-5080 calBCE (MAMS 16164)	Yes
I0054	F	J1c17	978534	1240k capture	MathiesonNature2015	5222-5022 calBCE (MAMS 21485)	Yes
I0056	M	T2b	367760	1240k capture	New libraries added to previously	5212-5006 calBCE (MAMS 21480)	half
I0104	M	U4b1a1a1	962767	1240k capture	MathiesonNature2015	2559-2296 calBCE (MAMS 21487)	Yes
I0116	M	W3a1	629665	1240k capture	MathiesonNature2015	2134-1939 calBCE (MAMS 21495)	plus
I0118	F	HV	1048626	1240k capture	MathiesonNature2015	2471-2246 calBCE (MAMS 21492)	Yes
I0406	M	H1	724111	1240k capture	MathiesonNature2015	3900-3600 BCE	plus
I0410	M	T2c1d or T2c1d2	521870	1240k capture	MathiesonNature2015	5295-5066 calBCE (MAMS 16161)	plus
I0413	F	V	638681	1240k capture	MathiesonNature2015	5302-5074 calBCE (MAMS 16166)	plus
I0723	M	X2m2	702849	1240k capture	MathiesonNature2015	6400-5600 BCE	Yes
I0724	M	K1a4	234996	1240k capture	MathiesonNature2015	6400-5600 BCE	Yes
I0727	M	K1a2	183324	1240k capture	MathiesonNature2015	6400-5600 BCE	Yes
CB13	F	K1a2a	717337	Shotgun	OlaldeMBE2015	5470-5320 calBCE (Beta-384724)	minus
ATP2	M	U5b3	1160576	Shotgun	GuntherPNAS2015	2899-2678 calBCE (Beta-386394)	minus
ATP9	F	U5b1b	430551	Shotgun	GuntherPNAS2015	1750-1618 calBCE (Beta-386395)	minus
ATP16	F	X2c	803446	Shotgun	GuntherPNAS2015	3261-2916 calBCE (Beta-368289)	minus
Matojo	M	H3c	884208	Shotgun	GuntherPNAS2015	3010-3879 calBCE (Beta-368295)	minus
KK1	M	H13c	1180629	Shotgun	JonesNatureCommunications2015	9890-9550 calBP (RTT5246)	minus
SATP	M	K3	816136	Shotgun	JonesNatureCommunications2015	13380-13130 calBP (OxA-346322)	minus
Bichon	M	U5b1h	1176132	Shotgun	JonesNatureCommunications2015	13770-13560 calBP (OxA-277632)	minus
bally	F	HV0+195	1150887	Shotgun	CassidyPNAS2016	3343-3020 calBCE (UB-7059)	minus
rath1	M	U5a1b1e	1181857	Shotgun	CassidyPNAS2016	2026-1885 calBCE (UBA-8707)	minus
rath2	M	U5b2a2	946668	Shotgun	CassidyPNAS2016	2024-1741 calBCE (UBA-8705)	minus
rath3	M	J2b1a	673050	Shotgun	CassidyPNAS2016	1736-1534 calBCE (UBA-8706)	minus
I0520	M	U5a2c	34803	1240k capture	This study	3360-3100 BCE	half
I0518	M	K1a3a1	64753	1240k capture	This study	3360-3100 BCE	half
I0519	M	X2b+226	164190	1240k capture	This study	3360-3100 BCE	half

Population Label	PCA label	Location	Latitude	Longitude	Country
Samara_Eneolithic	Steppe_LNBA	Khvalynsk II, Volga River, Samara	52.22	48.10	Russia
Samara_Eneolithic	Steppe_LNBA	Khvalynsk II, Volga River, Samara	52.22	48.10	Russia
Samara_Eneolithic	Steppe_LNBA	Khvalynsk II, Volga River, Samara	52.22	48.10	Russia
EHG	European_HG	Lebyanzhinka IV, Sok River, Samara	53.40	50.40	Russia
Yamnaya_Samara	Steppe_LNBA	Ekaterinovka, Southern Steppe, Samara	52.42	48.24	Russia
Yamnaya_Samara	Steppe_LNBA	Ishkinovka I, Eastern Orenburg, Samara	51.27	58.18	Russia
Yamnaya_Samara	Steppe_LNBA	Kurmanaevka III, Buzuluk, Samara	52.30	52.05	Russia
Yamnaya_Samara	Steppe_LNBA	Kutuluk I, Kutuluk River, Samara	53.10	51.13	Russia
Yamnaya_Samara	Steppe_LNBA	Lopatino I, Sok River, Samara	53.38	50.39	Russia
Yamnaya_Samara	Steppe_LNBA	Lopatino I, Sok River, Samara	53.38	50.39	Russia
Yamnaya_Samara	Steppe_LNBA	Lopatino I, Sok River, Samara	53.38	50.39	Russia
Yamnaya_Samara	Steppe_LNBA	Luzkhi I, Samara River, Samara	53.38	50.38	Russia
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
WHG	European_HG	La Brana-Arintero, León	42.91	-5.38	Spain
Central_MN	European_MN_Chalc	Esperstedt	51.42	11.68	Germany
Central_MN	European_MN_Chalc	Quedlinburg IX	51.79	11.14	Germany
Central_MN	European_MN_Chalc	Quedlinburg IX	51.79	11.14	Germany
Central_LNBA	Europe_LNBA	Benzingerode-Heimburg	51.82	10.91	Germany
Central_LNBA	Europe_LNBA	Benzingerode-Heimburg	51.82	10.91	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany

Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Central_LNBA	Europe_LNBA	Halberstadt-Sonntagsfeld	51.89	11.04	Germany
EHG	European_HG	Yuzhnyy Oleni Ostrov, Karelia	61.65	35.65	Russia
EHG	European_HG	Yuzhnyy Oleni Ostrov, Karelia	61.65	35.65	Russia
Central_LNBA	Europe_LNBA	Karsdorf	51.27	11.66	Germany
LBK_EN	European_EN_Anat_N	Halberstadt-Sonntagsfeld	51.90	11.05	Germany
LBK_EN	European_EN_Anat_N	Halberstadt-Sonntagsfeld	51.89	11.04	Germany
LBK_EN	European_EN_Anat_N	Halberstadt-Sonntagsfeld	51.89	11.04	Germany
LBK_EN	European_EN_Anat_N	Halberstadt-Sonntagsfeld	51.89	11.04	Germany
LBK_EN	European_EN_Anat_N	Halberstadt-Sonntagsfeld	51.89	11.04	Germany
LBK_EN	European_EN_Anat_N	Karsdorf	51.28	11.65	Germany
LBK_EN	European_EN_Anat_N	Karsdorf	51.27	11.66	Germany
Hungary_EN	European_EN_Anat_N	Szemely-Hegyes	46.40	18.74	Hungary
Central_MN	European_MN_Chalc	Salzmuende-Schiebzig	51.52	11.85	Germany
Iberia_EN	European_EN_Anat_N	Els Trocs	42.50	0.50	Spain
Iberia_MN	European_MN_Chalc	La Mina	41.25	-2.33	Spain
Hungary_EN	European_EN_Anat_N	Alsonyek-Bataszek, Mernoki telep	46.20	18.70	Hungary
Central_LNBA	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Central_LNBA	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Central_LNBA	Europe_LNBA	Eulau	51.17	11.85	Germany
Central_LNBA	Europe_LNBA	Eulau	51.17	11.85	Germany
Central_LNBA	Europe_LNBA	Halberstadt-Sonntagsfeld	51.89	11.04	Germany
Central_LNBA	Europe_LNBA	Quedlinburg VIII	51.79	11.14	Germany
LBK_EN	European_EN_Anat_N	Viesenhaeuser Hof, Stuttgart-Muehlhausen	48.78	9.18	Germany
LBK_EN	European_EN_Anat_N	Viesenhaeuser Hof, Stuttgart-Muehlhausen	48.78	9.18	Germany
SHG	European_HG	Motala	58.54	15.05	Sweden

SupplementaryDataTable2

SHG	European_HG	Motala	58.54	15.05	Sweden
SHG	European_HG	Motala	58.54	15.05	Sweden
SHG	European_HG	Motala	58.54	15.05	Sweden
SHG	European_HG	Motala	58.54	15.05	Sweden
SHG	European_HG	Motala	58.54	15.05	Sweden
Hungary_EN	European_EN_Anat_N	Berettyóújfalu-Morotva-Liget	47.32	21.53	Hungary
Hungary_EN	European_EN_Anat_N	Kompolt-Kigyoser	47.17	20.83	Hungary
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Barcın	40.30	29.57	Turkey
Hungary_EN	European_EN_Anat_N	Garadna	48.52	21.17	Hungary
Hungary_BA	Europe_LNBA	Kompolt-Kigyoser	47.17	20.83	Hungary
Hungary_CA	European_MN_Chalc	Apc-Berekalya I	47.17	19.83	Hungary
Hungary_EN	European_EN_Anat_N	Apc-Berekalya I	47.17	19.83	Hungary
Hungary_EN	European_EN_Anat_N	Debrecen Tocopart Erdoalja	47.52	21.59	Hungary

