Thèse de doctorat

de l'Université Sorbonne Paris Cité

Préparée à l'Université Paris Diderot

Ecole doctorale Bio Sorbonne Paris Cité

Equipe Epigénome et Paléogénome

UMR 7592 – Institut Jacques Monod

Paleogenomics of human population dynamics on the French territory between 7000 and 2000 before present

Par Samantha Brunel

Thèse de doctorat de Génétique

Dirigée par Eva-Maria Geigl

Présentée et soutenue publiquement à Paris le 14 novembre 2018

Président du jury :	Reiner Veitia	Professeur	Université Paris Diderot - CNRS
Rapporteurs :	Christine Keyser	Professeur	Université de Strasbourg - CNRS
	Wolfgang Haak	Docteur	Max Planck Institute for the Science of Human History
Examinateurs :	Etienne Patin	CR HDR	Institut Pasteur - CNRS
Directeur de thèse :	Eva-Maria Geigl	DR HDR	Université Paris Diderot - CNRS
Co-directeur de thèse :	Thierry Grange	DR HDR	Université Paris Diderot – CNRS
Membre invité :	Mélanie Pruvost	CR	Université de Bordeaux - CNRS

Aknowledgments

Difficile de résumer en quelques mots ces trois années de thèse tant elles furent hétéroclites. Riche de découvertes et de rencontres, mais aussi d'enseignements, cette expérience a été l'aboutissement d'un cursus universitaire orienté vers l'évolution de l'être humain et de son génome, bien qu'elle ne soit pas une fin en soi. Je suis reconnaissante envers mes directeurs de thèse, Eva-Maria Geigl et Thierry Grange, ainsi que Mélanie Pruvost et tous les collaborateurs du projet ANR ANCESTRA, pour avoir rendu ce travail possible.

Mes remerciements vont également :

A chaque membre du jury, et particulièrement à Christine Keyser et Wolfgang Haak, pour avoir accepté d'endosser le rôle de rapporteurs.

A Etienne Patin et Fabien Fauchereau, membres du comité de thèse, pour leur bienveillance et leurs conseils tout au long de ce périple.

Aux anciens du laboratoire Epigénome et Paléogénome : Mélanie Pruvost, les désormais docteurs Diyendo Massilani, Giulia Lizzo, Nathalie Coté et Olivier Gorgé, des présences adorables et des chercheurs talentueux, Silvia Guimaraes et Andrew Bennett. Une pensée également à ceux qui restent, en particulier Wejdene Ben Dhafer, à qui je souhaite force de caractère pour les trois années à venir.

Aux rencontres Monodiennes : Estelle Gauquelin, Joseph d'Alessandro, Alexandros Glentis, Marc-Antoine Fardin, ainsi qu'aux habitués du coin détente du 5ème étage, pour leur bonne humeur contagieuse, les pauses café salvatrices, les soirées et une certaine virée en Savoie.

A Cécile Gaston, colocataire et complice au soutien indéfectible.

A Nicolas Coënt, Jean-Baptiste Gelé, Louise Brasseur, Ansgar Bohmann, Simon Bouloy et Maëva Mehaye.

Au binôme d'Apprenties Chercheuses Kimberly Plancy et Maryse Yapo, ainsi qu'à Chaïma Benaksas, pour m'avoir permis de faire ce qui me plaît le plus : partager un peu de connaissances et beaucoup de curiosité.

Aux membres de l'association Science Ouverte de Drancy, en particulier Pauline Drapeau, Julien Rastegar et François Gaudel, pour cette mission formidable qu'est la vôtre. Ces nombreux samedis après-midi de tutorat à vos côtés ont été une bouffée d'air frais, et j'ai pris grand plaisir à participer à vos conférences.

Au professeur François Guérineau de l'Université de Picardie Jules Verne, sans qui mon parcours universitaire aurait probablement été tout autre.

A mes parents pour leur confiance et leur présence discrète mais essentielle ; à mes grands-parents, à qui je dois sans aucun doute la persévérance et la résilience qui m'ont accompagnée durant toute cette entreprise.

Et enfin, à ma petite sœur Sarah.

Abstract

Paleogenomics of human population dynamics on the French territory between 7000 and 2000 before present.

The last 10,000 years in Western Eurasia were marked by cultural transitions that profoundly transformed human societies: the advent of the Neolithic, the Bronze Age and the Iron Age. Paleogenomics, the analysis of ancient genomes, started to address the underlying demographic processes in various parts of the continent. In France, however, Late Prehistory is only known from the rich archaeological records and not yet explored through genetics at a territory-wide scale. We generated a large dataset comprising the complete mitochondrial genomes, Y chromosome markers and genotypes on a number of nuclear loci of interest obtained through a DNA enrichment approach of 193 Mesolithic, Neolithic, Bronze Age and Iron Age individuals sampled across the territory of present-day France. It was complemented with the low-coverage genomes of 58 individuals partially overlapping this dataset. This panel provides, for the first time, a high-resolution 5,000-year transect of the dynamics of maternal and paternal lineages in France as well as of autosomal genotypes. Both parental lineages and genomic data revealed different dynamics in the North and the South of the French territory during the Neolithic, with varying degrees of incorporation of autochthonous hunter-gatherers lineages into farming communities. They also revealed a mostly male-driven gene flow from individuals deriving part of their ancestry from the Pontic Steppe at the onset of the Bronze Age, a signature that then persisted through the Iron Age. The various nuclear phenotypic markers we studied evolved differently. While some harbor present-day European frequencies already at the Neolithic epoch indicating ancient episodes of positive selection of these specific traits, others show different evolutionary stages throughout the Neolithic and the Bronze Age allowing us the establish more clearly the origin and evolution of the phenotypic traits that characterize the present-day European population. This study further expands our understanding of the relationship between populations during late Prehistory in France and across Europe.

Keywords: ancient DNA, mitochondrial DNA, ancient genomes, Neolithic, Bronze Age, Iron Age, France

Résumé

Paléogénomique des dynamiques des populations humaines sur le territoire français entre 7000 et 2000 avant le présent.

Ces derniers 10000 ans en Eurasie occidentale ont été marqués par des transitions culturelles qui ont profondément transformé les sociétés humaines : l'apparition et la diffusion du Néolithique, de l'Âge du Bronze et de l'Âge du Fer. La paléogénomique, en analysant les génomes anciens, s'est attelée à en décrire les processus démographiques sous-jacents dans diverses parties du continent. En France cependant, la fin de la Préhistoire est seulement connue par le biais de l'archéologie, et n'a pas encore été explorée par la génétique à l'échelle du territoire. Nous avons produit un large jeu de données comprenant les génomes mitochondriaux, marqueurs du chromosome Y et génotypes d'une sélection de loci nucléaires d'intérêt via une procédure d'enrichissement pour 193 individus datant du Mésolithique, Néolithique, Âge du Bronze et Âge du Fer à travers le territoire de la France actuelle. Nous avons également généré les génomes à faible couverture de 58 individus répartis sur les mêmes périodes et recouvrant partiellement ce panel. L'intégralité de ces résultats offre, pour la première fois, un aperçu des dynamiques des lignées maternelles et paternelles ainsi que du génome nucléaire sur une période recouvrant 5000 ans. Que ce soient les lignées parentales ou le génome, différentes dynamiques apparaissent entre le nord et le sud de la France durant le Néolithique, avec un degré variable d'incorporation des populations de chasseurs-cueilleurs autochtones dans les communautés de fermiers. Ils mettent également en évidence, peu avant le début de l'Âge du Bronze, un flux de gènes dominé par des hommes dont la signature génétique des bergers de la Steppe Pontique, une signature qui ensuite persiste durant l'Âge du Fer, alors que la population montre peu de différentiation à l'échelle du territoire français. Certains marqueurs phénotypiques observés au Néolithique arborent une fréquence proche de celle observée dans la population européenne actuelle, indiquant des épisodes de sélection positive pré-datant le Néolithique, tandis que d'autres montrent des fréquences différentes, signe d'une sélection en cours sur ces loci. Cette étude accroit notre compréhension de relations entre les différentes populations de la fin de la Préhistoire, à l'échelle de la France et de l'Europe.

Mots-clés : ADN ancien, ADN mitochondrial, génomes anciens, Néolithique, Age du Bronze, Age du Fer, France

Table of contents

Table of figures

Table of tables

Part I. Introduction

Chapter I. Modern populations genetics

I. Variations in human populations

The world around us is a world of diversity of shapes and forms, with 1.5 million species identified and nearly a thousand times more yet to be observed (Larsen *et al.*, 2017). As humans, we are the results of a complex and seemingly ancient history that spans around 10 million years, from the moment our lineage diverged from that of present-day chimpanzees (Patterson *et al.*, 2006). Yet the 7.6 billion of us worldwide today are no exceptions to the rule of diversity, as humans indeed are a complex species that displays a wide range of features, some being remarkable while others remain invisible to the eye. These characters, or *phenotypes,* called for the establishment of categories and classifications among humans. Hence the most obvious traits such as skin and hair color, hair texture, head morphology, set the basis to concepts such as races that, stripped off their objective purposes, served to legitimate judgment of value among groups. The criteria used for classification evolved with the scientific knowledge available at the time, but always had to fulfill two main conditions: be hereditary and have a certain stability within a population from one generation to the other (Jacquard, 1978).

The existence of molecular genetic variation among human populations was first observed through the blood system by Karl Landsteiner in 1900, a discovery for which he was awarded the Nobel Prize in Physiology or Medicine in 1930. He found that when mixing together blood samples recovered from different individuals, they would sometimes agglutinate. Through his experiments, he identified the three main blood groups we are all familiar with today. The genetic inheritance of these groups was later demonstrated by Hirszfeld and Hirszfeld in a classic study published in 1910, resulting in the identification of the first human gene to be described: the ABO locus, which determines the presence of one, both, or neither of the A and B antigens on red blood cells (erythrocytes), defining ABO blood groups (Hirszfeld and Hirszfeldowa, 1919). In 1954, Arthur Mourant published *The distribution of the human blood groups*, a groundbreaking book in which he summarized allele frequencies for serological variation in various populations, as blood-group data was being abundantly produced in the context of blood

transfusion, and one of the very few human traits with clear genetic inheritance at the time. This large-scale study revealed that detectable genetic differences exist among human populations, as the frequency of the different groups varied across the world (Mourant, 1954).

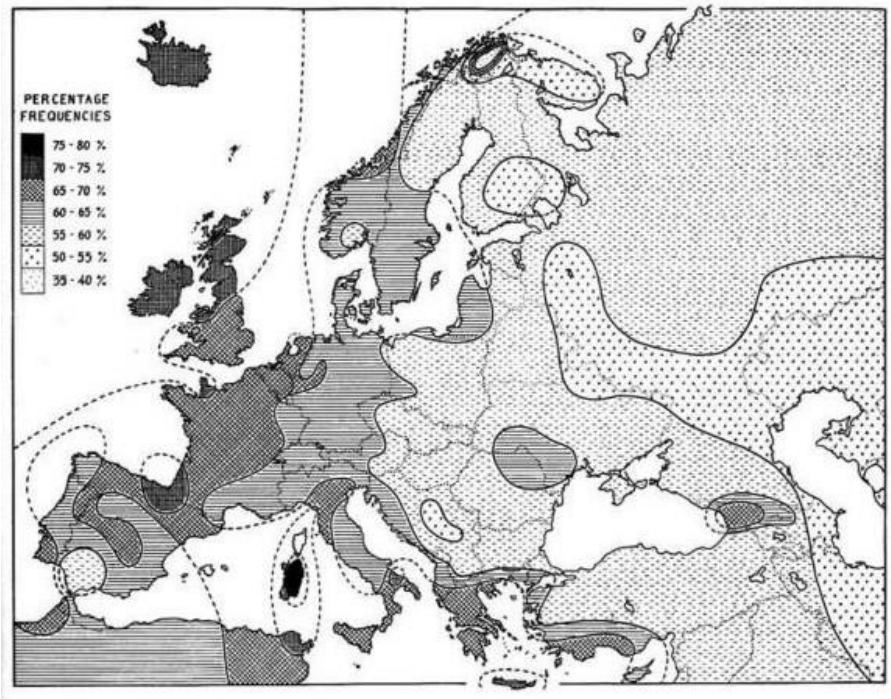


Figure 1: Frequency of the blood group O in Europe.
(Mourant, 1954)

As more and more genetic markers were discovered and studied within families, it soon became apparent that combinations of alleles at different genetic loci could segregate together from one generation to the next. This concept of haplotypes was introduced in a study of the major histocompatibility complex, where segregation of combinations of alleles was observed by Ceppellini and colleagues in 1967 (Ceppellini *et al.*, 1967). Altogether, these stepping stones contributed to our understanding of the mechanisms underlying both our similarity and our diversity, as analyzing the genotypes of various loci on a chromosome in populations and their inheritance within families could help us study humans and their evolution (Race *et al.* 1975).

In this introduction, I chose to present the milestones that ultimately shaped modern genetics, from the discovery of the DNA molecule to the theoretical concepts that provided the framework within which individuals and populations are studied today. Focusing on Western Eurasia, we will see how genetics aimed at answering questions

about when and how the European continent was populated, and what the origins of the first Europeans were. Finally, we will show how individuals from past populations, through multiple approaches involving specialists from a wide range of disciplines, ultimately allowed us to delve into our past. The outcome of these various studies will be put into perspective in order to discuss the results newly reported in this work.

1. From phenotypes to genotypes

As the first theories about the causes of phenotypic variations appeared, the underlying mechanisms of the transmission of characters across generations were still obscure. Hence the concept of *pangenesis* from Charles Darwin in 1868, stating that every part of the adult body emitted small organic particles called *gemmules* that migrated to the gonads, contributing heritable information to the gametes (Darwin, 1868), or Buffon's *seminal liquor,* a fluid "composed of parts which seek to be organized; that it in fact produces organized substances, but that they are not as yet either animals or organized substances, like the individual which produced them" (Buffon *et al.*, 1767). The rediscovery of Gregor Mendel's work in 1900 (Mendel G., 1866), 45 years after their first publication, set the basis for genetics by introducing concepts now known as *genes* and *alleles* and their transmission across generations. The identification of DNA as the support of genetic information over the course of the 20th century (Levene, 1919) and the description of its structure in the 1950s marked the beginning of the genetic era as we know it (Chargaff, 1950; Watson and Crick, 1953).

2. Internal forces altering the DNA sequence

As I briefly mentioned in the previous paragraph, the pioneer work of Phoebus Levene during the 20th century resulted in the description of the monomers of DNA, i.e. *nucleotides*, and their organization into the DNA sequence (Levene, 1919). Genetic variation is caused by changes that alter this sequence and are called *mutations*. Mutations occur spontaneously within cells, through non-repaired errors in DNA replication, or DNA damages induced by byproducts of the normal cell metabolism such as reactive oxygen species (Loeb, 1989). They may also be caused by external factors that alter the DNA sequence, such as ultraviolet light, X rays, or certain chemicals (Motulsky, 1984). There are many different types of mutations, the most commonly used for many

purposes being single nucleotide polymorphisms (SNPs), changes that only alter one nucleotide, replacing it with another. These mutations can only be identified through sequencing, when the DNA sequence is aligned with another sequence that serves as a reference. Every time a genome is passed from one generation to the next it accumulates 100–200 new mutations, according to a DNA-sequencing analysis of the Y chromosome (Xue *et al.*, 2009). Other genomic forces are then acting upon these variations. During meiosis, the cell divisions that lead to the production of gametes, newly duplicated chromosomes exchange parts through recombination, a process often mediated by homology. This recombination results in a new arrangement of maternal and paternal alleles on the same chromosome. Even if the order of genes along the chromosome does not change when this process occurs properly, the combination of parental alleles can be altered. The frequency of these crossovers varies across chromosomes, leading to the notion of "genetic linkage" between alleles to describe the tendency of alleles to be inherited together as a result of their location on the same chromosome. The stronger the so-called *linkage disequilibrium*, the more likely alleles at each of two positions will be found in association with one another. Since such loci are no longer considered as independent, description of human variation requires more than just the positions and characteristics of isolated SNPs. One rather needs to study how these SNPs vary simultaneously within a part of or a whole chromosome, which requires statistics of the correlations between the variations at different positions.

3. Other forces applying on existing variation

Allele frequencies can be estimated and compared among populations, revealing patterns of changes. These changes result from different forces that either affects single loci (evolutionary forces) or the genome as a whole (demographic forces).

(1) Genetic drift

Mutations set aside; each new generation draws a part of the genetic variability that existed in the previous one, in the form of two gametes picked at random. As this phenomenon occurs at each generation, any given allele has a chance to either disappear, or ultimately reach fixation in the population. The progressive modification of the genetic structure of a population solely by chance is called random genetic drift, and is due to a

finite number of individuals participating in the formation of the next generation, ultimately altering the frequency of an existing variant in the population (Kimura, 1983).

This process is slow in a large effective population, but its effect increases when a population goes through a bottleneck after a natural disaster or epidemic. The resulting loss of genetic diversity can persist long after the cause of the population decline disappeared, and reduces the potential of adaptation of the population. The reduction of genetic diversity that occurs when populations split into groups that afterwards remain separate characterizes a founder effect, a concept named after the term "founder", first introduced by Ernst Mayr in Systematics and *the Origin of Species from the Viewpoint of a Zoologist* to describe individuals starting a new population after a split from their parent group (Mayr, 1942). Such episodes are quite common in our history, as our species was born in Africa and every human population today was founded by a reduced number of individuals that originally stemmed from this initial group, carrying only a subset of its genetic diversity (Stringer and Andrews, 1988). As a consequence of the expansion of humans around the world through a series of founder effects, genetic variation in human populations decreases with the geographic distance to the source population in Africa (Ramachandran *et al.*, 2005).

(2) Migrations and gene flows

The study of genetic drift relies on one main assumption: the isolation of the group during the period that encompassed the studied generations. Yet such isolation over a long period of time is rarely encountered in nature, especially among humans.

Only a little more than a hundred tribes across the world still maintain limited contacts with other people according to Survival International, even though many more are probably still unknown (https://www.survivalinternational.org/uncontactedtribes). These tribes often live in regions that are difficult to reach, such as the deep interior of the Amazon, the Congo, and the mountains of New Guinea, but also islands off India such as the Andaman Islands, and despite being described as *uncontacted* people, they do in fact often have a history of contacts. Apart from these exceptions, there is no such isolation among most populations across the globe today, as technology enabled people to move relatively quickly and easily from one part of the world to the other. Increased contact between people despite geographical distance and cultural differences provide more opportunities to pass on genes from one population to another. As such gene flows

introduce new sets of alleles in the genetic pool of a population, only a low amount of migration (i.e. few individuals) is enough to counter the effect of genetic drift. The more geographically and genetically distant each parental population is, the stronger the effect.

(3) Natural selection

The concept of natural selection was formalized by Charles Darwin in his *On the Origin of Species* (Darwin, 1859) as follows: "If variations useful to any organic being do occur, assuredly individuals thus characterized will have the best chance of being preserved in the struggle for life; and from the strong principle of inheritance they will tend to produce offspring similarly characterized. This principle of preservation, I have called, for the sake of brevity, Natural Selection". This concept describes the differential survival and reproduction of individuals based on differences in their respective phenotypes, hence genotypes. Where one or several characters might be advantageous in a given environment and at a certain time, they might be preferably transmitted to the next generation as a result of a higher reproductive success, thus increasing in frequency in the population over time. Disadvantageous traits are removed from the population, as their carriers have fewer offspring each generation, reducing the frequency of the trait and its associated genotype in the population.

II. Signatures of past events in present-day people

The genetic variability of modern populations is not randomly distributed. Instead, it displays patterns. Within Europe for example, the first two principal axes of variation of the matrix of genotypes over about half a million genomic loci closely match latitude and longitude upon rotation of the axes (**Figure 2**) (Lao *et al.*, 2008; Novembre *et al.*, 2008), as would be expected if patterns of ancestry in present-day populations were mostly shaped by local migration. To the question "where does the genetic variation observed among modern Europeans come from?", an intuitive answer would be that it must reflect some combination of demographic phenomena that occurred in its past, since Europe has been continuously peopled since the arrival of Homo sapiens around 45,000 years ago (Benazzi *et al.*, 2011; Higham *et al.*, 2011; Trinkaus *et al.*, 2003). These phenomena would have acted upon both the genetic variation that accumulated before Europe was

colonized, as the first settlers had pre-existing genetic variation, but also the genetic variation that accumulated since then.

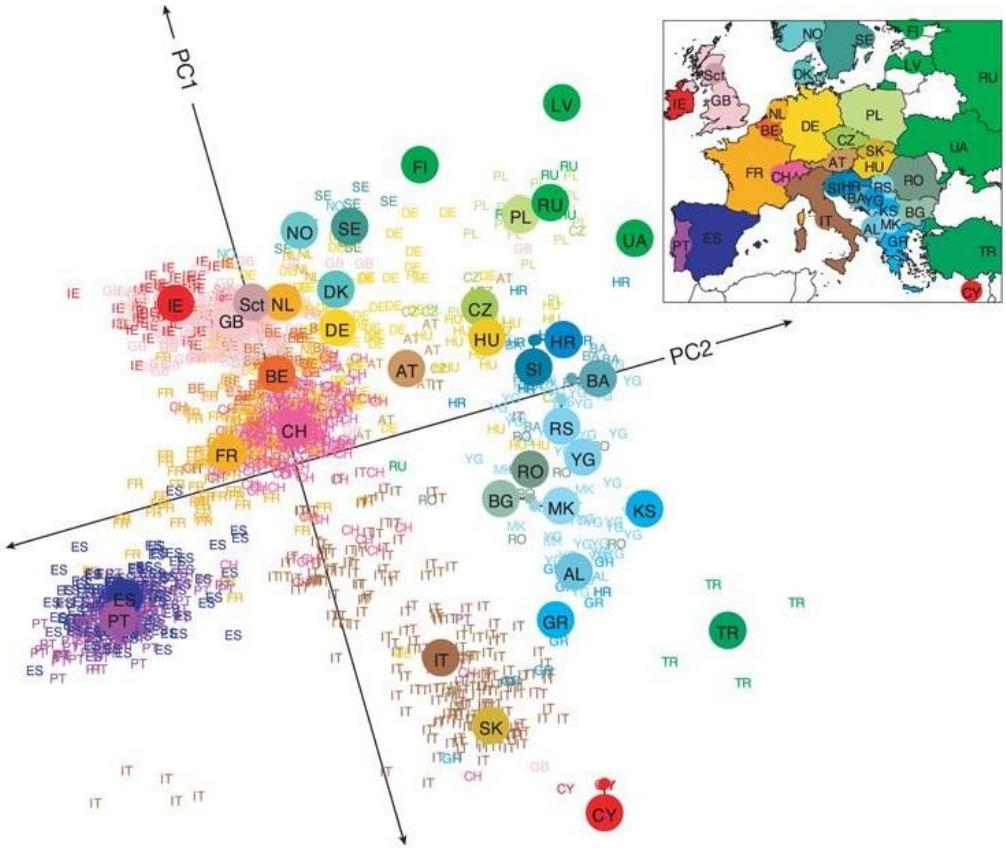


Figure 2: A statistical summary of genetic data from 1,387 Europeans based on principal component axis one (PC1) and axis two (PC2).
Small colored labels represent individuals and large colored points represent median PC1 and PC2 values for each country. The inset map provides a key to the labels. The PC axes are rotated to emphasize the similarity to the geographic map of Europe. (Novembre *et al.* 2008)

Indeed, the history of Europe is a history of migrations, sometimes at a very large scale. Thus, a common way to gain insight into these past population processes is to analyze and interpret patterns of genetic variation in present-day people. Few of us know our family histories more than a few generations back; it is therefore easy to overlook the fact that we are all distant cousins, related to one another via a vast network of relationships. Although most genetic relationship among individuals world-wide is very old, some individuals are related at far shorter time scales. Given that each individual has 2^n ancestors from n generations ago, theoretical considerations suggest that all humans are related genealogically to each other over surprisingly short time scales (Chang *et al.*, 1999; Rohde *et al.*, 2004).

Chapter II. From genetics to paleogenomics

I. Genetics of modern populations

1. Classical markers

An ever-increasing dataset of individual and population-wide biological information and developments in the mathematical framework associated with population genetics profoundly transformed the methodology of the field over the past century. Before the advent of DNA-based testing, proteins and blood groups constituted what was known as *classical markers*, and attempts were made at correlating their allelic frequencies in modern populations with spatial data. In Europe, clines in many of these genetic markers were therefore identified, spanning from southeast to northwest Europe (**Figure 3**). The correlation between the allele frequencies in those clines and the dates of origin of agriculture inferred from the archaeological record led Ammerman and Cavalli-Sforza to propose that the European genetic structure was determined mainly by population dispersal during the Neolithic: a *demic* diffusion (Ammerman and Cavalli-Sforza, 1984). In this model, Neolithic farmers expanded from the Near East into Europe, presumably as a result of population growth, spreading their genes along with novel technologies for farming and animal breeding over the entire continent (Sokal and Menozzi, 1982), and mixed very little with the preexisting hunting–gathering communities. In this scenario, the genes of the latter should represent only a small fraction of the present European gene pool. This contradicts the hypothesis broadly adopted among archaeologists that advocated in favor of the development of agriculture in Europe or, alternatively, the adoption of plants, animals and related technologies from the Near East with only limited and geographically restricted human migration. Indeed, computer simulations have shown that a Neolithic expansion is the simplest, but not the only, possible cause of the continent-wide clines of allele frequencies. Paleolithic population dispersal may have been accompanied by founder effects, resulting in clines under successive short-distance gene flow (Barbujani *et al.*, 1995). Thus, the current European gene pool could have been established during the Paleolithic, with relatively few Near Eastern genes incorporated during the Neolithic, and the farming technologies would mostly have spread through cultural transmission, rather than by population movements.

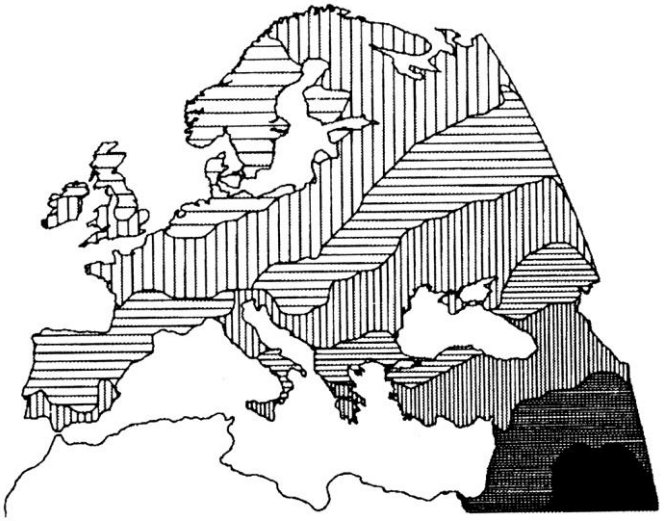


Figure 3: A summary of genetic variation in Europe's first principal component
Different shades of gray represent different values of a synthetic variable summarizing allele frequency at 120 protein loci. (Cavalli-Sforza, L. L., *et al.*, The History and Geography of Human Genes)

2. DNA markers

Methodological developments such as the polymerase chain reaction (PCR) in the 1980s coupled with Sanger sequencing in 1977 introduced the use of DNA markers to population genetics. It became possible to simultaneously study various genomic loci within theoretical frameworks that allowed us to introduce the notion of evolutionary distance between alleles, and study the genetic relationships among human populations, as well as their relatedness with other species. In this paragraph, I will introduce the different types of DNA found within a cell, and how they can be used to study relatedness among individuals and populations.

Mitochondrial DNA (mtDNA) is located within mitochondria, cellular organelles that are involved in the energetic metabolism, converting organic material into a form that cells can use, adenosine triphosphate (ATP). The number of these organelles varies among cells and cell types, but is usually comprised between several hundreds and one or two thousand. In humans, like in most species, mitochondrial DNA is transmitted exclusively from mothers to offspring, with almost no recombination events. It is a 16,569bp-long circular DNA molecule that codes for a total of 37 genes. These genes are not evenly distributed across the mitochondrial genome, which is divided into both a coding and non-coding region. The latter contains two specific "hypervariable" regions (HVR-I and HVR-II), that accumulate mutations at a faster rate, rendering it particularly useful for genotyping individuals. The mitochondrial genome as a whole accumulates

mutations, though at a slower pace. Combinations of these mutations define haplotypes, and when extended across the whole mitochondrial genome, fit into a broader haplogroup that characterizes a specific maternal lineage. All these mutations are defined through comparison with a reference sequence: the Cambridge revised sequence (rCRS, Anderson *et al.* 1981, Andrews *et al.* 1999).

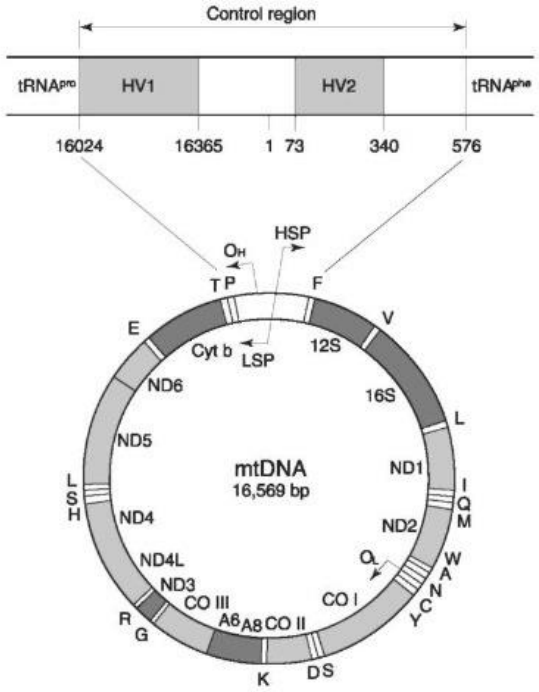


Figure 4: Structure of the mitochondrial DNA
(Parsons *et al.* 1999)

Nuclear DNA is composed of 22 pairs of autosomes numbered from 1 to 22 and one pair of allosomes (or sexual chromosomes) X and Y. The Y chromosome is found only in males, and is transmitted from fathers to sons only. It carries small regions that are homologous to the X chromosome (pseudoautosomal regions, PAR 1 and PAR 2) and therefore suitable for homologous recombination during meiosis (**Figure 4**). It also contains a male specific region called "non-recombining region of the Y chromosome" (NRY), which forms the bulk of the chromosome and like the mitochondrial DNA lacks the possibility to recombine. The mutations that accumulate over this large portion of the chromosome also define haplogroups that allow tracing of paternal lineages (Underhill *et al.*, 2000).

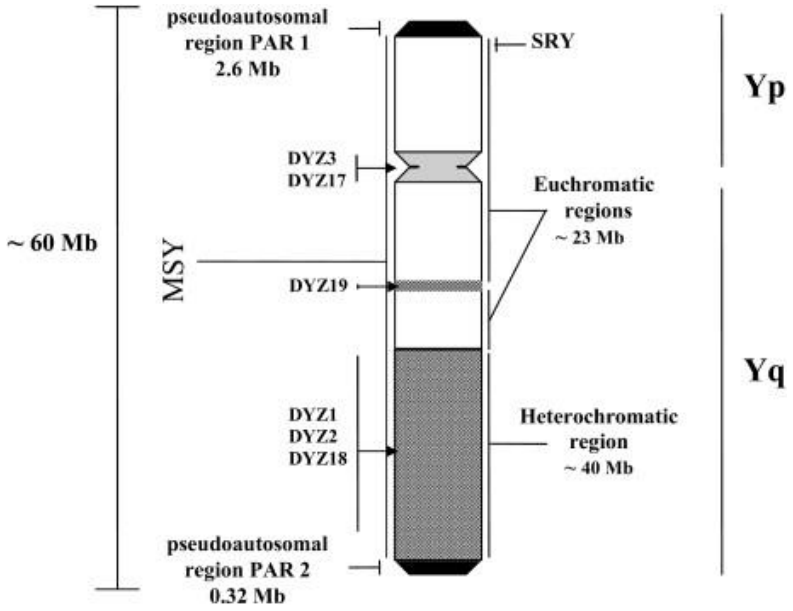


Figure 5: Structure of the Y chromosome
(Gusmão *et al.* 2008)

(1) Results from uni-parentally inherited markers

The first widely used molecular markers were variants of the mitochondrial DNA (mtDNA) and the NRY. Due to their peculiar mode of transmission, these markers are transmitted from one generation to the next intact (apart from new mutations), as they have no homologous partner to recombine with. As their mutation rate is known from the study of pedigrees, these markers allow a straightforward phylogeny construction and inferences of some aspects of population relationships, as they provide a uniquely male or female perspective of human evolutionary history (Pakendorf and Stoneking, 2005). Understanding the evolutionary path of both the female and male lineages has helped population geneticists trace their inheritance in modern humans back to Africa and their subsequent spread around the globe (Cann *et al.*, 1987). Due to their sex-specific nature, investigating both mtDNA and the NRY in parallel to data from the recombining autosomes remains important as they provide insights into social and demographic parameters, such as sex-biased introgression or mobility, which are crucial for the reconstruction of past societies (Haak *et al.*, 2008).

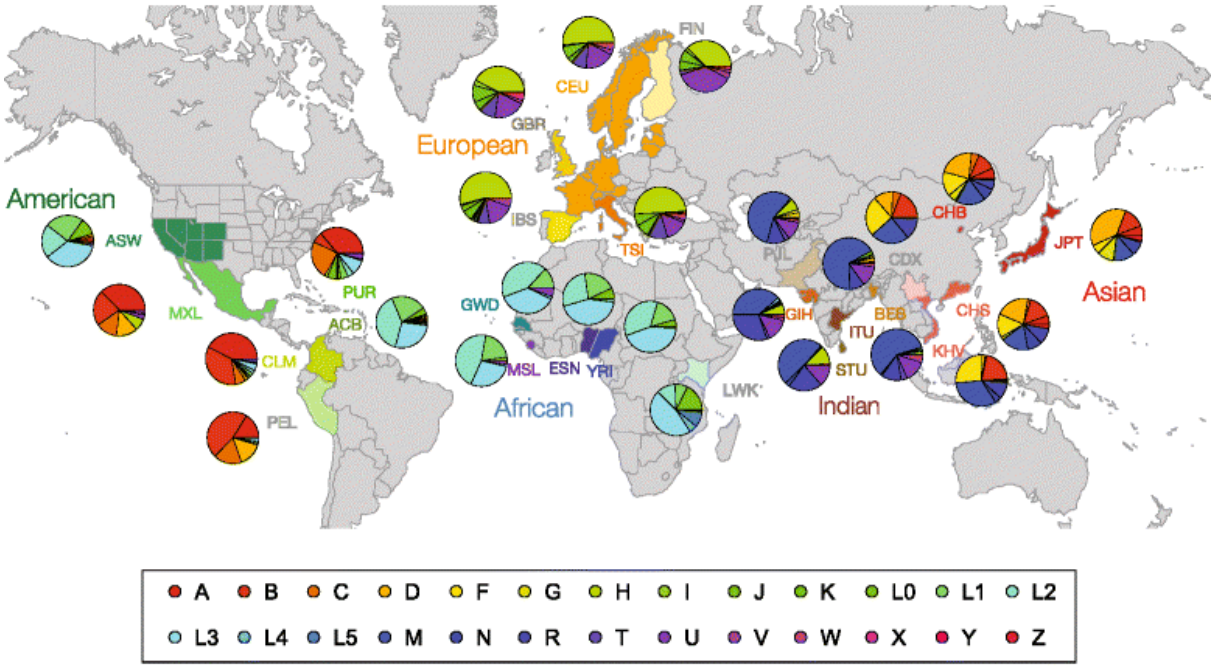


Figure 6: Global distribution of mtDNA haplogroups.
Map showing the names and locations of worldwide populations from the 1000 genomes project along with pie charts showing the relative frequencies of mtDNA haplogroups for each population. The haplogroups are color coded as shown in the key. (Lavanya *et al.* 2017)

Studies from uniparentally inherited markers have improved our understanding of human demographic history, as the distribution of mtDNA and Y chromosome haplogroups around the world reflects the early dispersion of humans out of Africa (**Figure 6**). At a more local scale, no significant structuring could be detected from the mtDNA in present-day Europe and Near East, in comparison with the gradients exhibited by classical markers (**Figure 3**) (Menozzi *et al.* 1978; Ammerman and Cavalli-Sforza 1984; Sokal *et al.* 1989, 1991; Cavalli-Sforza *et al.* 1994). A principal-component analysis of the hypervariable segment I (HVS-I) by Cavalli-Sforza and Minch (1997) indicated that the main detectable pattern was a shallow east-west gradient that accounted for only 23% of the variation. These studies brought out the limitation of this approach on mtDNA to infer the demographic history of Europeans, as selection and maternal gene flow across the continent potentially blurred ancient migration events (Barbujani and Chikhi, 2006).

Other approaches were used to deduce the patterns of the pioneer human peopling from the extant mtDNA diversity. One example was the calculation of the coalescence age of mtDNA haplogroups identified as being autochthonous from a given region. From the mtDNA variability of sampled individuals in the present-day population, and assuming

that no selection or gene flow had occurred, the age of the common ancestor to these lineages can be estimated. The *founder analysis* of mtDNA in Europe, undertaken by Richards and colleagues in 2000, was based around the idea that this coalescent age would reflect the time of the colonization of the region (Richards *et al.*, 2000). In this study, a founder type was defined as an ancestral node in the mitochondrial phylogenetic tree, that was either present in both the source and the destination area, or reconstructed from extant haplogroups (Richards *et al.*, 2000; Stoneking *et al.*, 1990). Through the fine tuning of this method and the extension of the Near Eastern, European, and Northern-Caucasus mitogenome databases, they observed that the largest fraction of extant mtDNA lineages date back to the reexpansion that followed the Last Glacial Maximum, 20,000 years ago, and that the maternal contribution of Near Eastern Neolithic farmers to the modern mtDNA pool did not exceed 20%. These results support the colonization of Europe by small groups of Near Eastern people combined with a wide-scale adoption of the Neolithic lifestyle by autochthonous Mesolithic populations (Richards *et al.* 2000). However, the coalescent time of two mitochondrial sequences drawn from both the source and the destination population is expected to be older than the split of the two populations (Barbujani *et al.*, 1998). Indeed, unless the colonizing group passes through a strong and long-lasting bottleneck (Nei *et al.*, 1975), it will harbor some degree of genetic diversity prior to migrating into the destination area. Under such scenario, the estimated age of founder types will predate the time of the migration.

An initial Y-chromosome analysis reached a similar conclusion (Semino, 2000), and the proposed scale of immigration was further reduced. A number of distinct episodes, including (i) migration from Northeast Africa, (ii) expansion of acculturated indigenous hunter-gatherers in the Balkans, or (iii) younger expansions, were proposed (Battaglia *et al.*, 2009). Studies of Western European mtDNA and Y-chromosome variation supported these conclusions, with a largely Mesolithic ancestry in various regions of Europe including the Basque country, the Iberian and Scandinavian Peninsulas and the British Isles (Luca Cavalli-Sforza *et al.*, 1995). These regions may still have received Neolithic immigrants, who integrated maternal lineages of autochthonous hunter-gatherers.

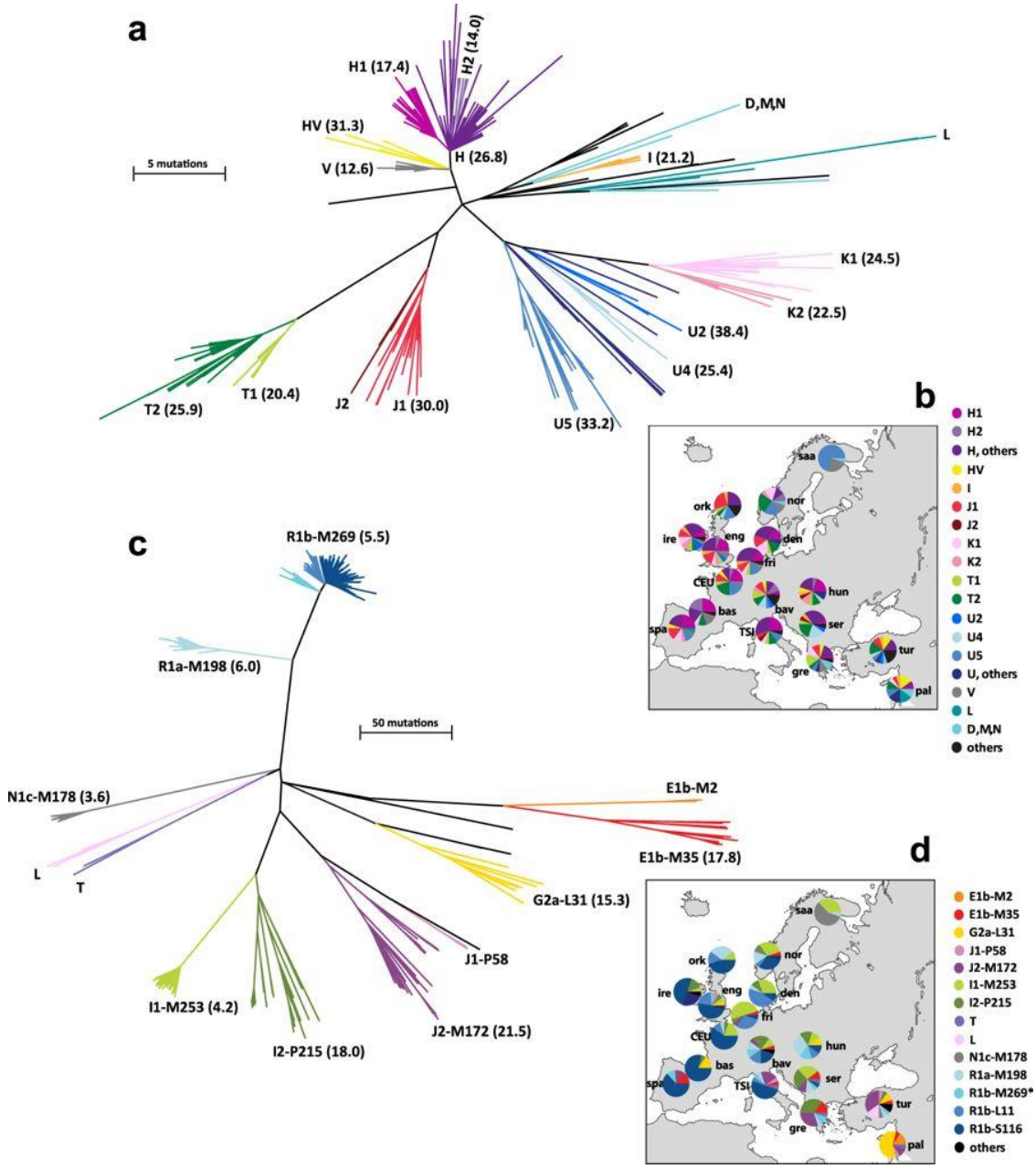


Figure 7: Phylogenies and geographical distributions of European mtDNA and Y chromosome lineages.
(Batini *et al.* 2017)

From the mtDNA and Y-chromosome it was concluded that: (i) the Neolithic most likely dispersed into Europe through human migration, accompanied by a spread of domestic plants and animals alongside the migrants, (ii) the immigration from the Near East was minor, and there was substantial adoption of farming by indigenous groups in many parts of Europe and (iii) internal European migrations that occurred after the Neolithic may have later considerably reshaped the genetic landscape. Thus, many mitochondrial lineages whose origin has been traced back to the Paleolithic period

probably reached Europe at a later time. Looking at three major Y chromosome lineages (I1, R1a, R1b) that represent more than 60% of present-day Europe's Y diversity, Batini and colleagues showed that they have a short coalescent time (3,500 to 7,300 years ago), indicating a widespread male-specific phenomenon during the onset of Bronze Age, contrasting with mitochondrial data (**Figure 7**)(Batini *et al.*, 2015).

(2) Results from autosomal markers

The study of genome-wide markers was initiated using microsatellites (short tandem repeats). Chikhi and colleagues used four of these markers in combination with other hypervariable genomic loci to study genomic variation across Europe. Their study highlighted broad clinal patterns of DNA variation that closely matched those described at the protein level. To the exception of the Saami, a semi-nomadic population from Northern Europe, all European populations were found to have split within the last 10,000 years. This pattern was consistent with the westward and northward expansion of a Neolithic population originating in the Near East, thus supporting a Near Eastern origin for the ancestral population of present-day Europeans.

The development of single nucleotide polymorphism (SNP) arrays further increased the resolution of genome-wide studies, as they are less prone to drift and providing insight further back into human history. I mentioned earlier how genes mirrored geography in Europe (Lao *et al.*, 2008; Novembre *et al.*, 2008). The study from Lao and colleagues revealed that, although the amount of differentiation within the European autosomal gene pool is small, the existing genetic differences correlated well with geographic distances, with a slight decrease in diversity running from south-to-north in Europe (**Figure 8**), with the highest haplotype and allelic diversity in the Iberian Peninsula, and the lowest haplotype diversity in England and Ireland (O'Dushlaine *et al.*, 2010). Results from these studies were compatible with expectations based on European population history, mainly the prehistoric population expansion from southern to northern Europe and/or a larger effective population size in the south as compared to the north of Europe.

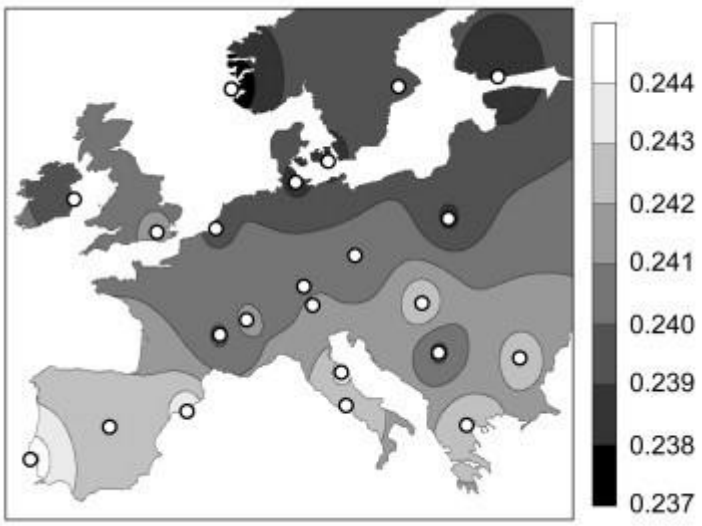


Figure 8: Isoline map of Europe based on the mean observed heterozygosity in each of 23 European subpopulations.
(Lao *et al.* 2008)

3. Limitations of the genetic approach in modern populations

When summarizing the results of the previously described studies, contradictions arise. While comparisons between archaeological findings and allele frequencies at protein loci suggest that most genes of current Europeans descend from populations that have been expanding in Europe in the last 10,000 years starting during the Neolithic period, recent mitochondrial data have been interpreted as indicating a much older, Paleolithic ancestry.

One has to bear in mind that when studying past events from variation in present-day populations, only the ancestral lineages of living subjects are available for inferences, thus lineages that became extinct in the past are usually not represented. Moreover, when considering nuclear data, the contribution of each of our ancestors to our nuclear genome becomes more diluted with every new generation, to the point where it is barely informative, leading to a progressive reduction of information the further back in time we go. Even if we focus on the last 50,000 years or so, it may seem that traditional genetics applied to modern populations can capture little of the dynamics of the growth and development of human communities. Most studies reflect current population structure and admixture during the last 3,000 years, but provide little insight into older population processes. Such insights can only be obtained through the analysis of ancient genomes.

II. A new discipline: paleogenomics

1. Background

The term "ancient DNA" (aDNA) refers to DNA molecules recovered from archaeological or paleontological samples. These samples are very diverse, ranging from tissues and bones, teeth, seeds and pollen for plant samples, coprolites and sediments (Haile, 2012; Loreille *et al.*, 2001; Pääbo *et al.*, 1988; Rollo *et al.*, 1988; Thomas *et al.*, 1989). The discovery that DNA could survive for thousands of years in favorable environmental conditions gave rise to a-new discipline called paleogenetics, a term that was initially invented by Emile Zuckerkandl and the physical chemist Linus Carl Pauling in 1963, when pondering possible applications in the reconstruction of past polypeptide sequences. The first ancient DNA sequence, isolated from a museum specimen of the extinct quagga, was published in 1984 by a team led by Allan Wilson (Higuchi *et al.*, 1984). Paleogeneticists, by piecing together ancient DNA sequences using various analytical methods, aim at providing a temporal dimension to genetic studies that would be inaccessible with present-day genomes alone. In the field of anthropology, the use of aDNA allowed geneticists to elucidate broad patterns of relationship among populations, providing answers to long-standing archeological and paleoanthropological questions. When, where, and from what source did specific human populations arise? Who admixed with whom and when did this admixture take place? Are obvious changes in the archaeological record the result of population replacement or cultural innovation? Did past cultures leave any genetic descendants? As we will show in the next part, analysis of aDNA has been successful in answering several of these questions, but also raised many new ones in the process.

2. Generalities about ancient samples

After an organism is buried and decays, minerals from mineral-rich groundwater invade empty spaces left by the decomposition of organic matter and start accumulating. The degree to which the remains are decayed when covered by sediment determines the later details of the fossil. Most fossils consist only of skeletal remains or teeth; but sometimes they can contain traces of skin, feathers or even soft tissues. Trace amounts of DNA can occasionally survive the decomposition of organic matter for long periods of time after the death of an organism, though it is at least partially degraded and chemically

modified (Hofreiter *et al.*, 2001). Little is known about the exact role the environment plays in DNA preservation. Bones are locally destroyed by bacteria and fungi (Bell *et al.*, 1996), and the diagenetic alteration that characterizes fossilization is a localized process leaving discrete "fossilizing regions" (Wess *et al.*, 2001). Thus, conditions that alter bacterial and fungal growth directly impact the chance for DNA preservation, and these conditions likely vary across and within bones. Temperature has been identified as a key factor in DNA preservation (Smith *et al.*, 2001), but it is likely that all factors influencing chemical reactions (*e.g.*, pH, chemical composition of bone and soil, hydrology...) also play a role. As a result, various types of microenvironments with different biological and physicochemical properties might enable long-term DNA preservation (i.e. so-called "molecular niches"(Geigl, 2002)), and such structures can form in the bones during fossilization. There, DNA might be protected from degradation (*e.g.*, by adsorption to the hydroxyapatite ($Ca_5(PO_4)_3(OH)$) matrix that composes the bone structure (Grunenwald *et al.*, 2014). The excavation of fossils, the washing procedures and other treatments they are subjected to as well as their transfer to museums or natural history collections constitute important changes in the physicochemical conditions of such microenvironments with dramatic consequences on DNA preservation (Pruvost *et al.*, 2007).

3. Signatures of ancient DNA molecules

Ancient DNA molecules are different from modern ones from many respects. The DNA molecule will degrade if not repaired. After death, cells no longer maintain their integrity, leading to the release of free radicals and nucleases from cell compartments that initiate the degradation of the DNA molecule. Cell rupture also releases nutrient-rich fluids, which encourage the growth of environmental microorganisms that contribute to further degradation of organic material. As the complex repair systems are no longer active in the dead cell, damages accumulate quickly, leading to fragmentation of the DNA molecule. If these enzymatic processes are reduced or stopped, traces of these organic molecules can survive. This is the case in exceptional circumstances, including rapid desiccation, freezing and high salt concentrations. It is thus the environmental conditions of preservation rather than the age of the cell that determines the rate of genomic

degradation (Gilbert *et al.*, 2005; Hofreiter *et al.*, 2001; Lindahl, 1993; Sawyer *et al.*, 2012; Willerslev and Cooper, 2005).

Apart from fragmentation, DNA also undergoes chemical changes (**Figure 9**). In particular, the glycosidic bonds that bind the nitrogenous base to the sugar-phosphate backbone are weak points in the structure of DNA. Accordingly, hydrolytic cleavage of the glycosidic bonds in DNA, often referred to as "depurination" as it occurs more easily in purines than in pyrimidines, will lead to the formation of abasic sites. From there, hydrolysis of the deoxyribose will result in the eventual fragmentation of the DNA into progressively smaller molecules with single-stranded overhangs (Gates, 2009).Another characteristic signature of ancient DNA is the conversion of cytosine residues into uracil through hydrolytic deamination, a process that occurs more frequently at the single-stranded ends of the molecules (Hofreiter *et al.*, 2001; Lindahl, 1993). Not only do these changes lead to an excess of C to T transitions in the DNA sequence, but they also hamper molecular approaches as they sometimes inhibit certain proofreading polymerases.

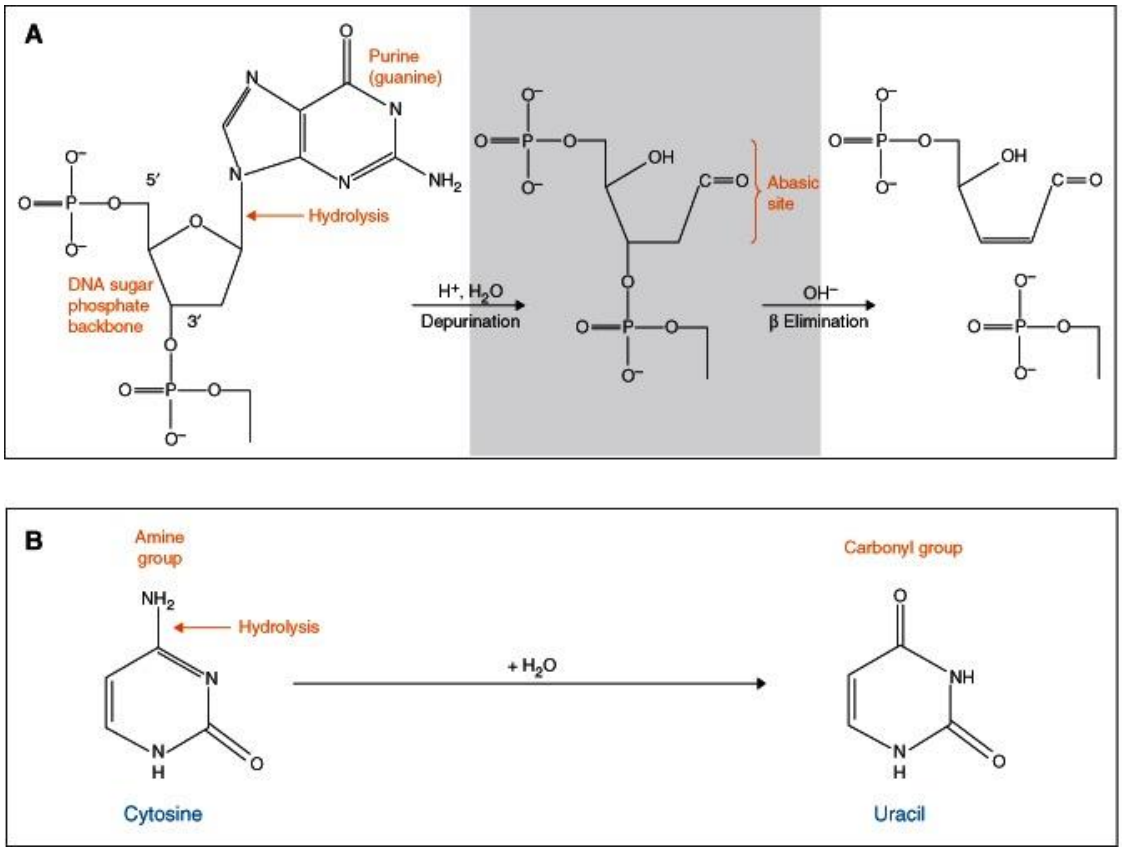


Figure 9: A. Fragmentation and B. deamination occurring in the aDNA molecule.
(Dabney *et al.* 2013)

4. Constraints and limitations

Using ancient DNA to achieve a better understanding of the events that occurred during the European prehistory is quite a recent approach. Despite promising results, several issues still limit the potential of the discipline.

(1) Sample availability

The number of well-documented archaeological specimens available to perform such analysis is limited. Indeed, the number of excavated skeletons dating from certain periods such as the pre-Neolithic is quite low. Archaeological samples can, depending on the temporal or geographical region of interest, be very scarce due to environmental or cultural factors leading to a poor preservation of remains such as funerary practices (*e.g.* cremation). Because of this taphonomic bias, it should not be forgotten that the ancient individuals studied are not necessarily representative of a given ancient population. Funerary practices such as the selection of buried individuals according to their sex, age, and social condition makes it likely that not the entire ancient population was represented within burial sites but only parts of it. The process of extracting DNA from archaeological remains is destructive and often these remains are precious and irreplaceable. Ancient DNA studies thus require close collaboration with archaeologists, anthropologists and paleoanthropologists. Moreover, not all bones are good sources of ancient DNA, in particular spongious bones, and the small amount of material obtained from bones allows only for a limited number of experiments. As a result, the main problem of genetic studies of populations from the past concerns the number and representativeness of the samples. Thus, increasing the data available is essential to reconstruct population histories while accounting for regional and local variability.

(2) DNA degradation

The ability to obtain authentic ancient DNA from archaeological specimens varies according to numerous endogenous and external factors which influence DNA preservation within skeletal remains.

As stated before, DNA undergoes degradation from the moment the cell stops its activity and dies and if the environmental conditions are unfavorable, DNA can end up completely degraded. The number of DNA fragments extracted from an ancient sample is therefore expected to increase exponentially as their size decreases (Allentoft *et al.*,

2012), which contrasts with the observed size distribution of DNA sequences from published ancient specimens, that instead shows a deficit in short sequences. Although this deficit might result from poor DNA preservation, the loss of such molecules during extraction and library construction cannot be excluded (*e.g.*, Dabney *et al*., 2013; Rohland and Hofreiter, 2007). Moreover, successfully retained very short fragments are less informative than long molecules as they can generate spurious alignments during post-sequencing analyses (Smith et al., 1985). The success rate of ancient DNA studies can thus vary from one specimen or one bone collection to another according to various additional parameters such as the age of the skeletal material, the burial circumstances, the storage conditions after excavation and the handling precautions taken during the analyses (Pruvost *et al*., 2007, 2008).

(3) Contamination

One of the major problems that prevented the widespread sequencing of hominin aDNA for several years was contamination. Genetic material extracted and sequenced from a tissue sample of a living individual will consist largely of DNA fragments from that individual (i.e., endogenous fragments) if standard laboratory practices are followed. In contrast, because aDNA is so scarce and fragmented, most of the genetic material extracted from fossils tends to be exogenous, usually mainly from environmental microorganisms. In addition, DNA from humans who handled the fossil can also be co-extracted and analyzed (*e.g.*, Green *et al.*, 2009). The latter type of DNA is especially troublesome, as present-day human DNA is similar in sequence to endogenous aDNA from hominin fossils, and can introduce biases in downstream analyses. Modern DNA, of a better quality, will often be favored by molecular approaches that rely on amplification such as the PCR even if it is less abundant than endogenous DNA. Although some of the first studies of nuclear aDNA from archaic hominins had problems with contamination (Green *et al.*, 2009; Wall and Kim, 2007), there have been substantial experimental and computational innovations to mitigate its effect in contemporary studies. In the past decade, researchers have developed two broad sets of approaches to correct for contamination in their aDNA samples, allowing for the study of previously unusable sequences.

First, it is now a standard practice to extract aDNA under strict clean-room conditions -including UV irradiation, bleach treatment of surfaces, and filtered air systems- so as to

minimize the proportion of exogenous DNA in the fossil extracts (Champlot *et al.*, 2010; Green *et al.*, 2009). Reagent contamination is also a problem that has yet to be broadly recognized and addressed, *e.g.*, through UV and enzymatic treatment of reagents prior to DNA library construction (Champlot et al., 2010). Additionally, during DNA library construction, unique adapters are incorporated to tag molecules that are present at the moment of extraction (Briggs *et al.*, 2007), preventing contaminant molecules accidentally added during subsequent steps from being confused with endogenous molecules.