SupplementaryDataTable2

WHG	European_HG	Tiszaszolos-Domahaza	47.93	21.20	Hungary
Hungary_BA	Europe_LNBA	Ludas-Varju-Dulo	47.82	19.95	Hungary
Hungary_EN	European_EN_Anat_N	Polgar Ferenci hat	47.88	21.19	Hungary
Hungary_EN	European_EN_Anat_N	Apc-Berekalya I	47.17	19.83	Hungary
Hungary_EN	European_EN_Anat_N	Polgar Ferenci hat	47.88	21.19	Hungary
MA1	..	Mal'ta	52.90	103.50	Russia
Kostenki14	..	Kostenki	51.23	39.30	Russia
Iceman	European_MN_Chalc	Tisenjoch	46.78	10.83	Italy
Corded_Ware_Estonia.SG	Europe_LNBA	Sope	59.41	27.03	Estonia
Northern_LNBA	Europe_LNBA	Sebber skole	56.97	9.55	Denmark
Northern_LNBA	Europe_LNBA	Kyndelöse	55.70	11.86	Denmark
Northern_LNBA	Europe_LNBA	Falshöj	56.68	10.03	Denmark
Northern_LNBA	Europe_LNBA	Viby	56.03	14.23	Sweden
Northern_LNBA	Europe_LNBA	Fredriksberg	55.56	13.06	Sweden
Northern_LNBA	Europe_LNBA	L Beddinge 56	55.38	13.45	Sweden
Central_LNBA	Europe_LNBA	Wojkowice	50.98	17.07	Poland
Central_LNBA	Europe_LNBA	Przeclawice	50.92	16.96	Poland
Central_LNBA	Europe_LNBA	Szczepankowice	50.95	16.94	Poland
Northern_LNBA	Europe_LNBA	Abekås I	55.40	13.60	Sweden
Northern_LNBA	Europe_LNBA	Abekås I	55.40	13.60	Sweden
Northern_LNBA	Europe_LNBA	Ängamöllan	56.00	14.10	Sweden
Yamnaya_Kalmykia.SG	Steppe_LNBA	Sukhaya Termista I	46.58	43.68	Russia
Hungary_BA	Europe_LNBA	Százhalombatta-Földvár	47.33	18.96	Hungary
Hungary_BA	Europe_LNBA	Százhalombatta-Földvár	47.33	18.96	Hungary
Northern_LNBA	Europe_LNBA	Trundholm mose II	55.91	11.57	Denmark
Hungary_BA	Europe_LNBA	Battonya Vörös Oktober	46.36	20.99	Hungary
Hungary_BA	Europe_LNBA	Szöreg - C (Sziv Utca)	46.22	20.20	Hungary
Hungary_BA	Europe_LNBA	Szöreg - C (Sziv Utca)	46.22	20.20	Hungary
Hungary_BA	Europe_LNBA	Szöreg - C (Sziv Utca)	46.22	20.20	Hungary
Corded_Ware_Poland.SG	Europe_LNBA	Leki Male	52.14	16.54	Poland
Corded_Ware_Germany.SG	Europe_LNBA	Tiefbrunn	48.93	12.26	Germany
Corded_Ware_Germany.SG	Europe_LNBA	Tiefbrunn	48.93	12.26	Germany