Second, after the DNA has been sequenced, several bioinformatic tools can be used to either remove contaminant reads or estimate the proportion of those reads present in a DNA library. A common practice is to estimate the rate of contamination using mitochondrial DNA, which is much more abundant than nuclear DNA and hence is sequenced to a much higher coverage. For highly divergent populations (*e.g.*, Neanderthals), one can use diagnostic positions that distinguish the two groups and assess how many discordant reads are present at each position (Renaud *et al.*, 2015). For modern human populations (*e.g.*, ancient Europeans), one can check for reads that diverge from the consensus sequence or that do not contain molecular signatures consistent with aDNA (Ginolhac *et al.*, 2011; Jónsson *et al.*, 2013). There are also more sophisticated contamination rate estimation methods that use larger subsets of the data, including sex chromosomes (Korneliussen *et al.*, 2014) and entire autosomal genomes (Racimo *et al.*, 2016). Additionally, one can use patterns of cytosine deamination at the ends of fragments (see II.2) to filter out sequenced reads that do not display this signature and are therefore not likely to be ancient (Skoglund *et al.*, 2014).

5. Methodological and technical improvements

Until 5 years ago, available genetic data consisted mostly of mtDNA studies, as the chances of retrieving nuclear data at a sufficiently high level of complexity decreases as we gradually move from cooler and temperate climate zones to warmer and more humid regions such as certain Mediterranean countries and regions in the Near and Middle East (Pruvost *et al.*, 2008). This is unfortunate since these regions are crucial for our understanding of certain population processes such as Neolithization (Pruvost *et al.*, 2008). Most of these studies were based on the targeted PCR amplification of regions of the mitochondrial genome, including the HVR, in order to recover genotypes of diagnostic

SNPs, crucial for haplogroup assessment. This approach, despite being highly cost effective, is prone to spurious results due to the high frequency of contamination with modern DNA when appropriate decontamination procedures are not followed, in particular when remains of H. sapiens are analyzed (Champlot *et al.*, 2010; Olalde *et al.*, 2018; Pääbo *et al.*, 2004).

With the rise of new generation sequencing (NGS) over the past decade, allowing the obtain millions of DNA sequences in only a few days of work, an ever-increasing number of genomes or at least partial genomic information are now available for study. A number of different sequencing technologies and platforms exist, but all require beforehand the conversion of DNA extracts into sequencing libraries. This process consists in the ligation of sequencing adapters to the ends of each DNA molecule found in the extract, enabling the amplification and sequencing of all molecules with a single pair of primers.

Being naturally fragmented, aDNA might at first glance seem like an ideal candidate for such an approach since sequencing library preparation often includes a DNA fragmentation step. In fact, sequencing allows the recovery of short molecules that would not be amplified by PCR. Indeed, only fragments that are long enough to allow for the hybridization of the two PCR primers around a sequence of interest are amenable to direct amplification, while the priming sites required for amplification and sequencing are added externally by ligating adaptor sequences to the molecule ends. Thus, benefits from this strategy are numerous when applied to ancient DNA, and already allowed several breakthroughs in our understanding of the past, by unraveling genomes of long extinct species such as Neanderthals and Denisovans (Briggs *et al.*, 2007; Reich *et al.*, 2010), and demographic processes of past human populations (*e.g.*, Gamba *et al.*, 2014; Haak *et al.*, 2015; Lazaridis *et al.*, 2014). First, the amplification of all molecules within the library renders it “immortal”, as it can be stored and used for several downstream applications without risking the loss of unique molecules. This is a major advantage, since the amount of available material to extract DNA from is a limiting factor in ancient DNA studies. Adapters can carry an index sequence allowing the tagging of each molecule in the library with one or a combination of specific indexes. This allows the identification of a sample in the pool of libraries that is used during the sequencing, but also to monitor potential cross-contamination of the libraries after their barcoding. But the use of ancient DNA material is far from straightforward, as it is scarce and degraded, and some of its features can limit or impair the proper conversion of the extract into a library.

Fortunately, recent major advances in laboratory procedures and sequencing methods have facilitated a shift in the field of paleogenomics, from the analysis of mitochondrial genomes for a small number of individuals to the generation of genome-wide datasets at the scale of large populations.

(1) Identification of preservation niches

The discovery by Gamba and colleagues in 2014 of a specific part of the skeleton as a niche for DNA preservation revolutionized ancient DNA studies. The petrosal part of the temporal bone houses a specific structure, the otic capsule, that is extremely dense and was shown to yield a systematically higher endogenous DNA content compared to other skeletal element (Gamba *et al.*, 2014; Pinhasi *et al.*, 2015). This discovery made aDNA analyses possible of samples from regions that were otherwise not amenable to ancient DNA analyses. More recently, Alberti and colleagues identified through Computed Tomography (CT) scanning the routinely discarded outermost layer of long bones as their densest region, and another promising source of endogenous DNA (Alberti *et al*., 2018).

(2) Better recovery of DNA fragments

Another significant leap forward came through the refinement of library preparation methods. In particular, where most protocols only convert double-stranded molecules into a library by blunt-end ligation of adapters, the *single-stranded* method converts each strand of the DNA fragments separately into library molecules. This provides huge benefits for highly degraded samples where a significant fraction of the remaining DNA is short and partially single-stranded, because such procedures don't rely on an initial repair step that would further shorten molecules presenting single-stranded overhangs. The first application of the single-stranded library preparation method was the generation of a 30-fold coverage genome sequence from the phalanx of a Denisovan individual, an extinct archaic human (Meyer *et al.*, 2012). The success of the method led to further attempts at recovering genomic information from other ancient hominins such as a Neanderthal from the Altai (Prüfer *et al.*, 2014) and a Pleistocene *Homo sapiens* from Ust Ishim in Siberia (Fu *et al.*, 2014). In combination with a DNA extraction method optimized for the recovery of extremely short DNA fragments (Dabney *et al.*, 2013; Glocke and

Meyer, 2017), this procedure also allowed the successful recovery of mitochondrial and nuclear DNA sequences from the hominin fossils of Sima de los Huesos, Spain (Meyer *et al.*, 2014), pushing back the temporal limits of DNA recovery from non-permafrost fossils to ~430,000 years. Although this method is not necessary for all ancient samples, direct comparisons between single- and double-stranded library preparation methods confirmed that single-stranded library preparation greatly improves the yield of library molecules, especially those shorter than 50 bp (Bennett *et al.*, 2014; Meyer *et al.*, 2012).

(3) Enrichment of DNA extracts prior to sequencing

In most ancient specimens, the amount of endogenous DNA is a mere fraction of all the recovered DNA, thus environmental DNA often takes up a high proportion of the sequencing capacity. To circumvent this limitation, and open genomic studies to samples with variable preservation levels, enrichment techniques were developed in order to increase the proportion of DNA of interest in a sample prior to sequencing. These strategies are multiple, and rely on the creation of either DNA or RNA biotinylated baits that cover specific regions (Briggs *et al.*, 2009) or the entirety of the genome (Carpenter *et al.*, 2013) from a modern reference individual, that hybridize to the ancient-DNA libraries and are then recovered with streptavidin-coated beads. The unbound DNA, corresponding mainly to the exogenous fraction of the extract, is washed away, and the captured endogenous DNA is eluted and amplified for sequencing.

Added together, these innovations allowed the generation of genomic information for more than 1000 individuals spanning the Middle Pleistocene (430,000 years BP) to the early 20th century, with a particular focus on the period between 10,000 and 2,500 years ago (**Figure 10**).

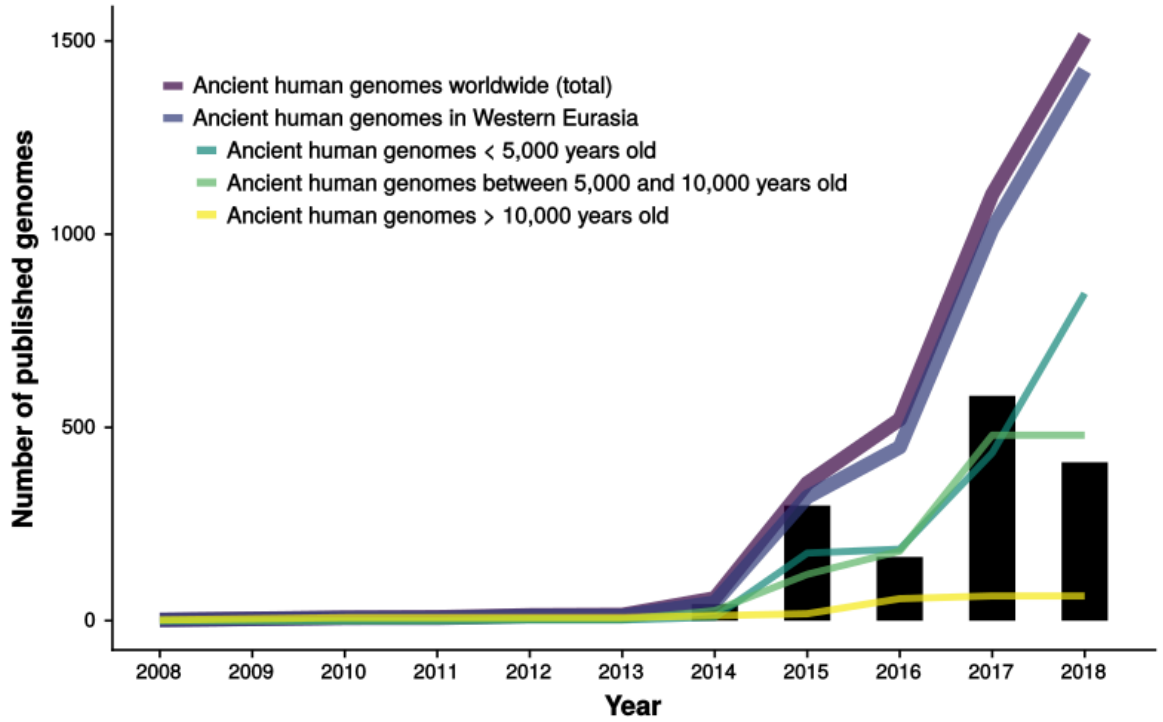


Figure 10: Evolution of the number of published genomes over the last 10 years.

Chapter III. The peopling of Europe during the Holocene

Venturing into the theme of this thesis requires providing some context beforehand. Through the following page, I will briefly introduce the main cultural transitions that occurred during late Prehistory in Europe: the most profound transition from a hunter-gatherer lifestyle to a sedentary lifestyle based on food production that occurred during the Neolithic and the transition from the Neolithic to cultures exploiting ores. This introduction is not an exhaustive description of the various underlying processes that drove the important changes that ultimately shaped our societies. Instead, I will try to give a brief overview of these transitions, calling upon both archeology and genetics, to better integrate the results later presented in this study.

I. Insight from archeology

The contribution of paleogenetics to understand the past and study population dynamics is invaluable but requires an interdisciplinary collaboration with archeologists and anthropologists. Thorough contextualization of sampled individuals is mandatory to interpret genetic patterns, as our ability to define past populations relies on accurate cultural assignment and dating in the first place. Here, I will present a very limited selection of those archeological contexts and hypotheses concerning these cultural transitions and periods, extracted from a prolific literature. These are the absolute prerequisite to interpret the obtained aDNA results and to produce conclusions that have the potential to solve some contentious archeological issues. Since my study concerns the peopling of France, I will focus on western Eurasia.

1. The Neolithic

The Neolithic, as it will be defined in the following section, is a complex process that encompasses both behavioral and cultural characteristics and changes that led to a profound transformation of lifestyles in various regions of the world. Indeed, within a few millennia, several centers of Neolithization appeared across the planet, among which the Fertile Crescent was the first to emerge (c. 12,000 BP) (Bellwood *et al.*, 2007).

(1) Neolithization of the Near East

Around 12,000 years ago, a series of innovations appeared in populations of Anatolia that would ultimately transform our way of life. The Natufian hunter-gatherers from the Levant had already become more sedentary during the Late Glacial, as people started relying on wild grains along with hunting and fishing (Boyd, 2006). The transition to sedentism progressively led people to start shaping the environment around them into a sustainable source of food for them to thrive for several generations. The flora was transformed, and signatures of domestication of plant species were first detected in what is called Pre-Pottery Neolithic B (PPNB, 7500 - 7000 cal. BC), a culture that followed the Natufian, even though the process might have started around 2,000 years earlier. From an economy based on harvest, people transitioned to a food production system relying on cultivation and transformation of wild species into crops (Willcox, 2005). The fauna also underwent important changes around this time, as it was progressively transformed to suit the needs of humans. While wolves (Canis lupus) were domesticated since the Late Glacial (18,000- 12,000 cal. BC, Savolainen *et al.* 2002), species used for food production were domesticated later, during the PPNB: wild goat (Capra aegagrus), bighorn (Ovis orientalis), auroch (Bos primigenius) and boars (Sus scrofa), all becoming very different from their wild ancestors (Zeder, 2008). After 7,000 cal. BC, the use of ceramics became widespread throughout the Near East, except in southern Levant and in the Syrian Desert (Mazurié de Keroualin, 2003).

These transformations have deeply altered prehistoric communities and their relationship with their environment, leading researchers to first use the term *revolution* to characterize the Neolithic transition for a long time (Childe, 1936). This term was later abandoned as the Neolithic transition was not a short-term event but took place over millenia, and the various elements mentioned above formed what is now described as a Neolithic *package* (Guilaine and Manen, 2004).

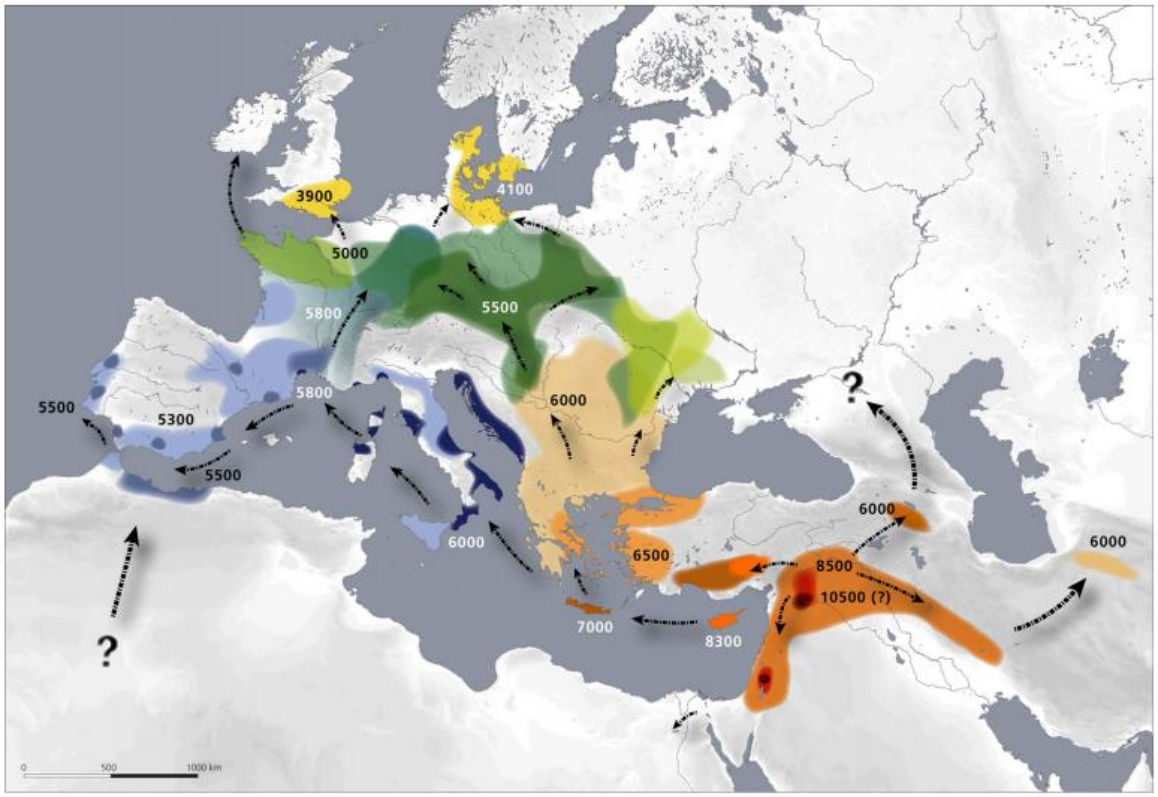


Figure 11: Neolithic cultures in Western Eurasia
The map shows the distribution of the first farming societies in western Eurasia together with the dates of major human migrations (in years calibrated B.C.) originating in the Fertile Crescent (Gronenborn, 2014).

(2) Spread towards Europe

The oldest traces of the Neolithic spread out of the Near East are observed in Cyprus during the IXth millennium BCE (Vigne *et al.*, 2012). The oldest settlements from the European continent are found in Thessaly and Crete around the VIIth millennium BCE, as the diffusion towards Europe goes through Greece and the Balkans (Reingruber and Thissen, 2011), by land as well as by the sea. It eventually reached both the westernmost and northernmost parts of Europe over the course of the following millennia, after dividing in two primary routes of diffusion: one along the Danube river corridor up to the Atlantic coast, the other along the Mediterranean coastline, relying heavily on boat and ultimately reaching the Iberian Peninsula (Bocquet-Appel *et al.*, 2009). Little is known about the respective part of migration or spread of ideas in this diffusion, and to what extent admixture with local populations occurred, if it did at all. What is known for sure is that the Neolithic diffusion occurred at a variable pace, probably due to environmental constraints, population dynamics and the autochthonous Mesolithic substrate. Ultimately,

from a number of identified source cultures: the Impressa and Cardial cultures for the Mediterranean Sea, the Starcevo and Linearbandkeramik (LBK) cultures for Central Europe, Neolithization spread and reached all parts of Europe, where it evolved into a wide variety of cultures.

The first farmers settled in France during the VIth millennium BCE. Far from being homogenous, the cultural, economic and symbolic components of these initial farming communities reflected their different historical trajectories. Chronology varies with the region considered, as the Impressa current appears to have arrived in Southern France about 500 year earlier than the oldest traces of the Danubian LBK in North-Eastern France, which was established in the Rhine basin around 5200 BCE, as the Mediterranean area transitions to the Cardial complex. Upon establishing themselves over the territory of France, these two pioneer fronts underwent cultural mutations as a result of their geographic expansion. This increasing regionalization ultimately led to both the Cardial and LBK complexes breaking up over the course of the Vth millennium BCE to give rise to a number of new cultures. (See **Supplementary Figures 1-3** for more details about the different Neolithic cultures found on the territory of present-day France)

2. The Bronze Age

The period from the late IVth millennium BCE saw many important social changes that varied from region to region but laid the foundations for the society of the upcoming Bronze Age.

The Chalcolithic Age, or Copper Age, refers to the initial use of pure copper (along with its predecessor tool making material, stone) and is considered by researchers as an intermediate stage between the Neolithic and the Bronze Age (Guilaine, 1972; Lichardus and Lichardus-Itten, 1985). Metal production from copper appeared in Neolithic Anatolia during the VIIth millennium BCE. Around the transition from the VIth to the Vth millennium BCE, the use of copper ore is attested in the Balkans and later spread towards the Carpathian basin, Hungary and Poland to reach the center of Germany. It then reached the Alps and Northern Italy around 3800-3500 BCE (Carozza *et al.*, 2007). In France, while the oldest copper items were recovered from the Paris Basin, they were shown to be only the result of trade as proper metallurgy only started in the region during the Bronze Age (Mille and Bouquet, 2004). In Southern France, on the contrary, copper is an abundant resource and is exploited during the IIIrd millennium BCE (Ambert and Carozza, 1996).

Simultaneous with such Copper Age cultures occured a number of Late Neolithic cultures in other regions. During the IIIrd millennium BCE, two new archaeological cultures expanded across Europe and replaced many of the more localized ones that had preceded them. The first complex, characterized by Corded Ware pottery and stone battle-axes, is found particularly in central and northern Europe (Childe, 1929). The second, equally expansive complex, dated to 2500–2200 BCE in Western Europe, is marked by the Bell Beaker pottery, and the frequent occurrence of copper daggers and stylized bell-shaped pots in the graves. The latter is found from Iberia, where the oldest radiocarbon sets the beginning of this culture to around 2750 BCE (Cardoso, 2014), to Britain and as far east as Hungary, with a place of origin still under debate to this day (Jeunesse, 2015).

Although the dates and the cultural roots of the Early Bronze Age vary, it is similarly defined by the use of copper alloys for tools throughout Europe. Appearing in Anatolia, the metallurgy of alloys then spread towards south-western Europe and Great Britain during the IIIrd millennium BCE. It is commonly accepted that Bronze Age started in France around 2300 BCE and ended sometime between 800 and 750 BCE. It is divided into three main phases: The Early Bronze Age (2300-1650 BCE), the Middle Bronze Age (1650-1350 BCE) and the Late Bronze Age (1350-750 BCE). During this period, the techniques of metalworking increased in sophistication (Armbruster and Pernot, 2006). Characterized by the adoption of new means of production and specialization, and both the accumulation and consumption of luxury items, Bronze Age is described as another crucial step in the evolution of human societies (Childe, 1950). Indeed, with the advent of metallurgy, society as a whole became more hierarchically organized. and everywhere along the Mediterranean basin, sites of central authority appeared and trade intensified, notably towards Mycenian Greece (Harding, 2000).

3. The Iron Age

While the discovery of sporadic iron items dates back to around 1700 BCE in Europe, Iron really began to supplant bronze in the production of tools and weapons at different times in various parts of the continent, and the transition to the Iron Age is embedded in local cultural developments. Archaeological research revealed that the use of iron was widespread in the eastern Mediterranean by 1200 BCE and that iron technology was established in Greece by 1000 BCE. It was conventionally dated between 800 and 50 BCE

in Central Europe. Indeed, a significant amount of this new metal appears towards the end of the VIIIth century, mostly in the form of swords, associated with a masculine elite. (Brun, 1987). The central European Iron Age has been divided into two sequential periods named after important archaeological sites. The earlier period (800–480 BCE.) is known as the Hallstatt period. The later period (480–25 BCE.) is known as the La Tène period and is characterized by a very distinctive style of decoration on metalwork.

Table 1: Compared chronologies of the Bronze and Iron Ages in the German and the French system

Date BCE	German system	French system	Period
1600	Bronze B1 Bronze B2	Middle Bronze I	Middle Bronze Age
1500	Bronze C1 Bronze C2	Middle Bronze II	
1350	Bronze D	Late Bronze	Late Bronze Age
1250	Hallstatt A1	Late Bronze IIa	
1150	Hallstatt A2	Late Bronze Iib	
1020	Hallstatt B1	Late Bronze IIIa	
930	Hallstatt B2/3	Late Bronze IIIb	
800	Gündlingen	Gündlingen	1st Iron Age
730	Hallstatt C	Early Hallstatt	
620	Hallstatt D1	Middle Hallstatt	
530	Hallstatt D2	Late Hallstatt I	
460	Hallstatt D3	Late Hallstatt II	
400	La Tène A	La Tène Ia	2nd Iron Age
320	La Tène B1	La Tène Ib	
250	La Tène B2	La Tène Ic	
180	La Tène C1	La Tène IIa	
150	La Tène C2	La Tène IIb	
90	La Tène D1	La Tène IIIa	
25	La Tène D2	La Tène IIIb	

Adapted from (Brun and Ruby, 2008)

The beginnings of ironworking represented a fundamental technological revolution for ancient Europe. While sources of copper and tin (which form bronze when alloyed together) were rare in prehistoric Europe, iron ores were ubiquitous. South and west-central Europe were now included in the periphery of the expanding Mediterranean civilization, as a vast trade network established with the Greek and Etruscan civilizations. (Brun, 1987). In parallel, societies became more complex in their political structure and

showed increasing social differentiation (Collis, 1981). Principalities arose, as evidenced by the wealth accumulated in certain large tumulus burials of the Hallstatt period, such as the Vix tumulus in Burgundy (Joffroy, 1979). The second Iron Age was marked by the first mentions of the Celts during the Vth century BCE, associated with the La Tène culture. With the establishment of Greek colonies along the Mediterranean Sea, such as the one in Marseilles in 600 BCE, trade with Greek and Etruscan city-states remained flourishing for Celtic chiefdoms throughout Europe. The end of the Celtic migrations that marked the IVth and IIIrd century BCE was associated with a demographic increase during the last 200 years of the Iron Age. Centralized settlement increased with the appearance of large defended sites, the so-called oppida (Collis, 1984). The development of technologies for the smelting and forging of iron led to the greater use of metals for everyday tools such as agricultural implements. As a results, subsistence production must have increased drastically (Audouze and Buchsenschutz, 1989). During the La Tène period, both archaeological and historical information can be used to reconstruct the Late Iron Age ways of life, as Late Iron Age peoples also appear in Greek and Roman texts such as historical and geographical works.

II. Ancient DNA studies helped to end longstanding debates

During the past five years, Western Eurasia has yielded more aDNA genomes than any other region in the world. The first aDNA complete genome sequence from Europe came from the Tyrolean Iceman, a 5,300-year-old (Late Neolithic or Copper Age) natural mummy discovered in 1991 in the Ötztal Alps. Surprisingly, the Iceman had more genetic affinity to present-day Sardinians than to the present-day populations inhabiting the region where he probably lived (Keller *et al.*, 2012), showing that major demographic changes have occurred in Europe after the Neolithic era. Furthermore, a genomic sequence from a 5,000-year-old farmer from Scandinavia was also found to have close genetic ties to Sardinians, unlike contemporaneous hunter-gatherers from the same region (Skoglund *et al.*, 2012).

Understanding the genetic implication of the process of Neolithization in Europe, or the later transition towards the Bronze Age, requires to define the genetic substrate, i.e. the genetic diversity that existed in Europe prior to these transitions, during the Upper Paleolithic (~50,000 to 11,000 years ago) and Mesolithic (~11,000 to 6,000 years ago).

Several mitochondrial sequences and nuclear genomic information from individuals dating from 45,000 to 7,000 years ago have been reported in various studies (Bramanti *et al.*, 2009; Fu *et al.*, 2014, 2016; Jones *et al.*, 2015; Malmström *et al.*, 2009; Raghavan *et al.*, 2014; Seguin-Orlando *et al.*, 2014; Skoglund *et al.*, 2014), revealing a Upper Paleolithic and post-glacial Mesolithic female substrate characterized by mtDNA haplogroup U, *e.g.* U2, U4, U5a, U5b, U8 across Europe. This observation is consistent with the previously mentioned study from Richards and colleagues (Richards *et al.*, 2000), that suggested that haplogroup U was one of the oldest Eurasian branches in the mitochondrial phylogeny, therefore a prime candidate for an expanding lineage during the initial peopling of Eurasia by anatomically modern humans some 40,000 years ago (Mellars, 2006). This particular haplogroup composition is in sharp contrast with that found in current populations, supporting the evidence that modern populations are not the direct descendants of the pre-Neolithic groups. It is also clearly different from that found in Neolithic individuals living in the same geographical area, in which haplogroup U is found to a lower frequency (Brandt *et al.*, 2013; Haak *et al.*, 2005; Lazaridis *et al.*, 2014). This supports the advent of farming in Europe as a demographic process involving the migration of people into Europe that ultimately replaced the autochthonous population, modifying the European genetic landscape.

1. The Neolithic

The first aDNA studies that addressed the question of the Neolithic diffusion in Europe were contrasted. On the one hand, Neolithic skeletons from various locations in Germany, Austria, and Hungary were found to harbor one specific mitochondrial DNA, N1a, rare in present-day Europeans, suggesting little maternal contribution from Neolithic farmers to modern populations, and lending support to an acculturation model for autochthonous individuals (Haak *et al.*, 2005). The first support in favor of a demic diffusion of the Neolithic into Europe came in 2007 with the study of 11 Neolithic remains from Granollers (Catalonia, northeast Spain) dated to 5,500 years BP (Sampietro *et al.*, 2007). The variety of maternal haplogroups displayed by these individuals was shown to be close to that found in modern populations from the Iberian Peninsula. Subsequent studies of the maternal composition of both autochthonous Mesolithic hunter-gatherers and Neolithic individuals further confirmed these results (Bramanti *et al.*, 2009; Brandt *et al.*,

2013; Haak *et al.*, 2010), revealing large genetic differences between these two groups across Europe.

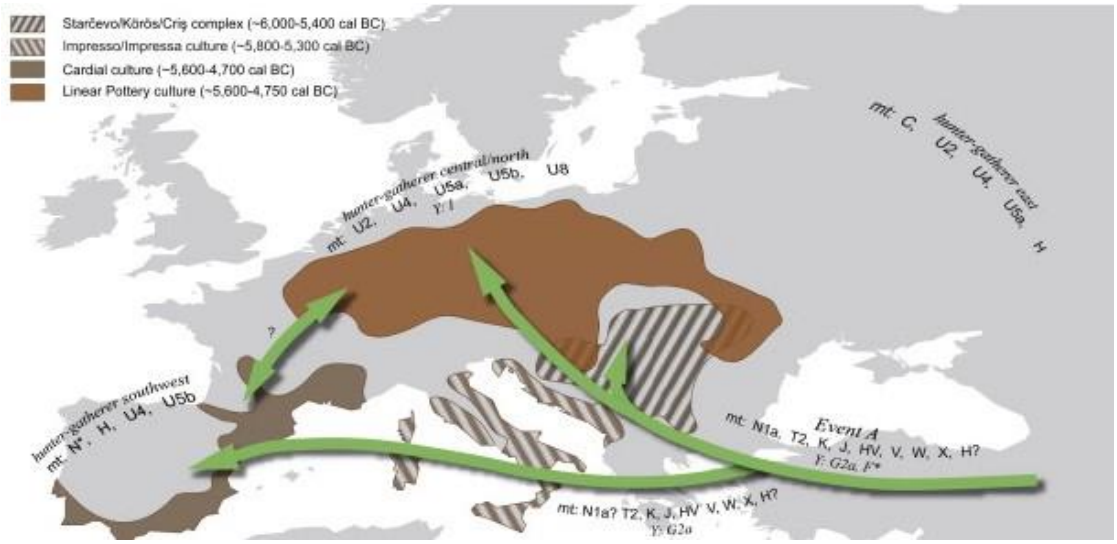


Figure 12: Spread of the Neolithic into Europe and associated haplogroups
(Brandt *et al.* 2013)

Genomic studies revealed that ~8,000-7,000 years ago, farmers from Anatolia had spread their Near Eastern ancestry across most of Europe, reaching as far as Scandinavia and Iberia. Early Neolithic farmers from Germany, Hungary, and Spain were shown to be closely related, and to have genetic ties to those of Greece and Anatolia (Hofmanová *et al.*, 2016; Lazaridis *et al.*, 2014), while being strikingly different from indigenous hunter-gatherers. In each of the three regions mentioned above, the arrival of farmers was shown to have prompted admixture with local hunter-gatherers, which unfolded over many centuries (Lipson *et al.*, 2017), resulting in an increase of the hunter-gatherer ancestry over time. Still, apart from present day Sardinians who retained close ties to Neolithic individuals such as Ötzi and the Scandinavian farmer mentioned before, modern Europeans are different from their Neolithic predecessors.

2. The Bronze Age

Back in 2014, the sequencing of a 24,000-year- old individual from Siberia (Raghavan *et al.* 2014) hinted to a more complex model for the origin of Europeans, as analyses revealed that at least three different ancient populations contributed to the genetics of present-day Europeans (Lazaridis *et al.*, 2014). The first two were the previously described autochthonous West European hunter-gatherers such as Loschbour (Lazaridis

et al., 2014) and the two La Braña individuals (Olalde *et al.*, 2014; Sánchez-Quinto *et al.*, 2012) along with early European farmers. But a third and, at the time, most mysterious component was drawn from ancient north Eurasians related to Upper Paleolithic Siberians (Seguin-Orlando *et al.*, 2014). The analysis of 69 additional Eurasians who lived between 3,000 and 8,000 years ago, including people from the Pontic Steppe (that stretches from the northern shore of the Black Sea as far east as the Caspian Sea, forming part of the larger Eurasian steppe adjacent to the Kazakh steppe to the east) shed more light on the unexpected ancestry of Europeans (Haak *et al.*, 2015). This study revealed that the arrival of the first farmers during the Early Neolithic from the Near East was followed by a massive migration from the Eurasian Steppe ~4,500 years ago, associated with the Late Neolithic *Corded Ware* culture, and involving herders from the Yamnaya culture. This Yamnaya population derived part of its ancestry from the hunter-gatherers who preceded them in this region, who distantly shared ancestry with the ancient Siberians. As such, it was probably one of the sources of the Ancient north Eurasian ancestry previously identified (**Figure 13**)(Allentoft *et al.*, 2015). Present-day European Y-chromosomal haplogroups illustrate that the migration of the Yamnaya across the continent was a massive one, as the signature haplogroup of the latter, namely R1b, is now predominant in Western Europe (Batini *et al.*, 2015).

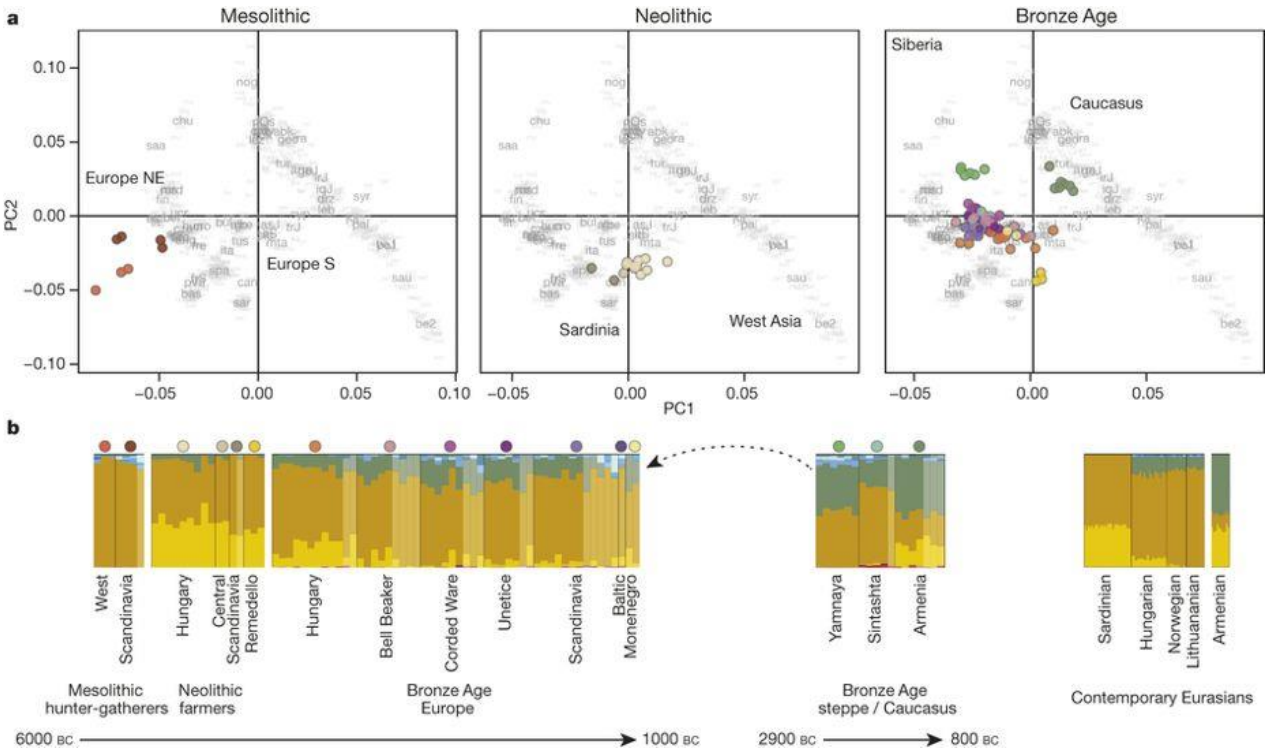


Figure 13: Yamnaya-related admixture into Bronze Age European populations

a. Principal component analysis (PCA) of ancient individuals from different periods projected onto contemporary individuals from Europe, West Asia, and Caucasus. **b**. ADMIXTURE ancestry components (K = 16) for ancient and selected contemporary individuals. Individuals with less than 20,000 SNPs have lighter colors. Colored circles indicate corresponding group in the PCA. Probable Yamnaya-related admixture is indicated by the dashed arrow. (Allentoft *et al.* 2015)

In parallel to the development of the Corded Ware complex, another large-scale complex with a distinctive pottery style became established in Western Europe: The Bell Beaker complex. Initially thought to be genetically homogenous, the Bell Beaker cultural complex expanded through a mixture of migration and cultural diffusion, which had different impacts depending on the region. Bell Beaker associated individuals from Iberia do not display as much of the Steppe-related ancestry that is present in both Central Europe and Great Britain, but instead retain genetic affinity with preceding Neolithic populations (**Figure 14**). They were also shown not to have contributed ancestry to Beaker individuals outside Iberia, thus excluding migration from Western Europe to the east as the sole possible cause for the spread of this culture (Olalde *et al.*, 2018).

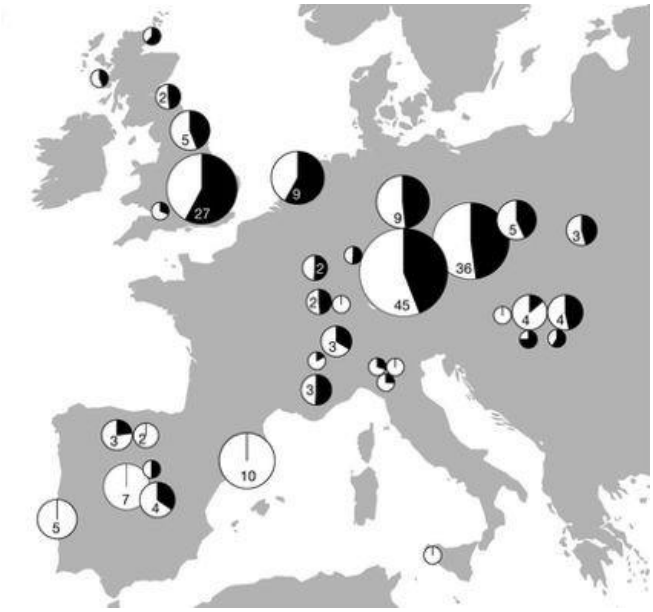


Figure 14: Proportion of steppe-related ancestry in Beaker-complex-associated groups in Western Europe.
Steppe related ancestry is displayed in black. The area of the pie is proportional to the number of individuals. (Olalde *et al.* 2018)

In summary, aDNA studies solved a longstanding archeological debate by showing that the Neolithic transition in Europe resulted from the movement of people from Anatolia. They also highlighted the incorporation of the autochthonous hunter-gatherer's gene pool into that of farmers, and unanticipated phenomena, such as the large-scale genetic contribution of the Yamnaya to the European gene pool at the onset of Bronze Age, or the genetic heterogeneity of the Bell Beaker cultural complex.

3. Natural selection

As modern humans expanded out of Africa to colonize the world, they came to inhabit various environments, each with their own constraints. Survival therefore depended upon adaptation to local conditions: phenotypic changes resulting from the fine tuning of the genomes of individuals, ultimately allowing them to thrive in their new habitat. Signatures of past selection events are retrieved from the genomes of present-day people, by searching for variants displaying differential frequencies across the continents.

Variation in skin color, one of the most discernible phenotypic traits among humans, was thought to be an adaptation to reduced sunlight regions, as the production of vitamin D in the skin is catalyzed by UBV radiation (Muehlenbein, 2010). The causing mutation was therefore expected to be present in the first inhabitants of Europe during the Paleolithic (Beleza *et al.*, 2013). While Scandinavian hunter-gatherers and Early European farmers indeed carried the derived allele for variants contributing to light skin (Mathieson *et al.*, 2015), western hunter-gatherers carried the ancestral allele at these genetic loci, implying that these individuals displayed dark skin tones (Brace *et al.*, 2018; Mathieson *et al.*, 2015; Olalde *et al.*, 2014). Therefore, light skin might not have been an essential adaptation for survival in Paleolithic Europe, and the cause behind its increase in frequency might have been else.

The transition from a hunter-gatherer lifestyle to one based on food production that occurred during the Neolithic deeply altered the interactions between humans and their environment. In particular, changes in diet and zoonotic infections with new pathogens resulting from plant and animal domestication were expected to have triggered episodes of selection on genetic variants associated with both metabolic pathways and pathogen resistance (Wolfe *et al.*, 2007). The genome of a Mesolithic hunter-gatherer from Spain revealed that some adaptive variants associated with pathogen resistance in modern Europeans were already present before the advent of agriculture (Olalde *et al.*, 2014). In fact, a large-scale paleogenomic study conducted on ancient Eurasians only revealed signatures of selection on a limited number of loci associated with diet and immunity dating back to the Neolithic (Mathieson *et al.*, 2015). The mutation responsible for lactase persistence, the strongest signal of positive selection in present-day Europeans, was not among them, as its first occurrence in Europe dates back to the Bronze Age, where it is found in 10% of studied individuals (Allentoft *et al.*, 2015; Mathieson *et al.*, 2015).

Chapter IV. Description of the thesis project

I. State of the art

As mentioned before, the last 10,000 years in Western Eurasia were marked by major cultural transitions: the most profound transition from a hunter-gatherer (HG) lifestyle to a sedentary lifestyle based on food production that occurred during the Neolithic (*e.g.*, Tringham, 2000; Bellwood, 2004) and the transition from the Neolithic to cultures exploiting ores (Kristiansen and Larsson, 2005). These cultural transitions were the result of demographic changes that deeply altered the genetic makeup of the Western Eurasian populations (Gamba *et al.*, 2014; Haak *et al.*, 2015; Lazaridis *et al.*, 2014; Olalde *et al.*, 2018; Skoglund *et al.*, 2012). The Neolithic culture expanded through the migration of early farmers from Anatolia into Europe starting about 7,500 years ago following two different routes. One route went through the Balkans and along the Danube northwards to the Hungarian plain and from there westwards to the Parisian Basin. The second moved along the coastline of the Mediterranean basin to arrive in Southern France and Spain (Lipson *et al.*, 2017). These early farmers brought two distinct Neolithic cultures to the territory of present-day France, the Linearbandkeramik (LBK) culture in the North and the Impressa and Cardial culture in the South. Later, over the course of the IIIrd and IInd millennia BCE, the Neolithic culture was replaced by the Bronze Age culture (Kristiansen and Larsson, 2005), which in turn evolved into the Iron Age during the last seven centuries BCE (Brun and Ruby, 2008).

In France, these transitions are known only from the rich archaeological records while the underlying demographic processes are not yet explored at a territory-wide scale. Only a handful of studies restricted to individual archeological sites have been performed relying on partial mitochondrial sequence information (Beau *et al.*, 2017; Lacan *et al.*, 2011; Rivollat *et al.*, 2015, 2016) or on partial Y chromosome sequences (Lacan *et al.*, 2011). Scattered ancient French genomes have started to be released as comparative material through large-scale studies, but no global conclusion has been drawn regarding the territory as a whole (Fu et al., 2016; Olalde et al., 2018; Posth et al., 2016). This gap needs to be closed since France is a geographic crossroads, providing multiple opportunities for interaction between populations of different origins, as is well documented by the archaeological record for later prehistory. In the Early Neolithic, it

was here that the central European and Mediterranean currents of neolithization eventually became established (Demoule, 2007). It is also an obliged passage towards both the Iberian Peninsula and the British Isles, both settled by Neolithic farmers during the VIth and Vth millennia BCE respectively. To what extent the "Central European" and "Mediterranean" Neolithic spheres interacted and admixed is an open question. Moreover, the territory of present-day France was part of a dense Europe-wide trade network during the Bronze Age and later during the Iron Age (Brun and Ruby, 2008; Carozza et al., 2007). The consequences on the French gene pool of the development and spread of metallurgy across Europe are still unknown.

II. Aim of the project

During this project, we aimed at shedding light on the peopling of present-day France, while providing a reliable and cost-effective analysis to screen a wide and heterogeneous panel of ancient individuals. We analyzed Mesolithic, Neolithic, Bronze Age and Iron Age individuals sampled from three interface regions of present-day France, the North, the East, and the South (Hauts-de-France, Grand-Est and Occitanie, respectively).

One of the major drawbacks when working with ancient DNA is its low relative quantity within a sample compared to any source of exogenous, environmental DNA, mainly microbial DNA from soil and bone-colonizing bacteria. Thus, the choice of a methodological strategy that offers the best compromise between cost, throughput, and depth of genetic analyses, allowing for an efficient analysis of a relatively large dataset that can contain a significant proportion of poorly preserved samples, was crucial.

Using a combination of targeted and genome-wide approaches, we set out to characterize the population across the three sampled regions during these periods and the transitions between them on a genetic level. The correlation of the obtained genomic information and their archeological context aims at characterizing major cultural transitions in the history of France, such as the Neolithization and the transition to ore exploitation during the Bronze and the Iron Age.

III. Strategy

For the majority of samples, we opted for a targeted approach through the enrichment of ancient DNA libraries with endogenous sequences of interest prior to sequencing. This allows a limited number of genomic targets to be sequenced at a medium coverage with only a moderate sequencing effort per sample, which permits to investigate specifically the regions mentioned above. This capture relies on two preliminary steps: the amplification of the sequences of interest using modern DNA as a template, and *in vitro* transcription of the PCR products to obtain biotinylated RNA baits to be used for capture.

A first part of the project thus consisted in selecting markers that were relevant from an archaeological point of view and setting up the protocol for their capture. As they are traditionally used in demographic studies, non-recombining markers such as the mitochondrial genome and Y chromosome, respectively maternally and paternally inherited, were investigated to provide some insight into the origins of the analyzed individuals. Nuclear DNA yields information about several characteristics, such as phenotypic or physiological traits. Therefore, changes in the genetic makeup of populations resulting from selective pressures or migrations can be assayed through a limited number of loci. We implemented in our study polymorphic sites present in genes involved in melanin biosynthesis and associated with eye or hair color (Walsh *et al.*, 2013). Moreover, the Neolithic transition in Europe represented a shift in lifestyle that was shown to be associated with a similar shift in pathogen load. Pathogens being the main selective pressure on human genomes (Fumagalli *et al.*, 2011), we decided to add several immunity-related markers to our study, either known to be risk factors or to allow resistance to the infection by several parasites, or simply identified as being under positive selection in ancient and modern non-African populations, thus advantageous at some point in our history (Barreiro and Quintana-Murci, 2010; Casals *et al.*, 2011; Mathieson *et al.*, 2015). To further determine the geographic origin of the samples, several ancestry informative markers were also investigated.

In parallel to these methodological developments, ancient DNA libraries were characterized for their suitability for the capture of nuclear markers in order to better understand the tight intertwining between the library preparation procedure and the capture requirements. After a shallow shotgun sequencing screening, I proceeded with the capture of samples from Mesolithic, Neolithic, Bronze Age and Iron Age backgrounds,

and performed a set of descriptive analyses on the combined dataset of ancient French and other ancient Western Eurasian individuals available from the literature.

Lastly, the best-preserved samples were subjected to a shotgun sequencing approach in order to obtain low coverage genomes, which upon being combined with the published genotypes of both modern and ancient Western Eurasians, opened the way to a wider range of population genetics analyses.

Part II. Material and methods

Chapter I. Material

In order to thoroughly cover the different periods encompassed in the 5,000-year transect of this project, a total of 342 samples were processed over the three years of this study. From the onset of the project, the sampling was mainly oriented towards temporal bones when available, which resulted in the following distribution for sample types:

- 270 temporal bones,
- 45 long bones,
- 27 teeth.

Most of the sampled individuals spanned the Neolithic to the Iron Age (~5500 to 25 BCE) and were collected from three regions of the present-day French territory: 138 from the Grand-Est (East), 123 from Occitanie (South) and 71 from the Hauts-de-France (North). The remaining 10 samples were associated with the Mesolithic period and sampled outside the geographic regions initially encompassed in the project: 4 from Île-de-France (North) and 6 from Nouvelle-Aquitaine (South-West). This resulted in the following dataset over the time transect spanning the Mesolithic to the Iron Age:

- 10 Mesolithic individuals (7100-6500 BCE),
- 219 Neolithic individuals, among which 37 Early Neolithic individuals (5500–4500 BCE), 32 Early Middle Neolithic (4500–4000 BCE), 111 Late Middle Neolithic (4000-3500 BCE) and 39 Late Neolithic (3500-2500 BCE) among which 4 were associated with the Late Neolithic Bell Beaker culture,
- 49 Bronze Age individuals (2500-800 BCE),
- 64 Iron Age individuals (800-25 BCE).

A list of all the samples included in the present study can be found on the following pages.

Table 2: List of all samples included in the present study

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
Mairy4249	4249		PB	les hautes Chanvières	Mairy	Grand-Est (East)	Ardennes	Middle Neolithic "Late Michelsberg"	3646-3379 cal BC	4770 +/- 45BP
BLI530	530		PB	Les Voies de Brienne	Blignicourt	Grand-Est (East)	Aube	Bell Beaker		
BRE447	447		TO	ZAC st Martin	Bréviandes	Grand-Est (East)	Aube	Early Neolithic "LBK"	5209-4912 cal BC	6100 +/- 40BP
BRE445A	445	A	PB	ZAC st Martin	Bréviandes	Grand-Est (East)	Aube	Late Neolithic	2500-2200 BC	
BRE445B	445	B	PB	ZAC st Martin	Bréviandes	Grand-Est (East)	Aube	Late Neolithic	2568-2299 cal BC	3940 +/- 40BP
BRE445C	445	V	PB	ZAC st Martin	Bréviandes	Grand-Est (East)	Aube	Late Neolithic	2500-2200 BC	
BRE445FK	445	FK	PB	ZAC st Martin	Bréviandes	Grand-Est (East)	Aube	Late Neolithic	2500-2200 BC	
BRE445HI	445	HI	PB	ZAC st Martin	Bréviandes	Grand-Est (East)	Aube	Late Neolithic	2500-2200 BC	
BUCH8	D40F0099		TO	PLA D39	Buchères	Grand-Est (East)	Aube	Early Neolithic		
BUCH3	D39F0269/1		TO	PLA D39	Buchères	Grand-Est (East)	Aube	Middle Neolithic 1	4464-4346 cal BC	5575 +/- 35BP
BUCH4	D39F0270/1		TO	PLA D39	Buchères	Grand-Est (East)	Aube	Middle Neolithic 1	4527-4365 cal BC	5625 +/- 35BP
BUCH5	D39F0276		TO	PLA D39	Buchères	Grand-Est (East)	Aube	Middle Neolithic 1	4400-4200 BC	
BUCH6	D39F0279		TO	PLA D39	Buchères	Grand-Est (East)	Aube	Middle Neolithic 1	4400-4200 BC	
BUCH2	D39F0274		PB	PLA D39	Buchères	Grand-Est (East)	Aube	Middle Neolithic 1	4400-4200 BC	
VAC4155-1	4155	1	PB	Pré Chevalier	La-Villeneuve-au-Chatelot	Grand-Est (East)	Aube	Middle Neolithic 2		
VAC4155-2	4155	2	PB	Pré Chevalier	La-Villeneuve-au-Chatelot	Grand-Est (East)	Aube	Middle Neolithic 2		
VAC5566	5566		PB	Pré Chevalier	La-Villeneuve-au-Chatelot	Grand-Est (East)	Aube	Middle Neolithic 2		
VAC1061	1061		PB	Pré Chevalier	La-Villeneuve-au-Chatelot	Grand-Est (East)	Aube	Middle/Late Neolithic		
PSS4170	4170		PB	Pont-sur-Seine	Pont-sur-Seine	Grand-Est (East)	Aube	Bronze Age	2196-1977 cal BC	3690 +/- 30BP
PSS4693	4693		PB	Pont-sur-Seine	Pont-sur-Seine	Grand-Est (East)	Aube	Middle Neolithic 2	3626-3351 cal BC	4650 +/- 45BP
PSS2042	2042		PB	Pont-sur-Seine	Pont-sur-Seine	Grand-Est (East)	Aube	Middle Neolithic 2		
PSS282	282		PB	Pont-sur-Seine	Pont-sur-Seine	Grand-Est (East)	Aube	Middle Neolithic 2		
PSS3072	3072		PB	Pont-sur-Seine	Pont-sur-Seine	Grand-Est (East)	Aube	Middle Neolithic 2		

Part II

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
PSS1	1		PB	Haut de Launois	Pont-sur-Seine	Grand-Est (East)	Aube	Middle/Late Neolithic	3634-3373 cal BC	4715 +/- 40BP
PSS2	2		TO	Haut de Launois	Pont-sur-Seine	Grand-Est (East)	Aube	Middle/Late Neolithic	3640-3379 cal BC	4755 +/- 40BP
PSS2323	2323		PB	Pont-sur-Seine	Pont-sur-Seine	Grand-Est (East)	Aube	Middle/Late Neolithic	3646-3382 cal BC	4775 +/- 40BP
SLPT201	201		TO	le château la Planche	Saint-Léger-près-Troyes	Grand-Est (East)	Aube	Early Neolithic	5216-4913 cal BC	6120 +/- 50BP
SLPT505	505		TO	le château la Planche	Saint-Léger-près-Troyes	Grand-Est (East)	Aube	Early Neolithic	5291-5026 cal BC	6190 +/- 40BP
ERS1164	1164		PB	Untergasse	Erstein	Grand-Est (East)	Bas-Rhin	Iron Age "La Tène"		
ERS86	86		PB	Untergasse	Erstein	Grand-Est (East)	Bas-Rhin	Iron Age "La Tène"		
ERS88	88		PB	Untergasse	Erstein	Grand-Est (East)	Bas-Rhin	Iron Age "La Tène"		
ERS83-2			PB	Untergasse	Erstein	Grand-Est (East)	Bas-Rhin	Iron Age "La Tène"		
MOR6	6		PB	ZAC des collines I IKEA, Hungeheuer Hoelzle	Morschwiller-le-Bas	Grand-Est (East)	Bas-Rhin	Early Neolithic "LBK"	5200-5000 BC	
MOR4	4		PB	ZAC des collines I IKEA, Hungeheuer Hoelzle	Morschwiller-le-Bas	Grand-Est (East)	Bas-Rhin	Early Neolithic "LBK"	5200-5000 BC	
MOR5	5		PB	ZAC des collines I IKEA, Hungeheuer Hoelzle	Morschwiller-le-Bas	Grand-Est (East)	Bas-Rhin	Early Neolithic "LBK"	5200-5000 BC	
MOR54	54		PB	ZAC des collines I IKEA, Hungeheuer Hoelzle	Morschwiller-le-Bas	Grand-Est (East)	Bas-Rhin	Early Neolithic "LBK"	5200-5000 BC	
NOR2B6	2B	6	PB	Buerckelmatt	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Halstatt C, early La Tène"		
NOR3-15	3	15	PB	Buerckelmatt	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Halstatt C, early La Tène"		
NOR4	4		PB	Buerckelmatt	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Halstatt C, early La Tène"		
NOR2B2	2B	2	PB	Buerckelmatt	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Halstatt C, early La Tène"		
NOR2B20	2B	20	PB	Buerckelmatt	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Halstatt C, early La Tène"		
NOR3-6	3	6	PB	Buerckelmatt	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Halstatt C, early La Tène"		
NOR3-10	3	10	PB	Nordhouse	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Hallstatt - Early Late Tène"		
NOR3-13	3	13	PB	Nordhouse	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Hallstatt - Early Late Tène"		
NOR36	36		PB	Nordhouse	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Hallstatt - Early Late Tène"		
NOR5-2	5	2	PB	Nordhouse	Nordhouse	Grand-Est (East)	Bas-Rhin	Iron Age "Hallstatt - Early Late Tène"		
OBE3722	3722		PB	PAEI	Obernai	Grand-Est (East)	Bas-Rhin	Early Bronze Age	2000-1700 BC	
OBE1450	1450		PB	PAEI	Obernai	Grand-Est (East)	Bas-Rhin	Early Bronze Age	1896-1695 cal BC	3485 ± 35 BP
OBE3626-2	3626	2	PB	PAEI	Obernai	Grand-Est (East)	Bas-Rhin	Early Bronze Age	2000-1700 BC	
OBE3628	3628		PB	PAEI	Obernai	Grand-Est (East)	Bas-Rhin	Early Bronze Age	2000-1700 BC	
OBE3629	3629		PB	PAEI	Obernai	Grand-Est (East)	Bas-Rhin	Early Bronze Age	2000-1700 BC	
OBE3722	3722		PB	PAEI	Obernai	Grand-Est (East)	Bas-Rhin	Early Bronze Age	2000-1700 BC	

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
OBE3626-1	3626	1	PB	PAEI	Obernai	Grand-Est (East)	Bas-Rhin	Early Bronze Age A2a	1926-1701 cal BC	3505 ± 35 BP
ROS102	102	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
ROS45	45	1999	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4730-4655 cal BC	5833 +/- 29BP
ROS78	78	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
ROS82	82	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4725-4640 cal BC	5816 +/- 28BP
Ros100	100	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4720-4635 cal BC	5789 +/- 33BP
ROS108	108	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4720-4635 cal BC	5787 +/- 33BP
ROS111	111	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4720-4635 cal BC	5791 +/- 33BP
ROS31	31	1999	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
ROS36	36	1999	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
ROS42	42	1999	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4680-4590 cal BC	5762 +/- 33BP
Ros43	43	1999	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
ROS47	47	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
ROS62	62	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
ROS66	66	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
ROS86	86	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
Ros115	115	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
Ros25	25	1999	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
Ros26	26	1999	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	

Part II

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
Ros40B	40 ind B		PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
Ros57	57	1999	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
Ros7	7	1998	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
Ros85	85	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
Ros88	88	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4720-4635 cal BC	5792 +/- 33BP
Ros91	91	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4730-4645 cal BC	5812 +/- 30BP
Ros96	96	2000	PB	Rosheim "Mittelfeld" "Rosenmeer"	Rosheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Grossgartach"	4700-4500 BC	
SCHW432	432		PB	Lotissement Les Terrasses de la Zorn	Schwindratzheim	Grand-Est (East)	Bas-Rhin	Early Neolithic "LBK"	5211-4962 cal BC	6130 +/- 40BP
SCHW72-15	72	15	PB	Lotissement Les Terrasses de la Zorn	Schwindratzheim	Grand-Est (East)	Bas-Rhin	Early Neolithic "LBK"	5383-5220 cal BC	6340 +/- 40BP
SCHW72-16	72	16	PB	Lotissement Les Terrasses de la Zorn	Schwindratzheim	Grand-Est (East)	Bas-Rhin	Early Neolithic "LBK"	5383-5220 cal BC	6340 +/- 40BP
SCHW72-17	72	17	PB	Lotissement Les Terrasses de la Zorn	Schwindratzheim	Grand-Est (East)	Bas-Rhin	Early Neolithic "LBK"	5383-5220 cal BC	6340 +/- 40BP
SCHW9	9		PB	Lotissement Les Terrasses de la Zorn	Schwindratzheim	Grand-Est (East)	Bas-Rhin	Middle Neolithic "Bischheim"	4450-4350 BC	
BERG351-2	351	2	PB		Bergheim	Grand-Est (East)	Haut-Rhin	Iron Age		
BERG643-2	643	2	PB		Bergheim	Grand-Est (East)	Haut-Rhin	Iron Age		
BERG643-1			PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Iron Age "Early La Tène"		
BERG02-2	2	2	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4037-3803 cal BC	5135 +/- 35BP
BERG157-2	157	2	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4300-3900 BC	
BERG157-7	157	7	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4315-4048 cal BC	5335 +/- 35BP
BERG02-3	2	3	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4200-3800 BC	
BERG02-5	2	5	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4000-3800 BC	