SupplementaryDataTable2

Corded_Ware_Germany.SG	Europe_LNBA	Tiefbrunn	48.93	12.26	Germany
Corded_Ware_Germany.SG	Europe_LNBA	Bergrheinfeld	50.01	10.18	Germany
Central_LNBA	Europe_LNBA	Untermeitingen	48.17	10.81	Germany
Hungary_BA	Europe_LNBA	Erd 4	47.34	18.90	Hungary
Hungary_BA	Europe_LNBA	Erd 4	47.34	18.90	Hungary
Hungary_BA	Europe_LNBA	Erd 4	47.34	18.90	Hungary
Hungary_BA	Europe_LNBA	Erd 4	47.34	18.90	Hungary
Remedello_CA	European_MN_Chalc	Remedello di Sotto	45.26	10.38	Italy
Remedello_CA	European_MN_Chalc	Remedello di Sotto	45.26	10.38	Italy
Remedello_CA	European_MN_Chalc	Remedello di Sotto	45.26	10.38	Italy
Yamnaya_Kalmykia.SG	Steppe_LNBA	Temrta IV	46.54	43.70	Russia
Yamnaya_Kalmykia.SG	Steppe_LNBA	Temrta IV	46.54	43.70	Russia
Yamnaya_Kalmykia.SG	Steppe_LNBA	Temrta IV	46.54	43.70	Russia
Yamnaya_Kalmykia.SG	Steppe_LNBA	Peshany V	46.56	43.68	Russia
Yamnaya_Kalmykia.SG	Steppe_LNBA	Ulan IV	46.62	43.33	Russia
Russia_EBA	Steppe_LNBA	Stalingrad Quarry	48.72	44.50	Russia
Central_LNBA	Europe_LNBA	Velke Prilepy	50.16	14.31	Czech Republic
Central_LNBA	Europe_LNBA	Moravska Nova Ves	48.80	17.02	Czech Republic
WHG	European_HG	Echternach	49.70	6.24	Luxembourg
Ust_Ishim	..	Ust'-Ishim	57.43	71.10	Russia
LBK_EN	European_EN_Anat_N	Viesenhaeuser Hof, Stuttgart-Muehlhausen	48.78	9.18	Germany
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
LBK_EN	European_EN_Anat_N	Halberstadt-Sonntagsfeld	51.89	11.04	Germany
LBK_EN	European_EN_Anat_N	Halberstadt-Sonntagsfeld	51.89	11.04	Germany
Iberia_MN	European_MN_Chalc	La Mina	41.25	-2.33	Spain
LBK_EN	European_EN_Anat_N	Viesenhaeuser Hof, Stuttgart-Muehlhausen	48.78	9.18	Germany
Yamnaya_Samara	Steppe_LNBA	Lopatino II, Sok River, Samara	53.38	50.38	Russia
Iberia_Chalcolithic	European_MN_Chalc	El Mirador Cave, Atapuerca, Burgos	42.33	-3.50	Spain
Anatolia_Neolithic	European_EN_Anat_N	Mentese	40.26	29.65	Turkey
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Central_MN	European_MN_Chalc	Esperstedt	51.42	11.68	Germany

SupplementaryDataTable2

Iberia_MN	European_MN_Chalc	La Mina	41.25	-2.33	Spain
Iberia_EN	European_EN_Anat_N	Els Trocs	42.50	0.50	Spain
LBK_EN	European_EN_Anat_N	Unterwiederstedt	51.66	11.53	Germany
LBK_EN	European_EN_Anat_N	Halberstadt-Sonntagsfeld	51.89	11.04	Germany
Corded_Ware_Germany	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Central_LNBA	Europe_LNBA	Esperstedt	51.42	11.68	Germany
Central_LNBA	Europe_LNBA	Alberstedt	51.45	11.63	Germany
Iberia_MN	European_MN_Chalc	La Mina	41.25	-2.33	Spain
Iberia_EN	European_EN_Anat_N	Els Trocs	42.50	0.50	Spain
Iberia_EN	European_EN_Anat_N	Els Trocs	42.50	0.50	Spain
Anatolia_Neolithic	European_EN_Anat_N	Mentese	40.26	29.65	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Mentese	40.26	29.65	Turkey
Anatolia_Neolithic	European_EN_Anat_N	Mentese	40.26	29.65	Turkey
Iberia_EN.SG	European_EN_Anat_N	Cova Bonica, Vallirana, Barcelona	41.38	1.93	Spain
Iberia_Chalcolithic.SG	European_MN_Chalc	El Portalon Cave, Sierra de Atapuerca	42.35	-3.52	Spain
Iberia_Bronze_Age.SG	Europe_LNBA	El Portalon Cave, Sierra de Atapuerca	42.35	-3.52	Spain
Iberia_Chalcolithic.SG	European_MN_Chalc	El Portalon Cave, Sierra de Atapuerca	42.35	-3.52	Spain
Iberia_Chalcolithic.SG	European_MN_Chalc	El Portalon Cave, Sierra de Atapuerca	42.35	-3.52	Spain
CHG	CHG	Kotias	42.13	43.12	Georgia
CHG	CHG	Satsurblia	42.24	42.92	Georgia
Switzerland_HG	European_HG	Bichon	47.01	6.79	Switzerland
Ireland_MN.SG	European_MN_Chalc	Ballynahatty, County Down	54.54	-5.96	Ireland
Ireland_BA.SG	Europe_LNBA	Rathlin Island, County Antrim	55.29	-6.19	Ireland
Ireland_BA.SG	Europe_LNBA	Rathlin Island, County Antrim	55.29	-6.19	Ireland
Ireland_BA.SG	Europe_LNBA	Rathlin Island, County Antrim	55.29	-6.19	Ireland
England_MN	European_MN_Chalc	Banbury Lane (Northampton)	52.22	-0.94	England
England_MN	European_MN_Chalc	Banbury Lane (Northampton)	52.22	-0.94	England
England_MN	European_MN_Chalc	Banbury Lane (Northampton)	52.22	-0.94	England