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
BERG118-1	118	1	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	3930-3654 cal BC	4970 +/- 40BP
BERG157-1	157	1	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4229-3972 cal BC	5245 +/- 35BP
BERG157-3	157	3	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4300-3900 BC	
BERG157-5	157	5	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4300-3900 BC	
BERG157-6	157	6	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4300-3900 BC	
BERG157-9	157	9	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4300-3900 BC	
Berg34-1	34	1	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4224-3824 cal BC	5190 +/- 40BP
Berg34-2	34	2	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4200-3800 BC	
BERG61	61		PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4200-3800 BC	
BERG79	79		PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4222-3956 cal BC	5205 +/- 35BP
BERG103-2	103	2	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	3943-3710 cal BC	5020 +/- 35BP
Berg157-4	157	4	PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4300-3900 BC	
BERG57	57		PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	4230-3978 cal BC	5255 +/- 35BP
BERG746	746		PB	Saulager	Bergheim	Grand-Est (East)	Haut-Rhin	Middle Neolithic "BORS-Michelsberg"	3941-3666 cal BC	4995 +/- 35BP
BIS130	130		PB	Muehlacker	Bischwihr	Grand-Est (East)	Haut-Rhin	Early Bronze Age A2a	2000-1800 BC	
BIS385	385		PB	Muehlacker	Bischwihr	Grand-Est (East)	Haut-Rhin	Early Bronze Age A2a	2000-1800 BC	
BIS159	159		PB	Muehlacker	Bischwihr	Grand-Est (East)	Haut-Rhin	Early Bronze Age A2a	2000-1800 BC	
BIS358			PB	Muehlacker	Bischwihr	Grand-Est (East)	Haut-Rhin	Early Bronze Age A2a	2000-1800 BC	
BIS382	382		PB	Muehlacker	Bischwihr	Grand-Est (East)	Haut-Rhin	Early Bronze Age A2a	2000-1800 BC	
BIS388-1	388	1	PB	Muehlacker	Bischwihr	Grand-Est (East)	Haut-Rhin	Early Bronze Age A2a	2000-1800 BC	
BIS388-2	388	2	PB	Muehlacker	Bischwihr	Grand-Est (East)	Haut-Rhin	Early Bronze Age A2a	2000-1800 BC	
COL11	11		PB	Jardin des Aubépines	Colmar	Grand-Est (East)	Haut-Rhin	Iron Age "La Tène B"		
COL153A	153A		PB	Jardin des Aubépines	Colmar	Grand-Est (East)	Haut-Rhin	Iron Age "La Tène"		
COL153i	153i		PB	Jardin des Aubépines	Colmar	Grand-Est (East)	Haut-Rhin	Iron Age "La Tène"		

Part II

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
COL11	11		PB	Colmar	Colmar	Grand-Est (East)	Haut-Rhin	Iron Age "La Tène B"		
NIED	1171		PB	Innere Allmende	Niederergheim	Grand-Est (East)	Haut-Rhin	Late Bronze Age IIIb		
RIX15	15		PB	Zac du Petit Prince	Rixheim	Grand-Est (East)	Haut-Rhin	Early Bronze Age A2a	1700-1600 BC	
RIX4	4		PB	Zac du Petit Prince	Rixheim	Grand-Est (East)	Haut-Rhin	Early Bronze Age A2a	1750-1630 cal BC	3415+-40BP
RIX2	2		PB	Rixheim	Rixheim	Grand-Est (East)	Haut-Rhin	Early Bronze Age		
RIX3	3		PB	Rixheim	Rixheim	Grand-Est (East)	Haut-Rhin	Early Bronze Age		
RIX8	8		PB	Rixheim	Rixheim	Grand-Est (East)	Haut-Rhin	Early Bronze Age		
Jeb8	8		PB	Jebsenboden	Sainte-Croix-en-Plaine	Grand-Est (East)	Haut-Rhin	Iron Age "Halstatt"		
CROI11	11		PB	Ste Croix en plaine	Ste Croix en plaine	Grand-Est (East)	Haut-Rhin	Iron Age "Early La Tène"		
CROI1-4	1	4	PB	Ste Croix en plaine	Ste Croix en plaine	Grand-Est (East)	Haut-Rhin	Iron Age "Early La Tène"		
WET370-1	370		PB	Wettolsheim	Wettolsheim	Grand-Est (East)	Haut-Rhin	Iron Age		
WET429			PB	Wettolsheim	Wettolsheim	Grand-Est (East)	Haut-Rhin	Iron Age		
ISM1	St 01 TR04	1	PB	Le Prieuré	Isle-sur-Marne	Grand-Est (East)	Marne	Late Neolithic	2800-2500 BC	
ISM2	St 01 TR04	2	PB	Le Prieuré	Isle-sur-Marne	Grand-Est (East)	Marne	Late Neolithic	2872-2583 cal BC	4130 +/-30BP
CHEP2	1	2	PB	chemin Royat	La Cheppe	Grand-Est (East)	Marne	Late Neolithic	2621-2468 cal BC	4015 +/-35 BP
CHEP3	1	3	PB	chemin Royat	La Cheppe	Grand-Est (East)	Marne	Late Neolithic	2600-2400 BC	
CHEP4	1	4	PB	chemin Royat	La Cheppe	Grand-Est (East)	Marne	Late Neolithic	2600-2400 BC	
LARZ4	4		PB	le champ Buchotte	Larzicourt	Grand-Est (East)	Marne	Early Neolithic "LBK"	5200-4700 BC	
ORC1	sep 3 Q2		PB	les Noues	Orcontes	Grand-Est (East)	Marne	Early Neolithic "LBK"		
ORC2	1 zone 2		PB	les Noues	Orcontes	Grand-Est (East)	Marne	Early Neolithic "LBK"		
Recy1	2309	1	PB	Parc de Référence	Recy	Grand-Est (East)	Marne	Middle/Late Neolithic	3632-3373 cal BC	4710 +/-35BP
SCPG2	2		PB	Gendarmerie	Sainte-Croix-en-Plaine	Grand-Est (East)	Haut-Rhin	Iron Age "Hallstatt - Late Tène"		
SCPG79	79		PB	Gendarmerie	Sainte-Croix-en-Plaine	Grand-Est (East)	Haut-Rhin	Iron Age "Hallstatt - Late Tène"		
BLP10	10		LB	la Plaine	Beaurieux	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		
BLP31	31		LB	la Plaine	Beaurieux	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		
BLP32	32		LB	la Plaine	Beaurieux	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		
BLP68	68	(2-3)	LB	la Plaine	Beaurieux	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
BLP35			LB	la Plaine	Beaurieux	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		
BLP9LF			LB	la Plaine	Beaurieux	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		
BCM137	137		LB	La croix Maigret	Berry au bac	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		
BFM265	265		PB	le fond du Marais	Bucy-le-long	Hauts-de-France (North)	Aisne	Iron Age		
BFT228	228		PB	La fosse Tournise	Bucy-le-long	Hauts-de-France (North)	Aisne	Iron Age "La Tène B"		
BLH447	447		PB	la Heronnière	Bucy-le-long	Hauts-de-France (North)	Aisne	Iron Age "La Tène B"		
BFM262	262		PB	Le fond du Marais	Bucy-le-long	Hauts-de-France (North)	Aisne	Iron Age "La Tène C1"		
CBV95	95		PB	La Bouche-à-Vesle	Ciry_Salsogne	Hauts-de-France (North)	Aisne	Bell Beaker	2574-2452 calBC	3970 +/- 30BP
CCF315-1	315	1	LB	Les Fontinettes	Cuiry-les-Chaudardes	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		
CCF315-2	315	2	LB	Les Fontinettes	Cuiry-les-Chaudardes	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		
CCT99	99		LB	Le champ Tordu	Cuiry-les-Chaudardes	Hauts-de-France (North)	Aisne	Middle Neolithic "Michelsberg"		
MDV188-2-1	188		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV188-3	188		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV188-4			LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV188-6-1	188		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV188-6-2	188		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV188-6-3	188		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV189-2-1	189	2-1	LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV189-2-2	189		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV254			LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV272	272		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV318	318		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		

Part II

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
MDV561	561		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV562V	562		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV563II			LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV563III	563		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV563IV	563		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV93-2	93		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV93-3	93		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV248	248		TO	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV249	249		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
MDV317	317		LB	Derrière le village	Menneville	Hauts-de-France (North)	Aisne	Early Neolithic "LBK"		
VAS524	524		PB	Dessus des Groins	Vasseny	Hauts-de-France (North)	Aisne	Iron Age "Early La Tène"		
VAS557	557		PB	Dessus des Groins	Vasseny	Hauts-de-France (North)	Aisne	Iron Age "Early La Tène"		
VAS75	75		PB	Dessus des Groins	Vasseny	Hauts-de-France (North)	Aisne	Iron Age "Early La Tène"		
VAS79-2	79	2	PB	Dessus des Groins	Vasseny	Hauts-de-France (North)	Aisne	Iron Age "Early La Tène"		
ATT26	26		PB		Attichy-Bitry	Hauts-de-France (North)	Oise	Iron Age		
Att52-1			PB		Attichy-Bitry	Hauts-de-France (North)	Oise	Iron Age		
ATT13			PB		Attichy-Bitry	Hauts-de-France (North)	Oise	Iron Age "La Tène"		
Es107-3		107-3	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es1-1		1-1	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es113-3	ST445 M17 C6	113-3	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es129-1		129-1	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es178-1		178-1	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
Es2-2		2-2	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es537-11	ST219 M24 C9	537-11	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es603-1	ST219 M27 C6	603-1	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es656-1	ST219 M30 C9	656-1	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es72-1		72-1	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es769-1	St219 M38 C17	769-1	LB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es48-8	ST219 M16 C25	48-8	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es97-1	ST445 M16 C6	97-1	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es1018-16	ST 445 M15 C7	1018-6	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es133-1	ST 219 M27 C6	133-1	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es168	ST445 M5 C11	168-1	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es278-1	ST219 M10 C8	278-1	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es42-3	ST219 M25 C13	42-1	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es454-1	ST219 M16 C15	454-1	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es48bis	ST219 M16 C25	48	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es626-27	ST219 M28 C9	626-27	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es637-1	ST219 M29 C5	637-1	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es75-5	ST445 M16 C6	75-5	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es791-1	ST219 M39 C4	791-1	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es90-9	ST219 M28 C9	90-9	PB	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es159-1		159-1	TO	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
Es37-2		37-2	TO	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		

Part II

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
Es74-1			TO	Mont d'Hubert	Escalles	Hauts-de-France (North)	Pas-de-Calais	Middle Neolithic "Micheslberg/Chasséen"		
NHI1			LB	La Haute Île	Neuilly-sur-Marne	Île de France	Seine-Saint-Denis	Mesolithic		
NHI11			LB	La Haute Île	Neuilly-sur-Marne	Île de France	Seine-Saint-Denis	Mesolithic		
NHI13			LB	La Haute Île	Neuilly-sur-Marne	Île de France	Seine-Saint-Denis	Mesolithic		
NHI20			LB	La Haute Île	Neuilly-sur-Marne	Île de France	Seine-Saint-Denis	Mesolithic		
PER1150C		1150c	PB	Les Perrats	Agris	Nouvelle-Aquitaine (South-West)	Charente	Mesolithic	7177-7057BC	
PER3023	46	3023	PB	Les Perrats	Agris	Nouvelle-Aquitaine (South-West)	Charente	Mesolithic	7177-7057BC	
PER3123	120	3123	PB	Les Perrats	Agris	Nouvelle-Aquitaine (South-West)	Charente	Mesolithic	7177-7057BC	
PER503	503	1146c	PB	Les Perrats	Agris	Nouvelle-Aquitaine (South-West)	Charente	Mesolithic	7177-7057BC	
PER1150B	185	1150b	PB	Les Perrats	Agris	Nouvelle-Aquitaine (South-West)	Charente	Mesolithic	7177-7057BC	
PER26	26	1150b	PB	Les Perrats	Agris	Nouvelle-Aquitaine (South-West)	Charente	Mesolithic	7177-7057BC	
BER36	F36		PB	Les Plots	Berriac	Occitanie (South)	Aude	Middle Neolithic "Chasséen"	4338-4172 cal BC	5400 +-30BP
BER37	F37		PB	Les Plots	Berriac	Occitanie (South)	Aude	Middle Neolithic "Chasséen"	4336-4076 cal BC	5390 +-30BP
BER49	F49		PB	Les Plots	Berriac	Occitanie (South)	Aude	Middle Neolithic "Chasséen"	4340-4180 cal BC	5410 +-30BP
SP1	1		PB	Le Champ du Poste	Carcassonne	Occitanie (South)	Aude	Middle Neolithic "Chasséen"	4450-4600 cal BC	5695 +/-40 BP
SP249	249		PB	Le Champ du Poste	Carcassonne	Occitanie (South)	Aude	Middle Neolithic "Chasséen"	4800-4500 BC	
SP386	386		PB	Le Champ du Poste	Carcassonne	Occitanie (South)	Aude	Middle Neolithic "Chasséen"	4700-4850 cal BC	5890 +/-35 BP
SP395	395		PB	Le Champ du Poste	Carcassonne	Occitanie (South)	Aude	Middle Neolithic "Chasséen"	4800-4500 BC	
QUIN234	7035	234	PB	Quinquiris	Castelnaudary	Occitanie (South)	Aude	Bronze Age		
QUIN58	7035	58	PB	Quinquiris	Castelnaudary	Occitanie (South)	Aude	Bronze Age		

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
QUIN59	7035	59	PB	Quinquiris	Castelnaudary	Occitanie (South)	Aude	Bronze Age		
EUG11	Deblais tamisage	11	PB	Dolmen de Saint-Eugène	Laure	Occitanie (South)	Aude	Bronze Age	2028-1878 cal BC	3580 +/- 30BP
EUG1	couloir	1	PB	Dolmen de Saint-Eugène	Laure	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"		
EUG10	Deblais tamisage	10	PB	Dolmen de Saint-Eugène	Laure	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"		
EUG12	Deblais tamisage	12	PB	Dolmen de Saint-Eugène	Laure	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"		
EUG2	couloir	2	PB	Dolmen de Saint-Eugène	Laure	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"		
EUG8	Deblais tamisage	8	PB	Dolmen de Saint-Eugène	Laure	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"	2925-2871 cal BC	4270 +/- 30BP
EUG3	couloir	3	PB	Dolmen de Saint-Eugène	Laure	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"	2942-2877 cal BC	4290 +/- 30BP
EUG5	Deblais tamisage	5	PB	Dolmen de Saint-Eugène	Laure	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"	2874-2621 cal BC	4140 +/- 30BP
FAD9			PB	Dolmen des Fades	Pépieux	Occitanie (South)	Aude	Bronze Age	1893-1700 cal BC	3490 +-30 BP
FAD1			PB	Dolmen des Fades	Pépieux	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"	2911-2705 cal BC	4240 +-30BP
FAD10			PB	Dolmen des Fades	Pépieux	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"		
FAD12			PB	Dolmen des Fades	Pépieux	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"		
FAD2			PB	Dolmen des Fades	Pépieux	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"		
FAD3			PB	Dolmen des Fades	Pépieux	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"		
FAD4			PB	Dolmen des Fades	Pépieux	Occitanie (South)	Aude	Late Neolithic "Chalcolithic"		
GAZEL			TO	Gazel	Sallèles-Cabardes	Occitanie (South)	Aude	Early Neolithic "Epicardial"	5296-5062 cal BC	6215 +-30 BP
PECH5	46101	fouille 46 portoir 273	PB	Pech Maho	Sigean	Occitanie (South)	Aude	Iron Age		
PECH8	47003	portoir 245 fouille 1969	PB	Pech Maho	Sigean	Occitanie (South)	Aude	Iron Age		
PECH10	75135	portoir1062	PB	Pech Maho	Sigean	Occitanie (South)	Aude	Iron Age		
PECH3	71289	Bloc CF portoir 1238	PB	Pech Maho	Sigean	Occitanie (South)	Aude	Iron Age		
PECH9	F64c	depotoir sur sol inf 1.30m portoir 1062	PB	Pech Maho	Sigean	Occitanie (South)	Aude	Iron Age		
BOU1	972	E10-C4b1	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"	3499-3126 cal BC	4590 +-30 BP

Part II

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
BOU10	1009	G09-C4a	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"		
BOU11	912	G15-	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"	3344-3035 cal BC	4485 +-30BP
BOU12	892	G17-C4b	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"		
BOU15	2	H5-10/20 cm	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"	3337-3024 cal BC	4465+-30 BP
BOU4	872	E15-C4a	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"		
BOU6	472	F13-	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"	3623-3370 cal BC	4680 +-30BP
BOU7	1849	F15-C4b	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"		
BOU8	2167	F15-	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"		
BOU9	344	F16-C4a	PB	Aven de la Boucle	Corconnes	Occitanie (South)	Gard	Late Neolithic "Néolithic Récent"		
PON1	1		PB	Tumulus de Pontel	Dions	Occitanie (South)	Gard	Iron Age		
PT2	A2	B18-17	PB	Oppidum du Plan de la Tour	Gailhan	Occitanie (South)	Gard	Iron Age		
MIT10-2032	10	2032	PB	MITRA 3	Garons	Occitanie (South)	Gard	Early Bronze Age	1882-1686 BC cal	3450 +/-35 BP
MIT1031	1031	diag	PB	MITRA 3	Garons	Occitanie (South)	Gard	Early Bronze Age		
MIT11	11	2006	PB	MITRA 3	Garons	Occitanie (South)	Gard	Early Bronze Age	1772-1608 BC cal	3390 +/-35 BP
MIT1030B	11	1030B	PB	MITRA 2	Garons	Occitanie (South)	Gard	Late Bronze Age II	1200-1000 BC	
MIT1059	1059		PB	MITRA 2	Garons	Occitanie (South)	Gard	Late Bronze Age II	1200-1000 BC	
MIT1167A	1167		PB	MITRA 2	Garons	Occitanie (South)	Gard	Late Bronze Age II		3005 +/-30BP
MIT1167B	1167		PB	MITRA 2	Garons	Occitanie (South)	Gard	Late Bronze Age II	1200-1000 BC	
MIT1167C	1167		PB	MITRA 2	Garons	Occitanie (South)	Gard	Late Bronze Age II	1200-1000 BC	
MIT1155	41	1155	PB	MITRA 3	Garons	Occitanie (South)	Gard	Late Neolithic "Neolithique final"	2502-2340 cal BC	3950 +/-35 BP
MIT83A	83	1026	PB	MITRA 3	Garons	Occitanie (South)	Gard	Late Neolithic "Neolithique final"	2471-2235 cal BC	3890 +/-35 BP
CLR01	rem18	X109-CLR03-020-RL75	TO		Le Cailar	Occitanie (South)	Gard	Iron Age		
CLR05	H9	CLR03-SB2084	TO		Le Cailar	Occitanie (South)	Gard	Iron Age "La Tène"		
CLR11	FS2383	88	TO		Le Cailar	Occitanie (South)	Gard	Iron Age "La Tène"		
CLR12	Q22	R9-208	TO		Le Cailar	Occitanie (South)	Gard	Iron Age "La Tène"		
CLR13	US2349	CLR08-Sect4	TO		Le Cailar	Occitanie (South)	Gard	Iron Age "La Tène"		
CLR6	Rem20	M16 R10-566	TO		Le Cailar	Occitanie (South)	Gard	Iron Age		

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
CLR8	H17	CLR03-SB2084 US2064 X101 Rem17	TO		Le Cailar	Occitanie (South)	Gard	Iron Age		
MAND1175	1175		PB	Manduel	Manduel	Occitanie (South)	Gard	Middle Bronze Age 1		
VIGN203	203		PB	Mas de Vignoles IV	Nîmes	Occitanie (South)	Gard	Final Bronze Age IIb		
VIGN3052	3052		PB	Mas de Vignoles IV	Nîmes	Occitanie (South)	Gard	Final Bronze Age IIb		
Sad1	1		PB	Tumulus du Sadoulet	Pompignan	Occitanie (South)	Gard	Iron Age		
Sad4	4		PB	Tumulus du Sadoulet	Pompignan	Occitanie (South)	Gard	Iron Age		
Red1161-64B			PB	Cureboussot	Redessan	Occitanie (South)	Gard	Bronze Age		
Red1041-1160	1041	116	TO	Cureboussot	Redessan	Occitanie (South)	Gard	Bronze Age		
Red1047	1047		TO	Cureboussot	Redessan	Occitanie (South)	Gard	Bronze Age		
Red1058-1519	1058	1519	TO	Cureboussot	Redessan	Occitanie (South)	Gard	Bronze Age		
Red1095	1095		TO	Cureboussot	Redessan	Occitanie (South)	Gard	Bronze Age		
Red1161-1515	1161	1515	TO	Cureboussot	Redessan	Occitanie (South)	Gard	Bronze Age		
Red1161-1528	1161	1528	TO	Cureboussot	Redessan	Occitanie (South)	Gard	Bronze Age		
PEI10	12611	C4-1	PB	Dolmen des Peirières	Villedubert	Occitanie (South)	Gard	Bell Beaker	2894-2678 cal BC	4200 +-30 BP
PEI2	69	C4-1	PB	Dolmen des Peirières	Villedubert	Occitanie (South)	Gard	Bell Beaker	2563-2308 cal BC	3935 +-30BP
PT7	B2	W18- 23-2	PB	Oppidum du Plan de la Tour		Occitanie (South)	Gard	Iron Age		
Cx161	161		PB	ZAC Agora	Cugnaux	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen classique ou récent"		
Cx165	165		PB	ZAC Agora	Cugnaux	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen classique ou récent"	4230-3970 cal BC	5250 +/-35BP
Cx166	166		PB	ZAC Agora	Cugnaux	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen classique ou récent"	4260-4040 cal BC	5325 +/-30BP
Cx19	19		PB	ZAC Agora	Cugnaux	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen classique ou récent"		
Cx20	20		PB	ZAC Agora	Cugnaux	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen classique ou récent"		
Cx30	30		PB	ZAC Agora	Cugnaux	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen classique ou récent"	4330-4040 cal BC	5340 +/-45BP
Cx13	13		PB	ZAC Agora	Cugnaux	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen"	4260-3990 cal BC	5310 +/-40 BP
VTQ14	Q14-1		PB	La Terrasse	Villeneuve-Tolosane	Occitanie (South)	Haute-Garonne	Early Bronze Age		
VT5-7	F5-7		PB	La Terrasse	Villeneuve-Tolosane	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen classique ou récent"		

Part II

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
VT7	7		PB	La Terrasse	Villeneuve-Tolosane	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen classique ou récent"		
VTP4-3	P4-3		PB	La Terrasse	Villeneuve-Tolosane	Occitanie (South)	Haute-Garonne	Middle Neolithic "Chasséen classique ou récent"	4050-3800 cal BC	5154 +/- 45 BP
CRE10-1	10	1	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE11A	11	A	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE11B	11	B	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE11C	11	C	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE14	14		PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4442-4261 cal BC	5485 +-30 BP
CRE20-A	20	A	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE20-B	20	B	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE21	21		PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4356-4258 cal BC	5460 +-30BP
CRE29	29		PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE3	3		PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE30	30		PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4366-4259 cal BC	5475 +-30 BP
CRE34	34		PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4446-4265 cal BC	5495 +-30 BP
CRE5B	5	B	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE5C	5	C	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE5D	5	D	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
CRE7-A	7	A	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4341-4233 cal BC	5415 +-30 BP
CRE8-C	8	C	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4441-4252 cal BC	5475 +-35 BP
CRE20-D	20	D	PB	Le Crès	Béziers	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
PIR3037AB	3037	12B	PB	Rec de Ligno	Valros	Occitanie (South)	Herault	Early Bronze Age		3663+/- 43BP
PIR3116B	3116	11B	PB	Rec de Ligno	Valros	Occitanie (South)	Herault	Early Bronze Age	2336-2135 cal BC	3790 +-30 BP
Pir6	160	6	PB	Le Pirou	Valros	Occitanie (South)	Herault	Early Bronze Age		3356+/- 43BP
PIR3117	3117	10	PB	Rec de Ligno	Valros	Occitanie (South)	Herault	Early Bronze Age		3606 +/- 43 BP

Lab ID	Structure	Individual	Bone type	Site	Location	Region	Department	Culture	Dates BC	Dates BP (C14)
Pir4	116	4	PB	Le Pirou	Valros	Occitanie (South)	Herault	Middle Neolithic "Chasséen"		5494 +/- 49 BP
Pir1	293	1	PB	Le Pirou	Valros	Occitanie (South)	Herault	Middle Neolithic "Chasséen"		5406 +/- 48 BP
Pir2	226	2	PB	Le Pirou	Valros	Occitanie (South)	Herault	Middle Neolithic "Chasséen"	4500-4100 BC	
Pir3	289	3	PB	Le Pirou	Valros	Occitanie (South)	Herault	Middle Neolithic "Chasséen"		5450 +/- 47 BP
Pir7	45	7	PB	Le Pirou	Valros	Occitanie (South)	Herault	Middle Neolithic "Chasséen"		5456+/- 49 BP
PEY163	163		PB	Le Peyrou	Agde	Occitanie (South)	Hérault	Iron Age		
PEY53	53		PB	Le Peyrou	Agde	Occitanie (South)	Hérault	Iron Age		
Pey74	74		PB	Le Peyrou	Agde	Occitanie (South)	Hérault	Iron Age		
PEY73			PB	Le Peyrou	Agde	Occitanie (South)	Hérault	Iron Age		
BES1248	1248	10519	PB	La Monédière	Bessan	Occitanie (South)	Hérault	Iron Age		
BES1096B	1096	10517	PB	La Monédière	Bessan	Occitanie (South)	Hérault	Iron Age		
BES1154	1154	10794	PB	La Monédière	Bessan	Occitanie (South)	Hérault	Iron Age		
BES1249	1249	11159	PB	La Monédière	Bessan	Occitanie (South)	Hérault	Iron Age		
DEV			PB	Barreau de la Devèze Cabrials	Béziers	Occitanie (South)	Hérault	Middle Neolithic "Chasséen"	4457-4143 cal BC	5454 +/- 89 BP
MAZ16	TR22 ST41 US414	16	PB	Mazeran 2	Béziers	Occitanie (South)	Hérault	Middle Neolithic 1	4785-4590 cal BC	6735-6540 cal BP

Chapter II. Sample processing

I. Sample preparation

In this section, I will detail the various steps that were conducted in the present study on ancient samples. The collection of remains we assembled over the course of this project came from collaborations with different archeologists and anthropologist. As some of these were long term collaborations, part of the most recently excavated remains had been handled according to recommendations in Pruvost *et al.* 2007. Still, excavation dates were very heterogeneous in the panel, ranging from the 1980s to the 2000s, which rendered impossible any control over the excavation procedures and post excavation treatments for most samples. DNA was prepared from a variety of bones ranging from long bones such as femurs, ulna and tibia, teeth and temporal bones from individuals between 7,000 and 2,500 years old. Processing of the samples systematically took place in a high containment laboratory located on the 6th floor of the Institut Jacques Monod in Paris, where protective clothing and decontamination of reagents, surfaces and equipment was employed (Champlot *et al.*, 2010). These procedures will be detailed in to following paragraphs.

1. Cutting and grinding

Samples were carefully labeled and photographed upon receipt, and stored in a dedicated storage room of the ancient DNA laboratory. Sample preparation procedures took place in the second room of the facility, within a dedicated Plexiglas chamber that was thoroughly decontaminated with both UV light and bleach. Procedures were slightly modified depending on the type of bone and will be detailed in the following sections. While part of the recovered powder was immediately used for extraction, the remaining bone powder was systematically stored at -20°C to reduce further degradation of the remaining DNA.

(1) Long bone

Long bones are heterogeneous structures that consist of a long shaft with two bulky extremities. While primarily composed of compact bone, they may contain a large amount of spongy bone towards the inside medullar cavity and at the extremities. To maximize

the recovery of DNA from such samples, we focused on the bone cortex, the dense bone located beneath the bone surface along the shaft. The density of the cortex can be variable, both along a single bone and among different samples, depending on their state of preservation. We first defined the sampling area by removing the outer layer of the bone with a clean razor blade, thus eliminating sediments and potential surface contaminants. In poorly preserved bones, we ground the underlying area into powder by low-speed drilling in order to avoid local heating that could damage DNA. Denser bones were cut into fragments with a diamond saw blade before performing cryogenic grinding using a freezer mill as a finer powder was shown to produce higher DNA yield (Rohland and Hofreiter, 2007).

(2) Temporal bone

Depending on the degree of preservation of the bone, the temporal bone can either be part of a partially intact skull or be found separate. The latter scenario is the most likely for a majority of ancient remains, and isolated fragments display a variable proportion of the porous and dense bone parts that composes the temporal bone. The particular structure we aimed at is the dense bone of the otic capsule which is encapsulated in a dense white bone (Pinhasi *et al.*, 2015). From fragmented bones, it can be exposed after scraping the outer bone layer with a razor blade and carefully removing the dense bone on the opposite side to the internal acoustic meatus (At the bottom right of **Figure 15**) through low-speed drilling. The cochlea appears in a yellowish, pearly hue and was cut off the bone piece using a diamond saw blade, then powdered using a freezer mill. As complete skulls can sometimes be fragile or filled with compacted sediments, we found that the easiest and least damaging way to access the inner ear was to carefully drill it out of the ear canal.

Figure 15: Picture of a fragment of the temporal bone

(3) Teeth

The tooth root was bleached and sediments from its outermost surface was gently removed with a sterilized razor blade. A diamond saw blade was then used to separate the crown and the root, and as much pulp and dentine as possible was carefully drilled out of the latter as previously described (see **Figure 16** and Damgaard *et al.* 2015). The remaining outer layer of the roots was then used for the DNA extraction.

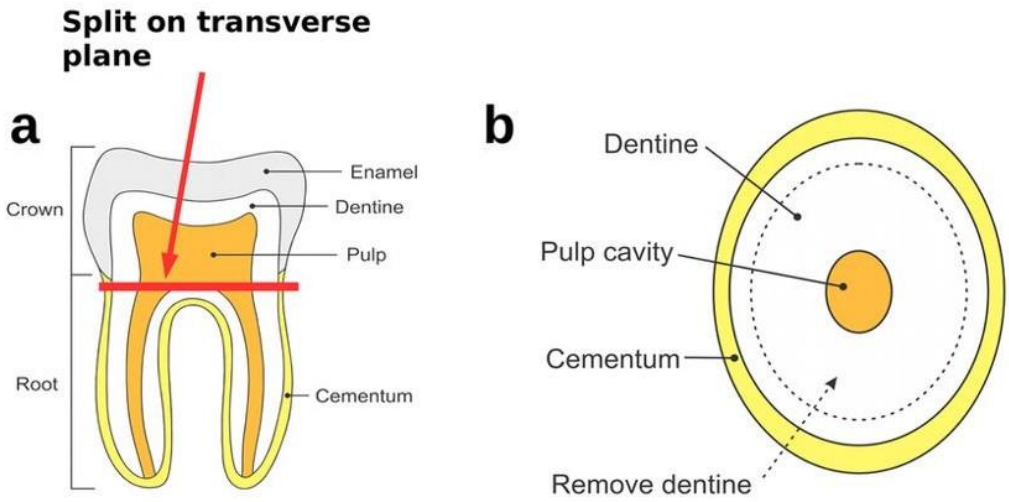


Figure 16: Sampling procedure of a tooth
(Daamgard *et al.* 2015)

2. DNA extraction

DNA extraction can be performed in several ways, and many commercial kits are available to greatly simplify these procedures. Most of these methods were designed to deal with fresh tissue containing intact cells and high molecular weight DNA, and ancient DNA usually fails to meet either of these two criteria. Getting the DNA into aqueous solution can be challenging, owing to its numerous modifications and the composition of ancient remains where cells structures are usually not preserved (Geigl, 2002). Besides, various components from the soil can be co-purified along with DNA. Also, as stated before, ancient DNA presents a variety of damages (Hofreiter *et al.*, 2001; Pääbo, 1989). Thus, in order not to degrade it further during extraction, aggressive treatments such as high temperatures or use of strong detergents should be avoided (Rohland and Hofreiter, 2007). Although these treatments might increase DNA release, they would decrease

overall DNA yield by inflicting further damage to the ancient DNA molecules. Extracting DNA from ancient remains is thus a challenging procedure that faces and needs to reconcile a number of problems, leading to specific optimizations.

We opted for a DNA extraction at 37°C relying on Ethylenediaminetetraacetic acid (EDTA), a chelating agent that sequesters metal ions such as calcium cations (Ca^{2+}), therefore weakening the hydroxyapatite matrix that composes hard tissues such as bones. We coupled EDTA with a source of phosphate, in the form of disodium hydrogen phosphate (Na_2HPO_4) to compete with DNA and prevent it from binding back to the hydroxyapatite upon release. β-mercaptoethanol was added to the extraction buffer as it helps in denaturing proteins by breaking the disulfide bonds between the cysteine residues, but mainly as it was shown to remove the tannins and polyphenols present in crude extract during plant DNA extraction, and these are frequently co-extracted along with DNA during the procedure (Arruda *et al.*, 2017).

Over the course of the project we also opted for a two-steps extraction procedure according to the rationale in Damgaard *et al.* 2015. Briefly, although colonization by microorganisms is likely to be heterogeneous within a bone (Collins *et al.*, 2002), it is expected that during the digestion of bone material with EDTA, surface contaminants would be released into solution first. This also applies to modern human DNA contamination, which would be deposited on the bone surface during sample handling. Endogenous DNA, preserved within local “niches” as I stated before (Geigl, 2002), might be better protected and therefore released after further degradation of the bone matrix after a longer incubation time. Thus, treating the ground bone material with a digestion buffer for a short period of time (a "pre-digestion") would remove a fraction of the exogenous non-target DNA and thereby enrich the DNA extract for endogenous DNA (Damgaard *et al.*, 2015). The addition of a pre-digestion step to the protocol resulted in the use, for the first incubation period, of both a detergent, in the form of N-Lauroylsarcosine sodium salt (NLS), and a proteolytic enzyme (Proteinase K), as they respectively degrade cell structures (here expected to be those of microorganisms) and degrade the proteins that would be released as a result. A schematic presentation that summarizes our extraction procedure can be found **Figure 17**.

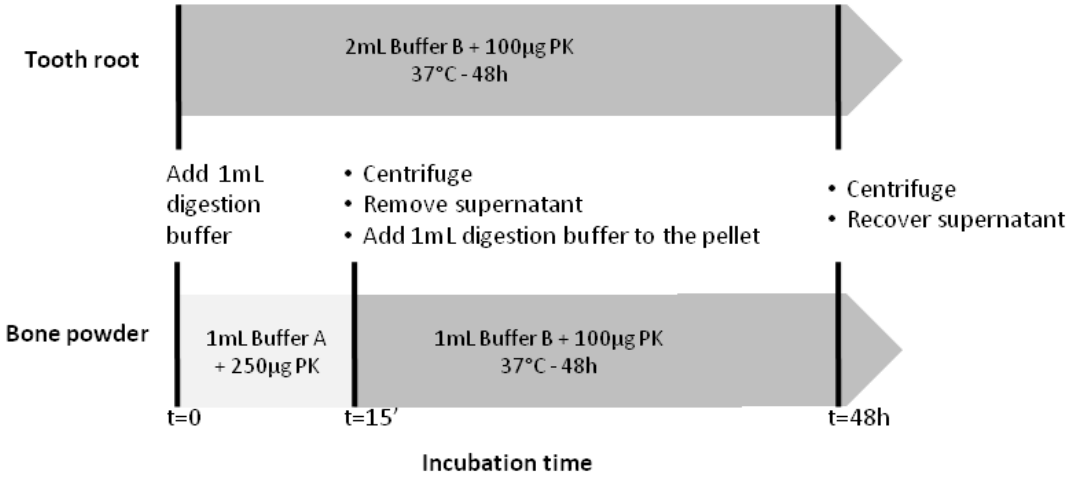


Figure 17: Extraction procedure for different type of samples.
Buffer A is composed of 0.25M Na_2HPO_4, 10mM EDTA, 0.5% NLS, 1% β-mercaptoethanol. Buffer B is composed of 0.25M Na_2HPO_4, 0.5M EDTA and 1% β-mercaptoethanol. PK: Proteinase K

The three extraction buffers used were prepared in small batches prior to the extraction and UV irradiated, then each was processed along with samples as an extraction blank. In total, 24 series of extractions were performed, and one extraction blank per buffer was carried out for each of them.

(1) Long bones and temporal bones

Between 100-250 mg of the recovered powder was incubated in 1 mL extraction buffer following a two-steps procedure: a 15min incubation at 37°C in 0.25 M Na_2HPO_4, 10 mM EDTA, 50mM Tris/HCl, 0.5% NLS, 1% β-mercaptoethanol and 250µg/mL proteinase K, followed by removal of the supernatant and re-suspension of the pellet in 0.5M EDTA, 0.25M Na_2HPO_4, 1% β-mercaptoethanol , 50 µg/mL proteinase K for 48-72h at 37°C.

(2) Teeth

The tooth root was incubated in 2mL 0.5M EDTA, 1% β-mercaptoethanol, 50 µg/mL proteinase K for 48h at 37°C.

3. DNA purification

Upon extraction, DNA molecules are released in the extraction buffer while bone powder and other residues can be pelleted through centrifugation. When purifying DNA, and when dealing with ancient DNA in particular, an ideal method for purification should

at the same time maximize DNA recovery and minimize the co-purification of chemical substances which are known to inhibit PCR. Several methods for DNA purification exist, including phenol extraction followed by ethanol or isopropanol precipitation (Kurosaki *et al.*, 1993), concentration and desalting of DNA using centrifuge filtration columns with defined pore sizes (Hagelberg and Clegg, 1991; Leonard *et al.*, 2000) and binding DNA to silica, either as a suspension (Hofreiter *et al.*, 2004) or as silica spin columns (Yang *et al.*, 1998). The latter was widely adopted by the ancient DNA community, as it performs best when clean DNA is required for downstream applications and when samples contain large amounts of potential PCR inhibitors. It is a two-steps approach where nucleic acids bind to the silica membrane in the presence of chaotropic salts, while polysaccharides and proteins do not bind well and residual traces are removed during alcohol-based wash steps, along with the salts. After binding and washing, nucleic acids are selectively eluted under low-salt conditions, using water or a Tris-EDTA buffer that protects DNA from further degradation after purification.

We chose a single silica-based DNA extraction method resembling the one described by Dabney *et al.* 2013, which was a modification of a protocol optimized for ancient bones and teeth (Rohland and Hofreiter, 2007), widely used in ancient DNA studies. In the Rohland and Hofreiter protocol, DNA was bound to silica, which was added as a suspension together with a binding buffer containing sodium acetate, sodium chloride, and guanidine thiocyanate. The silica particles with the conjugated DNA were desalted by using an ethanol wash buffer, and the DNA was eluted into a low-salt buffer. Dabney *et al.* aimed at equivalently recovering DNA fragments of all sizes, and made the following modifications to the previously described procedure: (i) the use of a binding buffer containing guanidine hydrochloride, and more isopropanol than used in standard procedures; (ii) an increase in the volume of binding buffer relative to that of extraction buffer; and (iii) the replacement of silica suspension with commercially available silica spin columns with a custom-adapted extension reservoir to enable large loading volumes. We also included changes (ii) and (iii) in our procedure, but our DNA binding step relies on the use of guanidine thiocyanate and isopropanol.

Insoluble debris from either the second extraction of bones or the incubation of tooth root were pelleted for 2 min at 13000 rpm, and DNA was purified from the collected supernatant using a method adapted from the QIAquick Gel Extraction purification kit protocol (Qiagen), developed in the lab. The modifications are as follows. A sample

volume of 1 mL was combined with 6 mL of QG binding buffer and 4mL of isopropanol and bound to the silica column on a vacuum manifold. 20 mL tube extenders (Qiagen) which had been submerged in bleach and rinsed with water were fixed to the columns, allowing to pour a larger volume, thus limiting handling. An additional wash with 2mL of QG was added, prior to the PE washing step, whose volume was increased from 750 µL to 2mL. After washing, columns were removed from the manifold and placed back in their respective 2mL collection tube for 1.5 min centrifugation at 13000 rpm on a bench-top centrifuge (Eppendorf). After being dry-spun for another 1.5 min at 13000, the columns were placed in a new Eppendorf tube and DNA was eluted twice in respectively 30 µL and 25 µL of TET buffer (EB buffer (Qiagen), 1 mM EDTA, 0.05% Tween 20) heated to 37°C.

II. Sample characterization

1. Total DNA yield of the extract

After purification, an aliquot of each extract was used for quantification of its total DNA yield. We used a Qubit 2.0 fluorometer (Invitrogen) that relies on a fluorescent dye to determine the concentration of nucleic acid, as the measurement of DNA's natural absorbance of light at 260 nm is often not sensitive enough for the concentration of most extracts. The particularity of this dye is that it only becomes intensely fluorescent upon binding to DNA, remaining at an extremely low fluorescence until then.

2. Inhibitory effect of ancient extracts on DNA amplification

Several classes of substances are known to inhibit DNA polymerases, therefore preventing or delaying the amplification of DNA. Among these, one can for example find humic acid that is present in the soil, but also melanin that can be extracted alongside DNA (Eckhart *et al.*, 2000; Matheson *et al.*, 2010). Inhibition can be assayed using real time-QPCR, by measuring the decrease in the efficiency and the amplification delay when adding several concentrations of ancient sample to modern target DNA, here a region of cow mitochondrial DNA. The PCR reaction was set up as follows: LightCycler Faststart DNA masterplus SYBRGreen I (*Roche*), forward and reverse primers (BB3r: TGCCCCATGCATATAAGCAAG, BB4m: TCACGCGGCATGGTAATTAG) 1 µM each, Faststart Taq DNA polymerase 1.15 u, and water up to 7 µL. This mix was then distributed in capillaries specific to the LightCycler Carousel based system (*Roche*). Then, 30 pg of

modern cow DNA were added to each capillary. I tested the inhibition conveyed by 0.5, 1 and 1.5 µL of ancient sample (and water up to 10 µL final volume) on the amplification of a modern cow mitochondrial fragment. The standard curve was obtained with a set of modern cow dilutions of around 300, 30 and 3 pg respectively. The amplification program was as follows: 10min of initial denaturation at 95°C, 15s at 95°C and 45s at 62°C for 50 cycles for the amplification, and a fusion program to estimate the melting temperature of the products. Results were analyzed using the LightCycler480 software v.1.51.62. and are shown, for four different samples, in **Figure 18**.

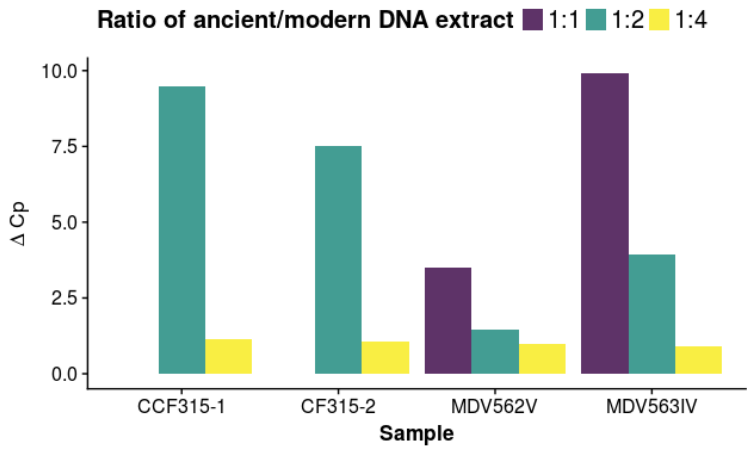


Figure 18: Results of an inhibition test

3. qPCR

The presence of endogenous DNA in a sample can also be estimated by qPCR. Here primers designed to amplify a 143 bp long fragment of human mitochondrial DNA were used, as any longer fragment is less likely to be recovered from ancient samples, or more likely to originate from modern contaminants. For this sensitive application, the PCR was set up in the 6th floor high containment facility. The first step consisted in treating some of the reagents with ethidium monoazide bromide (EMA), a fluorescent photoaffinity label that binds covalently to nucleic acids in solution and destroys it upon photoactivation (Massilani *et al.*, 2016; Rueckert and Morgan, 2007). UNG was added to the PCR mix according to the UQPCR procedure to prevent cross-contamination (Pruvost *et al.*, 2005). Combined with the incorporation of dUTP instead of dTTP during the amplification, it allows a treatment with a heat-labile uracil-N-glycosylase (UNG) to

degrade uridine-containing carryover products from previous PCR amplifications. A treated mix was thus prepared as follows: LightCycler Faststart DNA masterplus SYBRGreen I (*Roche*), Faststart Taq DNA polymerase 1.15 u, UNG 0.1 u according to the UQPCR procedure to prevent cross-contamination (Pruvost *et al.*, 2005), and EMA 1:1. The reaction was then incubated in the dark for 10 minutes, then exposed to light for 6 minutes to destroy the EMA-bound DNA as well as the remaining free EMA. During this intense light exposure, the tube was placed in a cryo-block to prevent the mix from heating, since most of the enzymes are heat-sensitive.

The PCR reaction was composed of the previously described treated mix, UV-treated GC-rich additive 1X, HCR1 (AACCGCTATGTATTTCGTACATTACT) and HCR2 (GGTTGATTGCTGTACTTGCTTG) primers 1 µM and irradiated water up to 9 µL. Then 1 µl of ancient sample was added. Capillaries were then closed and taken downstairs to the main lab in order to prepare additional capillaries containing the modern DNA dilution used to calculate the standard curve. For the amplification, the following program was used: 15 minutes at 37°C to activate the UNG, 10 minutes at 95°C to inactivate the UNG and activate the DNA polymerase, 15 seconds at 95°C followed by 15 seconds at 57°C and 20 seconds at 67°C for 80 cycles of amplification, and a fusion program as described above.

III. Library preparation

All the reaction buffers and enzymes used for the different steps of the library preparation were decontaminated beforehand using EMA. Each mix was decontaminated just prior to use as follows: 0.25 volume of EMA 24 µM was added to the reaction mix in the dark, using a red light. Then the mixture was incubated for 10 minutes in the dark, prior to exposing the Eppendorf tube to a bright light for 6 minutes.

1. DNA repair

The first step in preparing a double-stranded library out of ancient material is to repair miscoding lesions that appeared in the sequence as a result of the deamination of cytosines, leading to the presence of uracil. These ancient DNA signatures manifest as a high rate of C to T or G to A mismatches to the reference genomes after alignment of the sequenced reads, and are a double-edged sword. One the one hand, they are one of the

features expected from ancient DNA, together with a short fragment size, and are thus important to assess its authenticity. On the other hand, these mismatches can bias downstream analyses by being treated as genuine polymorphisms. Removal of these uracils is performed by the USER enzyme (*New England BioLabs*) that is a combination of Uracil-DNA-glycosylase that removes the uracil base, leading to the formation of an abasic site, and an Endonuclease VIII that breaks the phosphodiester backbone at the 3' and 5' of the abasic site. A "partial" treatment protocol was established that causes these characteristic ancient DNA damage to be restricted to the terminal nucleotides, while nearly eliminating it in the interior of the DNA molecules (Rohland *et al.*, 2015). To limit bias in the SNP calling due to ancient DNA damages while still being able to authenticate it, libraries destined to shotgun sequencing were prepared with an initial partial UDG treatment, in order to restrict damage patterns to the terminal bases. No treatment was performed on libraries subjected to in-solution targeted enrichment.

The results for both the absence of USER treatment and the use of a partial one are shown in **Figure 19**.

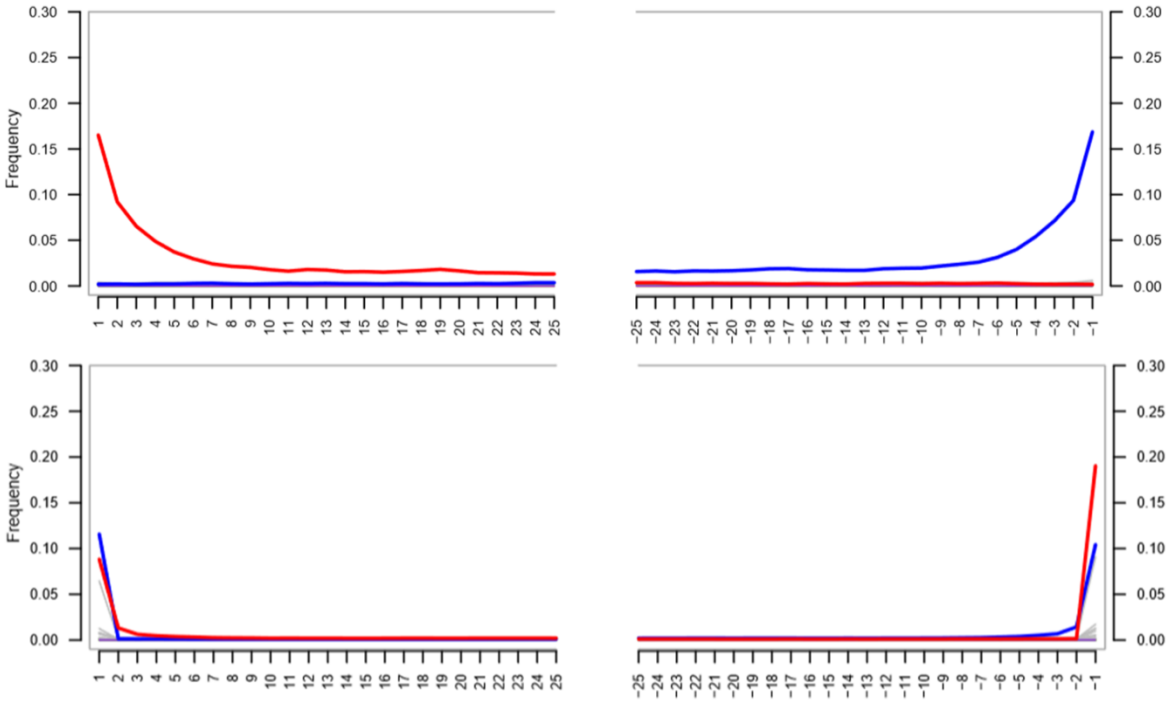


Figure 19: Deamination patterns in an ancient sample.
Results provided from the MapDamage analysis of the same sample without UNG treatment (top) and with a partial UNG treatment (bottom). Frequency of C to T transitions across the 25 terminal bases of each read that are consistent with aDNA damages is shown in red while G to A transitions are shown in blue.

The reaction mixture was set up as follows: 10µL of ancient DNA extract, 6.75µL of a decontaminated mix (3µL NEBNext Repair buffer 10X (New England BioLabs), 1.5µL USER enzyme, pretreated with 2.25µL EMA), and irradiated water up to 28.5µL. Incubation took place at 37°C for 1h, which resulted in. Then each double-stranded DNA fragment was converted to blunt ended DNA, with a 5'-phosphates and a 3'-hydroxyl allowing the ligation of double-stranded adaptors later on, using the NEBNext End Repair module (New England BioLabs). For this step, I just added 1.5µL of a decontaminated End Repair mix (1µL End Repair enzyme mix pretreated with 0.5µL EMA) and incubated at 20°C for 40 minutes. After a purification with a MinElute PCR purification kit (Qiagen) and elution in 2x 16.5µL, DNA was ready for adaptors ligation.

2. Adapter ligation and elongation

These adaptors were conceived from the P5 and P7 adaptors of the TruSeq DNA SamplePrep Kit by adding a unique 7 nt-long index sequence in their middle, thus allowing the identification of a sample in the pool of libraries that is used during the sequencing, but also to ensure that cross-contamination of the libraries did not occur. Prior to ligation, the single-stranded adaptors needed to be double-stranded at the end to-be-ligated (the 3'-OH end of the long adapter oligonucleotide). Such a partially double-stranded adapter was obtained by hybridizing the single-stranded adaptor (66 to 70 nt-long) with a 12 nt-long oligonucleotide complementary to its 3'end (AGATCGGAAGAG). 4µL of each oligonucleotide (100µM) was incubated in 2µL of annealing buffer 10X (10mM Tris, pH 7.5–8.0, 50mM NaCl, 1mM EDTA) and H2O up to 20µL, for a final concentration 20µM, for 3min at 95°C to fully denature any DNA secondary structure, then the temperature was lowered to 50°C, allowing both oligonucleotides to anneal.

For one sample, the ligation mix was composed of 10µL NEBNext Quick Ligation Reaction buffer 5X and 1.5µL Quick T4 DNA ligase (from the NEBNext Quick Ligation Module, New England Biolabs) and 5.75µL 24µM EMA. Decontamination was performed as described before, and 17.25µL of the ligation mix was added to 30µL of purified repaired DNA. Then 1µL of the 20 µM 50N and 70N adapter was added, and the ligation reaction was incubated for 30min at 20°C. The elongation step was used to obtain fully double-stranded adapters and was performed by incubating the ligation product with OneTaq HotStart 2X mastermix with Standard buffer (New England Biolabs) for 20min at 60°C.

3. Library quantification and amplification

The libraries were quantified by qPCR to estimate the initial complexity and the number of amplification cycles required for either DNA enrichment, sequencing or long-term storage. 1µL of a 1/20 dilution of each enriched library was amplified on a LightCycler 2.0 (Roche) using the LightCycler Faststart DNA masterplus SYBRGreen I kit (Roche) P7 and P5 Illumina primers at 1 µM each, Faststart Taq DNA polymerase 1.15 u, and water up to 9µL, for a final reaction volume of 10µL. This mix was then distributed in capillaries specific to the LightCycler Carousel based system. The cycling conditions were as follows: DNA denaturation for 5min at 95°C, then 40 cycles of 95°C for 20 seconds followed by 30 seconds of annealing at 60°C and 30 seconds of elongation at 72°C.

PCR amplification was then carried out as follows: DNA denaturation for 2min at 94°C, then 9 cycles of 94°C for 15 seconds followed by 30 seconds of annealing at 60°C and 60 seconds of elongation at 72°C. The libraries were then purified using NucleoMag NGS Clean-up and Size Select beads using a 1.3X volume (Macherey-Nagel), allowing a size-selection of the molecules longer than 130bp.

The DNA concentration and size distribution in each library was measured by qPCR, with a Qubit 2.0 fluorimeter (Life Technologies) and on a Bioanalyzer (Agilent), respectively.

Chapter III. Mitochondrial genome and nuclear SNPs enrichments

I. DNA capture

Sequences enrichment from both the whole mitochondrial genome and nuclear regions was performed through a capture approach using biotinylated RNA baits obtained through the in vitro transcription of PCR amplified target regions (Gnirke *et al.*, 2009). This approach was subjected to methodological developments for which details are given in **Part III**.

1. Mitochondrial baits synthesis

The capture of the human mitochondrial genome was adapted from the protocol used to capture bovine mitochondrial genomes in the laboratory (Massilani *et al.* 2016). Briefly, a total of 11 pairs of primers were designed using OLIGO6.67 that would generate as many overlapping PCR products, ranging from 1500 to 1800bp, covering entirely 16.5kb-long human mitochondrial genome. Annealing temperature was homogenized across pairs to be around 60°C, and a T3 polymerase promoter was added upstream of either the forward or reverse primer of the pair, enabling us to transcribe and generate complementary probes for both strands. Each PCR reaction was performed in a final volume of 50µL containing 4 ng of human genomic DNA, 1U of FastStart Taq, 1× FastStart buffer, 3mM MgCL2, 200µM dATP, dCTP, dGTP and dUTP, 1 mg/mL BSA, and 1µM of one of the 22 primer pairs (**Supplementary Table 1**) PCRs were performed on an Eppendorf MasterCycler epGradientS as follows: polymerase activation at 95°C for 10 min, 45 amplification cycles (95°C for 20 sec, 58°C for 30 sec, 72°C for 3 min), and a final extension step at 72°C for 5 min. The amplification of each fragment was visualized by agarose gel electrophoresis in a 1 % agarose gel (100 volts for 30 min), and the products were purified using the QIAquick PCR purification kit. The purified mitochondrial fragments were then quantified on a Nanodrop ND-1000 spectrophotometer. The products were pooled in equimolar amounts in two different lots: one containing the fragments with the T3 RNA polymerase promoter sequence at the 5′ end of the forward strands, and the other with fragments containing the T3 RNA polymerase promoter sequence at the 5′ end of the reverse strands.

2. Nuclear baits synthesis

For nuclear markers, in order to be able to capture independently both DNA strands without the necessity of two primer pairs per marker, we used a two-step PCR approach to attach the T3 promoter at the 5' of either the forward or the reverse strand via a 5bp-long linker sequence that differs on both sides. Results of the amplifications can be found in **Supplementary Table 2.**

The amplification of each marker was performed using qPCR on a LightCycler480 (Roche), in a 384-well plate. Two wells were dedicated to the sample amplification, and two other wells were negative controls where genomic DNA was replaced by nuclease free water. The PCR mix was as follows: LightCycler480 SYBRGreen I master 1X, 4ng of a male reference genomic DNA, and nuclease free water up to 7µL. The mix described above was distributed between the different wells using an Eppendorf robot, then 3µL of a 1.7µM primer dilution was added to the corresponding well. Cycling conditions were as follows: 8 minutes denaturation and Taq activation at 95°C, then 40 cycles of 15 seconds denaturation at 95°C followed by 45 seconds annealing and extension at 60°C. One microliter of each previously amplified product was transferred to a new Eppendorf tube to form a pool, and the volume was brought up to 100µL to obtain a 1/100 dilution of each individual product. Then, each individual product was diluted up to 1/100 000 in the final pool. The re-amplification step brings the T3 promoter upstream of either the forward or the reverse strand. In each case, the other primer of the pair brings a custom sequence, carefully designed not to be complementary to the sequence of any of the Illumina adapters used to prepare the ancient DNA libraries and to be uncommon, if not absent, throughout the human genome. Initial denaturation and polymerase activation took place for 5min at 95°C. The first five cycles of amplification consisted of 30 seconds denaturation, followed by 30 seconds annealing at 45°C and 60 seconds elongation at 60°C. Then for 10 cycles, annealing and elongation were performed together using a single 90 seconds step at 60°C.

In vitro transcription of the mitochondrial and nuclear pools, using as template the amplicons that harbor a T3 promoter upstream of either the forward strand or the reverse strand, took place at 37°C overnight. Four reactions were therefore set up as follows: T3 enzyme mix 1X, Transcription buffer 1X (Life Technologies), NTPs 0.5µM each with a 1 to 1 ratio of biotin-14-CTP/CTP (NTPs and biotin-14-CTP Life Technologies), template DNA

between 100ng (nuclear fragments) and 500ng (mitochondrial fragments) and nuclease free water up to 40µL. A the end of the transcription, this DNA template was digested by incubating the transcription reaction with 2µL of TURBO DNase I for 60 minutes at 37°C. Transcription products were purified by performing a phenol-chloroform extraction followed by ethanol precipitation. The quantity of RNA was measured before and after purification with a Qubit 2.0 fluorimeter (Life Technologies). Fragmentation of the long mitochondrial fragments took place in a magnesium rich buffer at 94°C for 4 minutes (NEBNext Magnesium fragmentation module).

3. Blocking oligos synthesis

Blocking oligos were obtained through the amplification of each N50X and N70X barcoded sequencing adapters of the NexteraXT series (Illumina) with the T7 RNA polymerase promoter sequence at the 5′ end of the forward primer. (N50X adaptor blocking oligos: 5′ATGTAATACGACTCACTATAGGGAATGATACGGCGACCAC3′ and 5′GGAAGAGCGTCGTGTAGG3′; N70X adaptor blocking oligos: 5′ATGTAATACGACTCACTATAGGGAGATCGGAAGAGCACACG3′ and 5′CAAGCAGAAGACGGCATAC3′. Each PCR was carried out in a 90µL final volume reaction containing 1nM of Nextera barcoded adapter, 1U Faststart Taq DNA polymerase, 1X Faststart buffer, 90µM of a mix of dNTPs, 1mg/mL of BSA and 0.5µM of each N50X and N70X adaptor blocking oligo primers. Cycling was performed on an Eppendorf MasterCycler epGradientS using a polymerase activation step at 95 °C for 10 min, followed by 25 amplification cycles (95 °C for 20 sec, 60 °C for 30 sec, 72 °C for 45 sec), and no final extension step. PCR products were visualized on a 2 % agarose gel (100 volts for 30min), and then quantified using a Qubit 2.0 Fluorimeter. Two equimolar pools were assembled from N70X and N50X PCR products respectively, and purified using the Qiagen Gel extraction protocol and eluted with 50 µL of EB (Qiagen). Each purified pool was quantified using a Qubit 2.0 Fluorimeter, and 50 ng were used as a DNA template in a 40 µL MAXIscript T7 in vitro transcription reaction (Ambion) containing 1 mM of each NTP. After an overnight incubation at 37 °C, a 1h TURBO DNase treatment was carried out at 37 °C to remove the remaining DNA template. The DNase, along with unincorporated nucleotides and buffer components, was removed through phenol-chloroform purification followed by ethanol precipitation. Purified transcripts were visualized on a 2

% agarose gel (100 volts for 30min), and quantified using a Nanodrop ND-1000 spectrophotometer.

4. Hybridization and hybrid pull-down

The capture is performed according to a published protocol (Massilani *et al.*, 2016), with subtle changes added over the course of the study. It consisted of either one or two steps of hybridization with a PCR amplification in between. The hybridization of the biotinylated RNA baits to the DNA took place in an Eppendorf MasterCycler epGradientS set up to 62°C, and consisted in mixing the following reagents while keeping a small volume (around 25 to 30µL): between 100 and 200ng of input DNA library, denatured for 5 minutes at 95°C beforehand, an equivalent amount of RNA baits, a five times excess of blocking oligonucleotides to prevent unspecific hybridization of the baits to the library adapters, and half the reaction volume of a 2X hybridization buffer prepared as follows: 10X Saline-Sodium Phosphate EDTA (SSPE) buffer, 10 mM EDTA, 0.2% Tween-20. Each reaction was then mixed by pipetting and incubated at 62°C for 48 h.

The capture of the RNA-DNA hybrids was performed using streptavidin-coated magnetic beads: either Dynabeads M-280 Streptavidin (Life Technologies) or Dynabeads® MyOne™ Streptavidin C1 ThermoFisher Scientific). 50/30 µL of beads per sample were washed three times in 1.5mL of 1 M NaCl, 10 mM Tris-HCl (pH7.5), 1 mM EDTA, 0.01% Tween-20, then re-suspended in the previous buffer and Denhardt's solution (0.5 X final concentration) and added to the hybridization mix and incubated 40 min at room temperature on a heatblock with constant agitation to prevent sedimentation of the beads at the bottom of the well. After the bead binding step, 5 post-hybridization washes were performed. The first two, lasting 15 and 10 min respectively, were performed at room temperature with 200µL of 5XSSC, 0.1% Tween 20. The three remaining washes were performed with 200µL 1X SSC, 0.1% Tween 20 for 5 min, the first one at room temperature, the last two washing steps at 62°C, using a pre-warmed washing buffer. For the second round of capture, the last washing step was performed using 0.5X SSC, 0.1% Tween 20 buffer for an increased stringency. To break the biotin-streptavidin bonds and release DNA fragments, the enriched libraries were eluted in EB buffer (Qiagen) complemented with 0.05% Tween 20 and heated at 95°C for 5 minutes, except for the earliest captured samples for which the elution was performed twice using 50µL

0.1M NaOH, and subsequent purification with QIAquick Gel Extraction kit (Qiagen) following manufacturer's instructions.

II. Library quantification and amplification

Eluted enriched libraries are quantified using qPCR to determine the number of amplification cycles required to either perform a second round of capture, or sequence and long-term store them. (See III.3. for the protocol)

Chapter IV. Sequencing and data processing

I. Sequencing

The libraries were sequenced on Illumina MiSeq and NextSeq platforms (75bp paired-end reads) at the Institut Jacques Monod and the Institut de Recherche Biomédicale des Armées in Brétigny-sur-Orge.

II. Paired-end reads trimming, merging and filtering

Adapter sequences were trimmed and overlapping paired-end reads were merged using LeeHom with the --ancientdna option (Renaud *et al.*, 2015). Cutadapt was then used to perform quality trimming with a threshold of 10 (-q 10), and subsequently filter out reads shorter than 30 bases (--minimum-length 30)(Martin, 2011).

III. Mapping and filtering

To screen sequenced libraries for their proportion of endogenous DNA, merged reads were aligned to the human 1000 Genomes project build hs37d5 with decoy contigs. (**Supplementary Table 3**)

To assess the efficiency of the mitochondrial genome capture, merged reads were mapped with BWA version 1.2.3 (Li and Durbin, 2009) against the revised Cambridge reference sequence (rCRS) (Andrews *et al.*, 1999) for human mitochondrial DNA with a duplication of its first 100 bases at the end to ensure mapping of the reads overlapping the junction resulting from the virtual linearization of the circular mitogenome. Consensus sequences were obtained using the software ANGSD v0.910, relying only on reads with base quality above 20, mapping quality above 30 and with more than 3x coverage. Mitochondrial haplotypes were determined based on the PhyloTree phylogenetic tree (van Oven Mannis and Kayser Manfred, 2008) build 17 and by using the HAPLOFIND web application (Vianello *et al.*, 2013), and Phy-Mer (Navarro-Gomez *et al.*, 2015). Unexpected or missing mutations were visually inspected in Geneious version R6 (Kearse *et al.*, 2012) to check if they could be results of misincorporations in low coverage regions. When the coverage was insufficient for reliable consensus calling, the BAM file was inspected visually in order to detect SNPs that were supported by reads with a total

coverage of 3x were reported and used for haplogroup frequency bases approaches only (**Supplementary Table 4**).

Merged reads from enriched libraries were also mapped to the human 1000 Genomes project build hs37d5 with decoy contigs. Target regions were summarized in a file, allowing their extraction using the mpileup tool from the SAMTools toolkit (v. 1.2). Proper variant calling file was generated using bcftools 1.2's call function (Li, 2011). Y chromosome haplogroups were determined using the Yleaf software (Ralf *et al.*, 2018) (**Supplementary Table 5**).

IV. Contamination and authenticity

Sequences authenticity was estimated on libraries subjected to either a partial or no USER treatment using MapDamage (Ginolhac *et al.*, 2011). Mitochondrial contamination was estimated after sequencing using schmutzi (Renaud *et al.* 2015), while we used heterozygous transversion sites on the X chromosome of male samples were used to obtain an estimate of nuclear contamination for those samples using ANGSD (Korneliussen *et al.* 2014). This analysis excluded the pseudoautosomal region of the X chromosome and a minimum base and mapping quality threshold of 30 was applied while no coverage filter was used. We report the result of ANGSD's method 1 and 2 in **Supplementary Table 8**.

V. Sex determination

Shotgun sequencing data also provides with an accurate identification of the biological sex of ancient remains, an information that is often important to archeologists as it is not always retrievable from fragmentary remains or juvenile individuals. Here sex assignment was performed by computing the number of alignments to the Y chromosome (nY) as a fraction of the total number of alignments to both sex chromosomes (nx + nY) according to a method published by Skoglund and colleagues. (Skoglund *et al.*, 2013). The ratio of these two values is therefore RY = nY/ (nx + nY). Assuming that each sequence read from the sex chromosomes is an independent draw from two possible outcomes (Y- or X-chromosome), a 95% confidence interval (CI) was computed by a normal approximation as RY 1.96 RY (1 - RY)/ (nx + nY).

Results across all studied individuals can be found in **Supplementary Table 7**, and an example of the use of this method is illustrated in **Figure 20** on the site of Bergheim in Alsace. This sepulture corresponds to a silo, a circular pit containing the remains of 8 complete individuals along with the skull fragment of an immature and a minimum of seven severed arms (not included in our panel). We successfully extracted DNA from 7 out of 8 individuals, which consisted in both adults (individuals 1,3 and 7) and immature individuals (individuals 2, 5, 6 et 9). Since these individuals are thought to have been deposited simultaneously (to the possible exception of individual 1), the results were expected to shed light on their possible relatedness (see Part. II for mitochondrial and Y chromosome haplogroups). Moreover, genetics provided the only opportunity to assign a sex to four of these individuals, since sex determination could not be performed from osteological remains of immature individuals. Individuals 2 and 6 were assigned as females, while 5, 7 and 9 were unambiguously males. Due to a lower coverage, individual 1 and 3 were found to be consistent with males (respectively 27/11355 and 50/21029 unique reads mapping to the Y chromosome). One was inconsistent with the osteological sex assignment (individual 1), as it was identified as a female, while individual 3 was identified as a male. Through our shotgun sequencing, we therefore report unprecedented sex assignment for the 4 immature individuals of this particular structure.

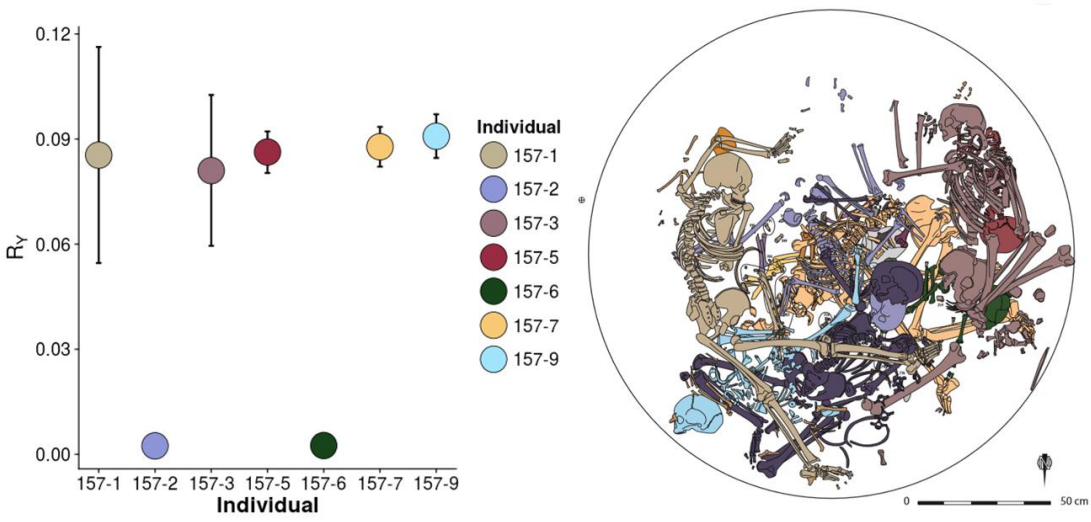


Figure 20: Sex determination of 7 individuals from the Neolithic tomb 157 of Bergheim

Chapter V. Population genetics analyses

I. Capture data

1. Population clustering

The location of individuals with complete mitochondrial genome data used for this study, including our newly sequenced samples, is shown in Figure 30. We combined our 6 Mesolithic hunter-gatherers with published Holocene hunter-gatherers from Western Europe (after 10,000 BCE) (Fu *et al.* 2016, Posth *et al.* 2016, Jones *et al.* 2015). The remaining dataset was split in five time periods: namely the Early Neolithic (5500–4500 BCE), Early Middle Neolithic (4500–4000 BCE) and Late Middle Neolithic (4000-3500 BCE), sometimes combined into one Middle Neolithic group, Late Neolithic (3500-2500 BCE), Bronze Age (2500-800 BCE) and Iron Age (800-25 BCE). This temporal breakdown combined with geographical groups depending on the sampling gave rise to the following groups for our data: Ros (Rosheim, Alsace), for the Early Middle Neolithic, Esc (Escalles, Pas-de-Calais), Mch (Michelsberg culture, Northern France) Brg (Bergheim, Alsace) for the Late Middle Neolithic, SFMN (Middle Neolithic, Southern France), SFLNCA (Late Neolithic/Chalcolithic, Southern France), and NFLN (Late Neolithic, Northern France) for the Late Neolithic, FBA for Bronze Age and FIA for Iron Age. (**Table 1**). As for published data, we split the 638 individuals according to geography: Anatolia, Iberia, Great Britain, Central Europe and the Pontic Steppe. These geographical groups were then further split in the same time periods mentioned above. The date interval covered by individuals within these groups was approximately 300 years on average. Further sub clustering was performed for the Late Neolithic period and the transition towards the Bronze Age according to existing population structures highlighted by nuclear data, thus separating Late Neolithic (LN), Chalcolithic (CA) and Bell Beaker (BB) associated individuals in Central Europe. The clustering was as follows:

- Early Neolithic Iberia (IEN) and Central Europe (CEN), the latter comprising Danubian cultures from Hungary (Starcevo, Koros, ALPc, LBKT) and LBK from Germany.
- Middle Neolithic Great Britain (GMN) and Central Europe, the latter divided in two groups: CEMN (“Early Middle Neolithic”) and CLMN (“Late Middle Neolithic”)

comprising individuals from Hungary (Lengyel, Sopot, Tizsa, Vinca) and Germany (Rössen) respectively.

- Late Neolithic Central Europe (CLN) that comprises Late Neolithic individuals from Germany and Corded Ware associated individuals from Czech Republic.
- Late Neolithic/Chalcolithic Iberia (ILNCA) and Central Europe (CCA), the latter comprising individuals from Chalcolithic Hungary.
- Bell Beaker Great Britain (GBB), Iberia (IBB) and Central Europe (CBB), the latter comprising individuals from Czech Republic, Germany, Hungary, Poland, Switzerland and the Netherlands.
- Bronze Age Central Europe (CBA) comprising individuals from Czech Republic and Germany.

This dataset is complemented with Western Hunter-Gatherers (HG), Anatolian Early Neolithic Farmers (ANE) and Bronze Age groups from the Pontic Steppe (SBA) (**Supplementary Table 6**).

2. Haplogroup frequency-based tests

To identify clusters among prehistoric populations with similar haplogroup composition, the frequency of 20 haplogroups that were found both in our dataset and in the ancient comparative dataset were estimated. Mean frequency and 95% confidence intervals were obtained for each group from 10,000 resampling of count data tables. Results for the overall French population from the Mesolithic to the Bronze Age are shown **Figure 35**.

To unravel groups with similar haplogroup composition, we performed cluster analysis with the different French groups and all comparative dataset based on the haplogroup frequencies of the corresponding datasets. Principal component analysis was performed in PAST (Hammer *et al.* 2001), plotted in R 3.3.3. Euclidean distance was estimated from haplogroup frequencies, and Ward's clustering was performed on the distance matrix using pvclust package in R 3.3.3 (Suzuki *et al.* 2014) using 10000 bootstrap replicates to assess significance through an unbiased p-values.

To assess the probability that the different groups are drawn from the same metapopulation, a statistical significance test, the Fisher's exact test was performed on

absolute frequencies, using the package fisher.test in R 3.3.3 (Fisher *et al.* 1922) on the complete series of pairwise populations.

3. Sequence-based analyses

Pairwise fixation indexes were computed in the Arlequin 3.5 software (Excoffier *et al.* 2010)) using the complete mitochondrial genomes of the ancient individuals. To reduce the chance of incorporating erroneous SNPs, we first trimmed the first and last 30 nucleotides from all of the consensus mtDNA sequences, because these regions were ambiguously sequenced in some samples from the compiled dataset. We computed the number of pairwise differences between the populations using Tamura and Nei's substitution model with 10000 permutations and p-value of 0.05. We performed post-hoc adjustment of the p-values following the Benjamini-Hochberg procedure to correct for the false discovery rate (Benjamini *et al.* 1995). This was achieved using the p.adjust function in R 3.3.3. F_{ST} values were used for multidimensional scaling (MDS), performed using the cmdscale function in R 3.3.3.

To examine the correlation between F_{ST} and geographic or temporal distance between groups, a distance matrix was computed on averages dates and geographical coordinates and a Mantel test was performed using the mantel.rtest function R 3.3.3, with 10,000 bootstrap replicates.

II. Shotgun data

To analyze the newly sequenced samples in context of modern and prehistoric European variation, we merged them with the published ancient individuals that we processed alongside them from the raw sequencing data (see **Supplementary Table 6** for the published dataset) as well as genotyping data from the modern populations of the Human Origins panel (Lazaridis *et al.* 2016). We only used published data obtained from UNG treated libraries. Base quality was rescaled over the whole ancient dataset with MapDamage to account for the probability of each transition to be a damage signature (Ginolhac *et al.* 2011). According to their MapDamage profile, we rescaled most samples over the three terminal bases, with the notable exception of MA1, ElMiron and Bichon, respectively rescaled over 5, 5 and 10 bases.

For this ancient dataset, genotype was called for the 1.2 million SNPs broadly used in aDNA in-solution capture procedures (Mathieson *et al.* 2015) A single allele was drawn at random for each position (minimum mapping and base quality of 30) using PileupCaller (https://github.com/stschiff/sequenceTools), rendering the individuals from the dataset homozygous for each locus. For genotyping data from modern individuals of the Human Origins panel, heterozygous sites in the eigenstrat file were randomly recoded as homozygous for the reference or the alternative allele using an awk script. The two files were then merged using plink v1.9 (Purcell *et al.* 2007)

We also merged this dataset with a subset of individuals extracted from Lazaridis *et al.* 2016 that were genotyped over the same positions: Kostenki14, KK1, SATP, mota, S_Mbuti-1, S_Mbuti-2, S_Mbuti-3 and Ust Ishim that were used as outgroups. We extracted alongside them individuals belonging to the following populations: Levant_N, Iran_N and Natufian (Lazaridis et al, 2016).

1. Population clustering

After mildly filtering the dataset for genotyping rate and linkage disequilibrium with plink v1.9 using the parameters –indep-pairwise 200 25 0.7, --geno 0.95 and –mind 0.95 (Purcell *et al.* 2007), principal component analysis was conducted where ancient individuals were projected onto the first 2 components of the PCA defined by modern Western Eurasian (lsqproject option from the smartpca package). The modern day populations from the Human Origins data set used here are: French, Druze, Sardinian, Palestinian, Orcadian, Russian, Italian_North, Basque, Adygei, Saami_WGA, Bulgarian, Hungarian, Lithuanian, Iranian, Syrian, Lebanese, Jordanian, Saudi, Balkar, Georgian, North_Ossetian, Chechen, Abkhasian, Armenian, Lezgin, Nogai, Kumyk, Belarusian, Ukrainian, Estonian, Mordovian, Czech, Icelandic, Greek, Scottish, English, Spanish, Spanish_North, Finnish, Canary_Islander, Croatian, Norwegian, Sicilian, Italian_South, Turkish, Albanian, Cypriot, Lebanese_Christian, Lebanese_Muslim, Iranian_Bandari, Romanian, Assyrian.

2. Admixture

We ran ADMIXTURE, a maximum likelihood method of estimating the ancestries of individuals, on the whole Human Origins dataset and all ancient (Alexander *et al.* 2009.

The data was pruned for linkage disequilibrium using plink v1.9 (--indep-pairwise 200 25 0.4) and a genotyping rate of at least 5% (--geno 0.95, --mind 0.95) (Purcell *et al.* 2007). Unsupervised ADMIXTURE was run for K=6 to 13 and the lowest cross validation error was obtained for K=11. Clustering results were plotted in R. Results on the full dataset for the different values of K are shown in **Supplementary Figures 4 and 5.** In addition to this unsupervised approach, we estimated the proportion of ancestry attributable to Anatolian Neolithic farmers, Western hunter gatherers and Steppe herders by running a supervised ADMIXTURE where Anatolia_Neolithic, WHG and Yamnaya_Samara were used as source populations.

3. f3-statistics

We used f3-statistics as introduced in Patterson *et al.* 2012 to measure shared drift among ancient individuals and populations relative to an outgroup. (Patterson *et al.* 2012). An f3-statistics of the form f3(A: B, C) measures the covariance if the difference in allele frequencies between populations A and B and A and C across various genomic loci. It is commonly used to test whether a target population A is the result of the admixture of the two others. If A is chosen as an outgroup population (African Mbuti hunter-gatherers in our case), the value of the statistics when different from zero instead reflects the length of shared drift between populations B and C. By setting B as each of our ancient individuals/populations and C to all other modern and ancient populations from the literature, the magnitude of the f3-statistics indicates which populations are closer to one another. In order to assess significance, standard error was computed using weighted blocks of jackknife over 5Mb blocks to compensate for heterogenous coverage of the different positions. To estimate how many standard errors the statistics deviates from 0, Z-scores are calculated, with statistical support for values strictly above 3 (or below -3). Both f3 and D-statistics (see below) were computed using Admixtools (Patterson *et al.* 2012).

4. D-statistics

D-statistics of the form D (A, B:C, D) represent a model-based approach for detecting gene flow between populations using multiple genomic loci, by testing for violation of the proposed tree topology. By setting A as outgroup (see above), the D statistics tests for the genetic similarity between population B and the clade formed by C and D. In a case where

B indeed branches out of (C, D) or shares an equivalent number of alleles with both C and D, the proposed tree topology holds and the statistics equals zero. In cases where B shares significantly more alleles with C or D, the statistics deviates from zero and becomes either negative (if B shares more alleles with C) or positive (if B shares more alleles with D).

Part III. Methodological developments

Chapter I. Ancient DNA capture

I. Generalities about the capture

The development of NGS technologies and optimization of ancient DNA library sequencing protocols increased the number and the success rate of paleogenomic studies over the years, but even though sequencing prices have been lowering, costs are still prohibitive. Indeed, the fact that most aDNA extracts show a minority of endogenous templates resulted in the necessity of huge sequencing cost, and has led to the development of enrichment approaches aimed at reducing sequencing costs and improving the sequence quality of the targeted loci.

Primer extension capture was the first of such methods and succeeded in recovering full mitochondrial sequence information from five Neandertal specimens (Briggs *et al.*, 2009) and from an approximately 30 kyr old modern human (Krause *et al.*, 2010). It was later superseded by other in-solution enrichment methods relying on biotinylated baits, either designed from known sequences and manufactured commercially (Ávila-Arcos *et al.*, 2011), or prepared from modern DNA extracts (Maricic *et al.*, 2010). These baits are subsequently used to target complementary library inserts (Gnirke *et al.*, 2009). Targeted enrichment by capture has become a major tool in many ancient DNA laboratories, as it reduces the sequencing effort per sample while retrieving a great deal of valuable information. This way, complete mitochondrial genomes, exomes and even genomes have been produced (Carpenter *et al.*, 2013; Castellano *et al.*, 2014; Krause *et al.*, 2010). Although all these methods achieve their goals, they often require specific equipment and involve probes that have to be purchased from manufacturers at substantial costs, as they cannot be directly regenerated in the laboratory.

The procedure we developed is summarized **Figure 21** and is directly inspired by Maricic *et al.* 2010. Briefly, it consists in using PCR products to generate probes that will capture targets for sequencing from pooled sequencing libraries of multiple individuals, using standard laboratory equipment. The procedure varies slightly from the original Maricic protocol and between mitochondrial and nuclear probes, and will be detailed further in the following pages.

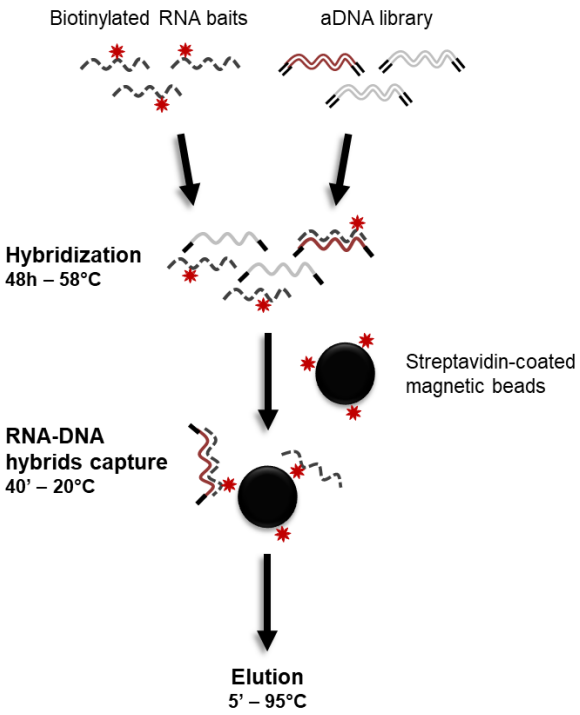


Figure 21: Summary of the enrichment procedure used during this project

II. Marker selection and primer design

1. Complete mitochondrial genomes

Upon starting this project, mitochondrial primers had already been designed by Laurent Cardin, a former master student from the laboratory, but optimization of the PCR conditions had to be performed. For each of the 22 PCR reactions, we optimized both hybridization temperature, as sequence length and composition vary across the mitochondrial genome, and $MgCl_2$ concentration, as it is a cofactor for the Taq polymerase while also binding to dNTPs and helping their proper incorporation in the newly generated sequence. Results are shown in **Table 3.**

Table 3: Optimized PCR conditions for the amplification of the complete mitochondrial genome

Primer ID	Amplicon length (bp)	Annealing temperature (°C)	Conc. MgCl (mM)
Hs-800-2500	1777	58 (58)	3 (3)
Hs-2400-4200	1811	58 (58)	4 (3)
Hs-4100-5700	1599	58 (58)	5 (4)
Hs-5400-7100	1618	58 (58)	3 (3)
Hs-6800-8500	1741	56 (58)	3 (3)
Hs-8200-9900F	1584	58 (58)	4 (3)
Hs-9800-11400F	1626	58 (58)	3 (3)
Hs-11300-12800	1525	58 (58)	5 (3)
Hs-12700-14200	1547	58 (58)	3 (3)
Hs-14200-15700	1584	58 (58)	3 (3)
Hs-15700-800	1728	58 (58)	3 (3)

Optimized annealing temperature and MgCl concentrations are given for each pair of primers where the T3 promoter is in 5' of the forward primer. Bracketed values correspond to optimal conditions for pairs that harbor the T3 promoter upstream of the reverse primer.

2. Nuclear regions of interest

(1) Phenotypic and ancestry-informative markers (AIM)

I selected in the literature 47 single nucleotide polymorphisms (SNP) considered to be relevant to answer archeological questions for both the Mesolithic-to-Neolithic and the Neolithic-to-Bronze Age transitions. Among them, markers giving information about phenotypic traits such as eye and hair color (Fortes *et al.*, 2013; Walsh *et al.*, 2013), but also physiological adaptations, mostly related to changes in diet (Hancock *et al.*, 2010; Hervella *et al.*, 2015; Krüttli *et al.*, 2014; Tishkoff *et al.*, 2007), and continental origin (Gettings *et al.*, 2014) can be found (**Table 4**).

Table 4: List of SNPs associated ancestry, physiological and physical trait

Gene	SNP	Alleles	Phenotype
CSMD1	rs10108270	C/A	AIM
RAAN	rs4891825	G/A	AIM
LIMGH1	rs10007810	G/A	AIM
HTBP2	rs4918842	T/C	AIM
LOC442008	rs1876482	G/A	AIM
BRF1	rs3784230	A/G	AIM
DARC	rs2814778	T/C	AIM
AIM_rs2065982	rs2065982	T/C	AIM
AIM_rs6451722	rs6451722	G/A	AIM
AIM_rs730570	rs730570	G/A	AIM
AIM_rs2714758	rs2714758	A/G	AIM
MCM6_rs41525747	rs41525747	G/C	Lactase persistence
MCM6_rs4988235	rs4988235	G/A	Lactase persistence
MCM6_rs41380347	rs41380347	A/C	Lactase persistence
MCM6_rs145946881	rs145946881	C/G	Lactase persistence
PLRP2_rs4751995	rs4751995	G/A	Lipid metabolism
PLRP2_rs4751996	rs4751996	G/A	Lipid metabolism
SLC24A4	rs12896399	G/T	Eye color
TYR	rs1393350	G/A	Eye color
OCA2	rs1800407	C/T	Eye color
ASIP	rs6119471	C/G	Eye color
BNC2	rs2153271	C/T	Freckles
KITLG	rs12821256	T/C	Hair color
TYR	rs1042602	C/A	Hair color
EXOC2	rs4959270	C/A	Hair color
HERC2	rs12913832	A/G	Hair color (blond) - eye color (blue)
IRF4	rs12203592	C/T	Hair color (blond) - eye color (blue)
AR	rs2497938	T/C	Male baldness
AMELO	rs2106416	C (X) / T (Y)	Sex marker
AMEAB	rs9462529	5bp deletion	Sex marker
SLC24A5	rs1426654	A/G	Skin pigmentation
SLC45A2	rs16891982	C/G	Skin pigmentation - and eye color
MC1R_rs885479	rs885479	G/A	Skin pigmentation - hair color (red)
MC1R_rs1110400	rs1110400	T/C	Skin pigmentation - hair color (red)
MC1R_rs11547464	rs11547464	G/A	Skin pigmentation - hair color (red)
MC1R_rs1805006	rs1805006	C/A	Skin pigmentation - hair color (red)
MC1R_rs1805007	rs1805007	C/T	Skin pigmentation - hair color (red)
MC1R_rs1805008	rs1805008	C/T	Skin pigmentation - hair color (red)
MC1R_rs1805009	rs1805009	G/C	Skin pigmentation - hair color (red)
MC1R_rs2228479	rs2228479	G/A	Skin pigmentation - hair color (red)
ASIP_rs2378249	*rs2378249*	*A/G*	*AIM*
DHCR7/NADSYN1	*rs7944926*	*G/A*	*Circulating vitamin D level*
SLC22A4_rs1050152	*rs1050152*	*C/T*	*Coeliac disease*
SLC22A4_rs174546	*rs272872*	*G/A*	*Coeliac disease*
FADS1-FADS2	*rs174546*	*T/C*	*Decreased triglyceride levels*
EDAR	*rs3827760*	*A/G*	*Hair morphology*
GRM5	*rs7119749*	*A/G*	*Skin pigmentation*

(2) Immunity-related markers

Increasing population densities and the intimate contact with domesticated animals during the Neolithic are believed to have considerably increased zoonotic infections (that is, the transmission of infectious diseases from animals to humans) (Wolfe *et al.*, 2007). Thus, pathogens are expected to have shaped the human genome as a result of selective pressures acting upon genes involved in host resistance and immunity response. I selected a subset of such genes, harboring polymorphisms that influence susceptibility to parasitic and viral infections in European populations (Beutler, 1994; McKenzie *et al.*, 1998; Peisong *et al.*, 2004) (**Table 5**). Among these, some are involved in the innate immunity response, with functions spanning from pathogen recognition all the way to downstream effectors (Olalde *et al.*, 2014). Finally, to unravel to which extent modern diseases can be related to past adaptations, several common variants that have been associated with present-day diseases through genome-wide association studies were also added to the study, such as coronary heart disease, atherosclerosis and celiac disease (Hunt *et al.*, 2008; Yasuda *et al.*, 2007; Zhou *et al.*, 2012).

Table 5: Immunity-related variants incorporated in the current study

Gene	ID	Alleles	Associated phenotype
CDKN2B-AS1	rs2383206	A/G	Coronary heart disease
EDNRB-AS1	rs5351	T/C	Atherosclerosis
HLA-DQ1	rs2187668	C/T	Coeliac disease
HFE	rs1800562	G/A	Hemochromatosis
CCR5	CCR5-Δ32/ rs333	GTCAGTATCAATTCTGGAAGAATTT CCAGACA/ -	HIV-1 resistance
G6PD	rs5030868	G/A	Thalassemia risk factor - Malaria resistance
STAT6	rs324013	T/C	Resistance to Schistosome/Ascaris infection - Asthma
STAT6	rs324015	T/C	Resistance to Schistosome/Ascaris infection - Asthma
IL13	rs1800925	T/C	Resistance to Schistosome/Ascaris infection - Asthma
IFIH1/MDA5	rs10930046	C/T	Antiviral response
CASP12	rs497116	G/A	Susceptibility to sepsis
SOCS2	rs10745657	A/G	Cytokines and cytokine modulators
TOLLIP	rs5743899	C/T	Cytokines and cytokine modulators
IL29	rs30461	G/A	Cytokines and cytokine modulators
TGFB2	rs900	A/T	Cytokines and cytokine modulators
CCL18	rs2015086	G/A	Cytokines and cytokine modulators
CCL18	rs14304	C/T	Cytokines and cytokine modulators
NOS2A	rs7215373	C/T	Cytokines and cytokine modulators
SH2B3	rs3184504	C/T	Intracellular molecular adaptors
TLR1	rs4833095	C/T	Pattern recognition receptors
TLR3	rs3775291	C/T	Pattern recognition receptors
TLR6	rs5743810	G/A	Pattern recognition receptors
TLR1-6-10	rs4833103	C/A	Pattern recognition receptors
PPT2-EGFL8	rs2269424	G/A	Major histocompatibility complex

(3) Y chromosome markers.

The Y SNPs selected from the literature and their corresponding haplogroup are listed in the table below. I selected diagnostic SNPs for the assignment of various haplogroups reported in prehistoric Europe (Haak *et al.*, 2008, 2010; Lacan *et al.*, 2011). Other selected SNPs were used to define nodes and draw a decisional tree.

Table 6: Y chromosome SNPs and associated haplogroups

Marker	ID	Alleles	Haplogroup	Marker	ID	Alleles	Haplogroup
M42	rs2032630	A>T	BT	M522	rs9786714	G>A	IJK
M216	rs2032666	C>T	C	M304	rs13447352	A>C	J
V86	rs192993367	G>A	C*	*M267*	*rs9341313*	*T>G*	*J1*
V20	rs182352067	G>A	C1a	L559	rs373707621	A>G	J2a
M38	rs369611932	T>G	C1b2	*M102*	*rs2032608*	*G>C*	*J2b*
M217	rs2032668	A>C	C3	*M221*	*rs2032667*	*G>A*	*J2b*
M168	rs2032595	C>T	CT	P131	rs9786043	C>T	K
M96	rs9306841	C>G	E	P326	rs372687543	T>C	LT
M33	rs368762706	A>C	E1a	M231	rs9341278	G>A	N
P2	rs9785756	G>A	E1b1	P186	rs16981290	*C>A*	O
M215	*rs2032654*	*A>G*	*E1b1b*	M119	rs72613040	T>G	O1
L336	*rs112779735*	*G>A*	*E1b1b*	M122	rs78149062	A>G	O3
M78	*rs368977028*	*C>T*	*E1b1b1a1*	M45	rs2032631	G>A	P (QR)
M75	rs2032639	G>A	E2	M242	rs8179021	C>T	Q
P14	rs9786420	C>T	F	M207	rs2032658	A>G	R
M201	rs2032636	G>T	G	M173	rs2032624	A>C	R1
U5	rs2178500	G>T	G2a	*L146*	*rs17250535*	*T>A*	*R1a*
L31	rs35617575	C>A	G2a	*M449*	*rs17222279*	*G>A*	*R1a*
M52	rs376769460	A>C	H	*M459*	*rs17316227*	*A>G*	*R1a1*
m258	rs9341301	T>C	I	*L562*	*no SNP*	*G>A*	*R1b-S21*
L68	rs35547782	C>T	I2	*M415*	*rs9786194*	*C>A*	*R1b1*
M438	*rs17307294*	*A>G*	*I2*	*P202*	*rs369616152*	*T>A*	*S1*
M436	*rs17315680*	*G>C*	*I2a2*	*M184*	*rs20320*	*G>A*	*T*
P219	*rs17221964*	*T>G*	*I2a2a*				

3. Primer design

All the different primer pairs were designed using the Oligo 7 software in order to obtain pairs with a similar annealing temperature of 60°C, and to minimize potential dimer formation. To reduce the propensity of formation of secondary structures of the RNA baits synthesized from the PCR products, I restricted the product size between 70 to 160bp. The selectivity of each primer was verified in silico with a BLAST analysis (Altschul *et al.*, 1990). To ensure that capture baits specifically hybridize to their target, the selected region was also scanned with Repeatmasker to discard portions of DNA harboring repeats within the bait's sequence (Smit *et al.*, 2015).

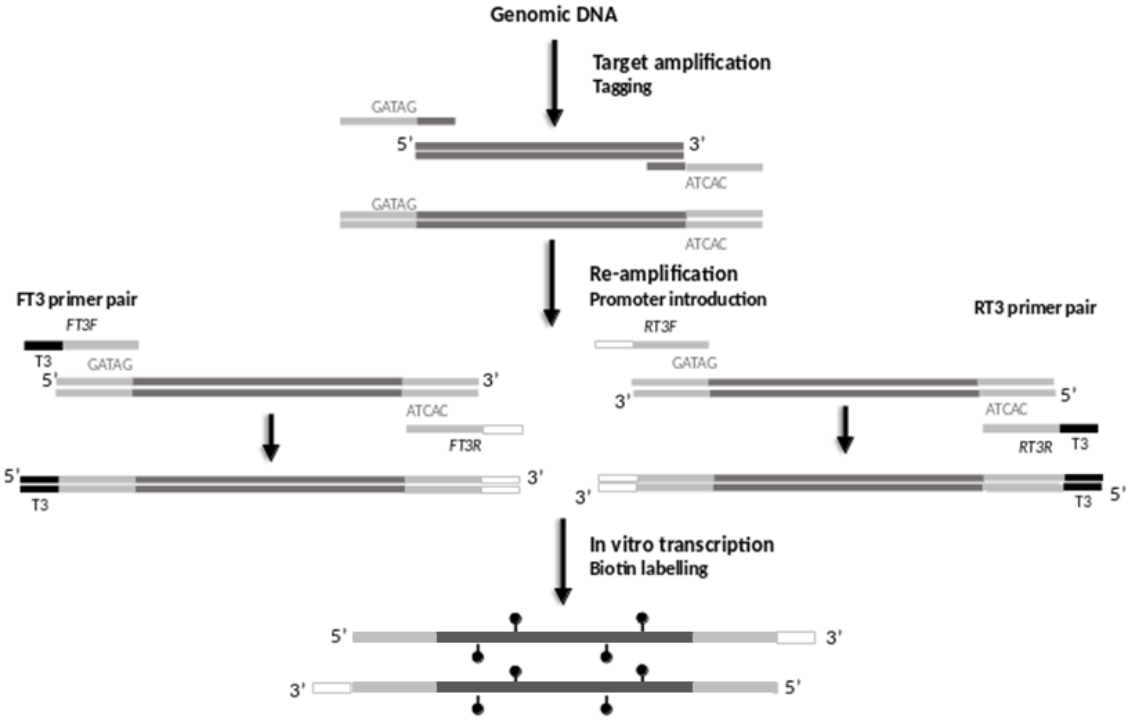


Figure 22: Strategy used to obtain biotinylated RNA baits from modern human DNA.

Forward and reverse locus-specific primers (dark grey) were modified by including short non-complementary sequences at their 5' end (light grey) that will act as primer-binding site for a pair of universal re-amplification primers, carrying either the T3 promoter (black) or a custom sequence (white) to amplify the first-PCR product (Clarke et al., 2014). In vitro transcription of these PCR products in the presence of biotin-CTP results in randomly labeled RNA capture probes.

III. Target PCR amplification

The amplification of each marker was performed using qPCR on a LightCycler480 (*Roche*), in a 384-well plate, and two quantities of template genomic DNA were assayed: 4ng and 800pg. For each pair of tagged primers, two wells were dedicated to the sample amplification, and two other wells were negative controls where genomic DNA was replaced by nuclease free water. A specific mix containing a dilution of this oligonucleotide was prepared and added manually. The number of copies per microliter was calculated according to the following formula: *(Quantity (g) * Avogadro constant (mol^{-1})/ length (bases)* molecular weight of a nucleotide ($g.mol^{-1}$)) / volume (µL)*.

Most primers were behaving as expected, except for a few that required some troubleshooting as described here. The Y chromosome marker P131 harbored two distinct melting peaks when amplified from 800 pg genomic DNA. One of those corresponded to the melting temperature of the negative control (**Figure 23a**.), thus indicating that primer dimers were formed along with the expected product. It was therefore decided to use 4ng of modern human DNA as a template for each PCR reaction. Another Y chromosome marker, M231, harbored a bimodal melting peak that could indicate the presence of several products (**Figure 23c.**), but there was only one product to be seen in the gel electrophoresis of the PCR product and of the re-amplification from the pool assay (as described in the next part) (**Figure 23d.**). This bimodal tendency might be specific from this product and reflect a bimodal distribution of the AT vs. GC % along the sequence.

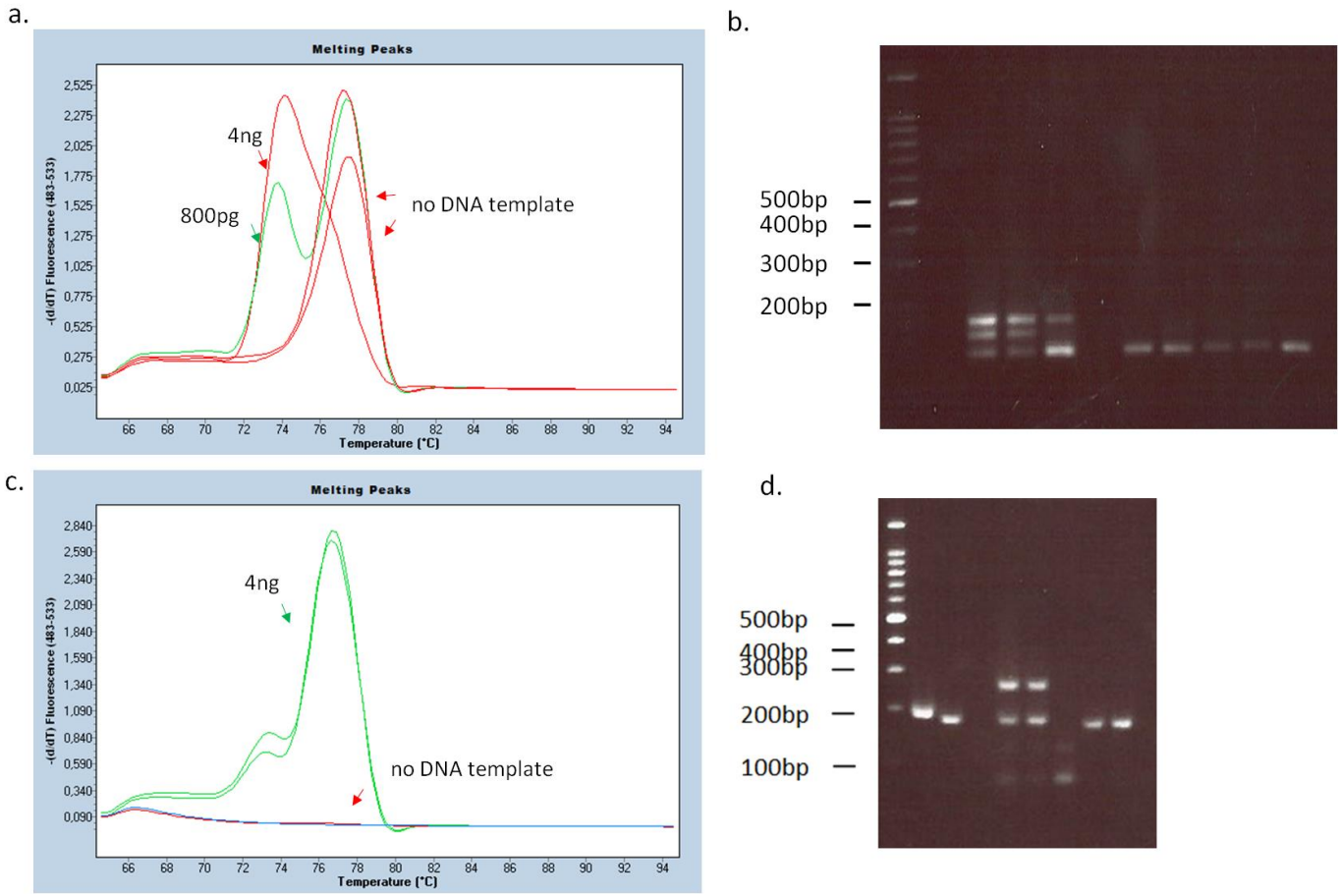


Figure 23: qPCR amplification of P131 and M231.
(a) P131 melting peaks obtained from the amplification of respectively 4ng (red), 800pg (green) modern human DNA. The two remaining red peaks correspond to negative controls. (b) Agarose gel electrophoresis of the qPCR products of CCR5, obtained through the amplification of 20ng (lane 3), 4ng (lane 4) and 800 pg (lane 4) of modern human genomic DNA. Lane 6 – 10 corresponds to negative controls. (c) M231 melting peaks obtained from the amplification of 4ng modern human DNA (green). The two remaining curves correspond to negative controls. (d) Agarose gel electrophoresis of the qPCR products of P131 (lane 2) and M231 (lane 3), obtained through the amplification of 4ng of modern human genomic DNA (2-3), the amplification of P131 (lane 3-5) and M231 (10-13) from a pool of PCR products that were re-amplified with universal primers carrying a T3 promoter. Lane 6 and 9 correspond to negative controls.

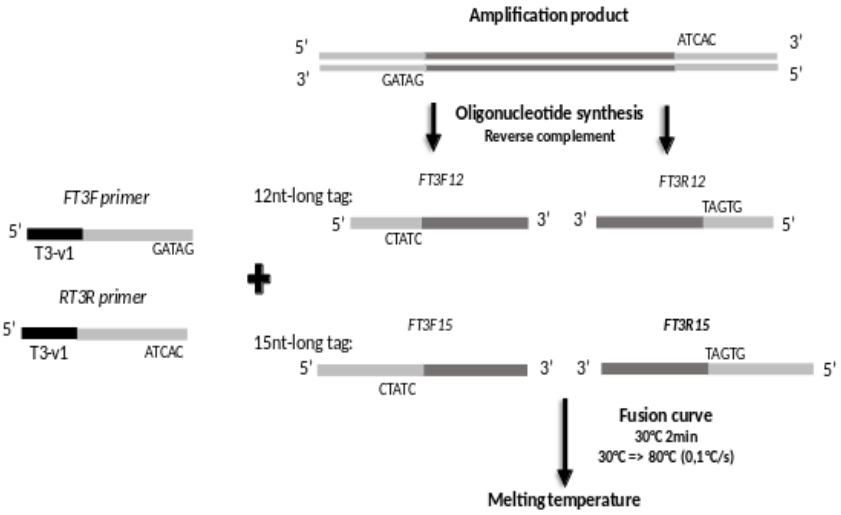
Since no amplification was observed for CCR5, another qPCR was performed in capillaries with a newly made primer dilution. Upon amplification, it showed multiple products on agarose gel (**Figure 23b.**) despite a single well-defined melting peak, indicating that this primer pair allows the amplification of at least three distinct products. Since the allele to be screened for this marker is a 32bp deletion, two distinct products on the gel could correspond to heterozygosity for this marker; however, none of the lowest bands correspond to the size of the expected product harboring the deletion. The three slices were cut out of the gel and purified following a QIAgen gel extraction kit and sent

for Sanger sequencing, and as expected none correspond to the deletion, and each sequence only partially aligned to the reference CCR5 allele. Thus, the primers are able to amplify different genomic regions. Since the allele we are interested in is mainly the deletion, I designed an oligonucleotide corresponding to the deleted region along with new primers to amplify it (see **Supplementary Table 1**).

IV. Second amplification

One microliter of each previously amplified products was transferred to a new Eppendorf tube to form a pool, and the volume was brought up to 100 µL to obtain a 1/100 dilution of each individual product, further diluted to 1/100 000. The re-amplification step brings the T3 promoter upstream of either the forward or the reverse strand. In each case, the other primer of the pair brings a custom sequence, carefully designed not to be complementary to the sequence of any of the Illumina adapters used to prepare the ancient DNA libraries and to be uncommon, if not absent, throughout the human genome. Three different sets of candidate primers were tested for the re-amplification step, varying in either the position (forward: FT3 or reverse: RT3), the length (26 or 33 nt) of the T3 promoter or the length of the 5' tag (see **Figure 24**).

My first aim was to test the specificity of the hybridization of primers to their tag. This was achieved by determining their respective melting temperature for different lengths of the annealing portion (*i.e.* the length of the tag). I mixed one of the primers (FT3) with different short complementary oligonucleotides either corresponding to the proper tag or the opposite one. Both these oligonucleotides were synthetized in two different lengths. Double-stranded oligonucleotides production was measured using SYBR Green I-containing PCR reaction buffer in a LightCycler. Temperature was then raised up to 80°C while monitoring the fluorescence. The procedure is summarized in **Figure 24** (see also **Figure 22**).



	reverse complement of the forward tag		reverse complement of the reverse tag	
	annealing over 12 nt FT3F12	annealing over 15nt FT3F15	annealing over 12 nt FT3R12	annealing over 15nt FT3R15
FT3 forward primer GCAAATTAACCCTCACTAAAGGG AGA GATAG	CTATCTCTCCCT	CTATCTCTCCCTTTA	GTGATTCTCCCT	GTGATTCTCCCTTTA
RT3 reverse primer GCAAATTAACCCTCACTAAAGGG AGA ATCAC				

Figure 24: Conducted tests to estimate the specificity and the melting temperature of universal re-amplification primers depending on the annealing length.

Using 1 µM of each oligonucleotide in a 10 µL reaction volume, the obtained melting temperature was 45.7°C when annealing over 12 nucleotides, and 51.6°C when annealing over 15 nucleotides. No proper melting peak was computable when FT3 was incubated with the reverse complement of the opposite tag, indicating that the specificity provided by a 5-nt mismatch was sufficient. In order not to lower the annealing temperature of the re-amplification step below 45°C (to avoid non-specific hybridizations that could occur at low temperature), I therefore decided to add a 15 nt–long tag to the target specific primers. To then test the re-amplification efficiency, the purified product from the amplification of one nuclear marker alone (LP1) was used as a template for 3 qPCR reactions, each corresponding to a universal re-amplification primer pair. In parallel, the same amount of the product was amplified with the primers that were used during the first PCR. There was no significant variation in crossing point (Cp) between the different pairs, and no dimers were found along with the products. (**Figure 25**) As the longest version of the promoter was expected to increase the efficiency of the transcription, I used it for both strands.

Figure 25: Re-amplification of LP1 tagged PCR product.
Amplification products from the target specific tagged primers are shown lanes 4 and 5. Lanes 7-12 correspond to the re-amplification with respectively: the FT3-v1 pair (7, 8), the FT3-v2 pair (9, 10) and the RT3-v1 pair (11, 12).

To assess the representation of each marker when pooled together, I conducted a test consisting in using the re-amplified pool as a template for each of the 88 initial primer pairs. First, I tried two dilutions of each individual PCR product in the pool, 1/1000 and 1/100 000 respectively, to see if manipulating certain products concentration could allow me to counter the small Cp variations between them during the first PCR. Upon re-amplification of these pools, I also tested two dilutions of the re-amplified products. Average Cp for a 1/100 000 dilution of a 1/100 000 diluted pool was 23.95 (SD +/- 1.16) for 83/88 markers, and 29.64 (SD +/- 1.88) for the remaining 5.

To compensate for this lower efficiency of amplification, these outliers were less diluted in the final pool, resulting in a dilution scheme when multiplexing the target amplification products as follows: 1 of each reaction with Cp around 25 was transferred to a new Eppendorf, and volume was brought to 100 µL by adding nuclease free water. Then 1 µl of this pool was transferred to a new tube, along with 1 µL of each PCR products with a Cp around 29 and volume was brought to 100 by adding nuclease free water. In the end, this pool was again diluted to the tenth. This results in either a 1:1000 dilution or a 1:100 000 dilution depending on the product.

V. In vitro transcription

The main difference in our approach compared to that of Maricic and colleagues (Maricic *et al.*, 2010) is that we use RNA probes instead of DNA. RNA presents a number of advantages over DNA when it comes to generating probes. First, it is easily degraded after capture, which allows for the elimination of all probes from the reaction before amplifying the enriched libraries, where DNA would persist and potentially generate chimeric molecules during PCR, thus creating artefacts. We can also generate single-stranded RNA probes so that hybridization among probes won't compete with the hybridization of RNA to its DNA complementary target. Moreover, in vitro transcription produces high amount of RNA relative to the amount of input DNA template, thus representing an extra step of amplification of the number of probes: from a single PCR reaction, we can generate several micrograms of RNA that can be stored and used for several downstream captures.

For mitochondrial probes, 5µL of either forward (1-11F) and reverse (1-11R) PCR reactions were used to create two separate pools (i.e. one for each strand), which were subsequently purified and concentrated on Qiagen columns. For nuclear probes, the re amplified pool (see **Selection of nuclear markers**) was purified using Qiagen columns**.** In vitro transcription is performed using the T7 polymerase, with a ratio of 1:1 CTP and biotin-CTP, and either 100ng or 500ng of purified PCR productions as a template, for nuclear and mitochondrial probes respectively.

After purification (see **Material and methods**), mitochondrial RNAs are fragmented by a 4 minutes incubation at 94°C in a magnesium-rich buffer, resulting in a size distribution of 200 to 500 bases for our overlapping mitochondrial RNA probes.

Chapter II. Results of a preliminary capture

To investigate the ability for each bait to hybridize with its target, the capture was conducted on a human DNA library prepared from a 150-year old sample from a grave that yields 8% of endogenous DNA. This library was prepared following Illumina's recommendations and amplified for 6 cycles, for a final DNA concentration of 2 ng/µL. A 5 µL aliquot was amplified for 5 additional cycles and eluted in 60 µL. For the first round, I captured both strands separately, and mitochondrial and nuclear baits were separated as well, for a total of 4 capture reactions for the sample. At the end of the first round of capture, each of the 4 libraries was quantified by qPCR, then amplified for 20 or 21 cycles, the latter corresponding to the mitochondrial forward strands capture. The four PCR reactions were then pooled and upon purification, two capture reactions were set up with forward baits and reverse baits respectively.

Around 1.24 million paired-end reads passed the Illumina quality filter and were successfully merged with LeeHom. After removal of PCR duplicates that accounted for 72% of the library, 148,972 reads remained that could be aligned to the human reference genome hg19, among which 5% mapped to the mitochondrial genome. The resulting coverage was of 36,9X on average ((SD +/- 10.8), allowing to unambiguously assign the T1a1k haplogroup.

Only 159 reads could be mapped to a concatenation of the different target regions, resulting in coverage of 35.45 % (SD +/- 32) on average. Y chromosome markers were excluded from the calculations since no reads mapped to these regions, consistent with the individual being a woman.

The low coverage of nuclear markers compared to that of the mitochondrial genome could have several explanations. In human cells, for each copy of the nuclear genome, there are several hundreds to thousands of copies of the mitochondrial genome, which makes it more likely to be recovered from ancient remains (Bogard *et al.*, 2005). This ratio is variable among sample types, and varies with their endogenous DNA yield (Furtwängler *et al.*, 2018). The inherent complexity of the captured library might have been too low, with too little different molecules covering regions of interest. Second, hybridization conditions might not be as optimal for the recovery of single copy markers as they are to capture the mitochondrial genome. Hybridization time, in particular, might be too short to allow fully efficient capture of the rarer molecules. Since the amount of

endogenous DNA per se could not be directly measured, it had to be estimated based on the quantity of DNA obtained after amplification of the initial library. In this case, the library quantity was 50 ng after 6 cycles of amplification, indicating that the library initially contained around $50/2^6$= 10.8 ng of amplifiable DNA, corresponding to 10.8*8.9%= 69pg of endogenous DNA. Estimating one complete haploid nuclear genome at 3.3pg, the initial library thus contained around 20 haploid copies of the genome, which should be enough to ensure the representativeness of each individual locus. The present estimation might thus be an overestimate of the useful copy number, because not all read length might be captured with the same efficiency. To estimate whether there was a bias in favor of the longer reads during capture, the distribution of the reads mapping to human genome (including the mitogenome) was compared before and after capture using boxplots (**Figure 26**)**.** It can be seen that capture enriches for longer fragments (there is about a 10-15bp increase in both the median size and the upper and lower quartile), thus selecting a subpopulation of the longer fragments. This implies that not all of the initial estimated 20 haploid copies are captured.

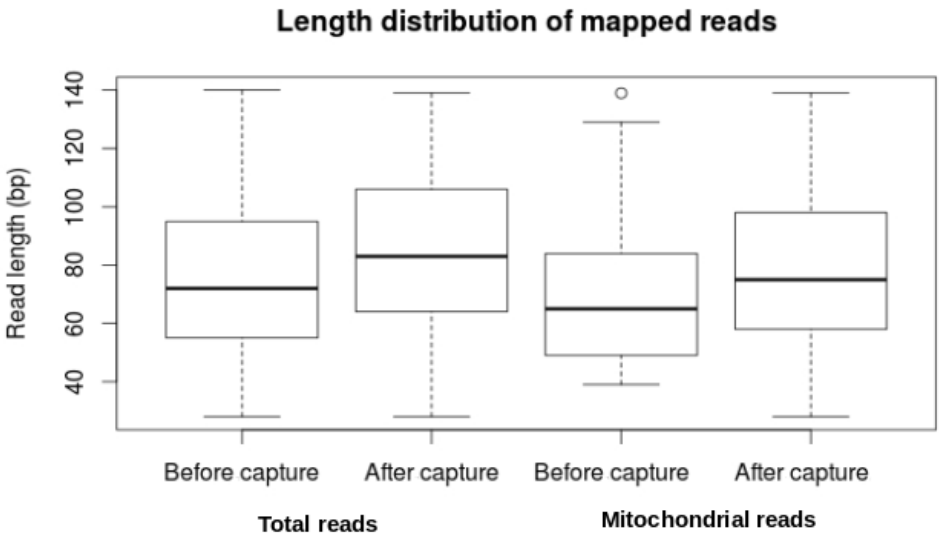


Figure 26: Boxplots showing the size distribution of reads mapping to human genome and mitogenome before and after capture.

As previously discussed, another possible cause for the low recovery of nuclear markers could be that they are less efficiently captured than the mitogenome over a similar hybridization time. To estimate if this was the case, I used the following reasoning: The combined length of all the captured regions was 10 kb long, i.e., representing a $1/3.10^5$ fraction of the complete genome. Before capture, around 163,000 reads were successfully mapped to the human genome while 77 specifically aligned to the mitogenome. After capture, the latter fraction increased to 7830, which is 100 time higher than prior to capture. If the capture was as efficient for the nuclear targets as it was for the mitogenome, I would have expected $(163{,}000/3.10^5)$ *100 = 54 reads on target. As stated above, 159 independent reads were mapped to the target region, indicating that the nuclear genome was captured as efficiently as the mitochondrial genome, if not better. Thus, the most plausible interpretation is that the complexity of the sufficiently long fragments in the initial library is a limiting factor, which implies that a particular attention has to be given to library preparation. In addition, the quality of the prepared libraries must be carefully evaluated in order to select for samples that give rise to libraries of sufficient initial complexity to ensure that a satisfactory coverage of the desired targets will be achieved following capture.

Chapter III. Conclusion and discussion

Trace amounts of DNA can “survive” the decomposition of organic matter for long periods of time after the death of an organism, and the recent field of Paleogenomics made the retrieval of these ancient molecules possible, making them accessible to genetic analysis. But the minute amount and degradation state in which it subsists make it a methodological challenge to work with.

The aim of this part of the project was to elaborate a technique to capture genomic regions of interest, selected based on their relevance to the question raised above, in order to enrich a sample in those regions prior to high-throughput sequencing. Therefore, I had to develop a protocol to capture nuclear regions of interest with biotinylated RNA baits, while adapting the capture of the mitochondrial genome to human samples.

Biotinylated RNA baits were transcribed from 70-160bp-long genomic target regions that were amplified by PCR beforehand, which provides the economy of multiple oligonucleotide synthesis foreach bait. The whole procedure was performed with a two-step PCR, using a pair of primer per genomic target, and a universal pair of primers to incorporate the T3 promoter upstream of either the forward or the reverse strand, thus allowing the synthesis of baits complementary to either DNA strands. This strategy reduces by half the costs of the oligonucleotides during the capture set up and minimizes the risk of the formation of primer dimer artefacts that is increased with long 5’tags. Altogether, this strategy could be a valuable economic alternative to sequence capture using commercial microarray-derived oligonucleotides for a medium size number of markers captured.

The attempt made at capturing nuclear SNPs alongside the mitochondrial genome revealed that the baits were efficient at recovering their target region, but the initial complexity of the ancient DNA library is a limiting factor.

Part IV. A French population from the Neolithic to the Iron Age: insight from uni-parentally inherited markers and phenotypes

Chapter I. Introduction

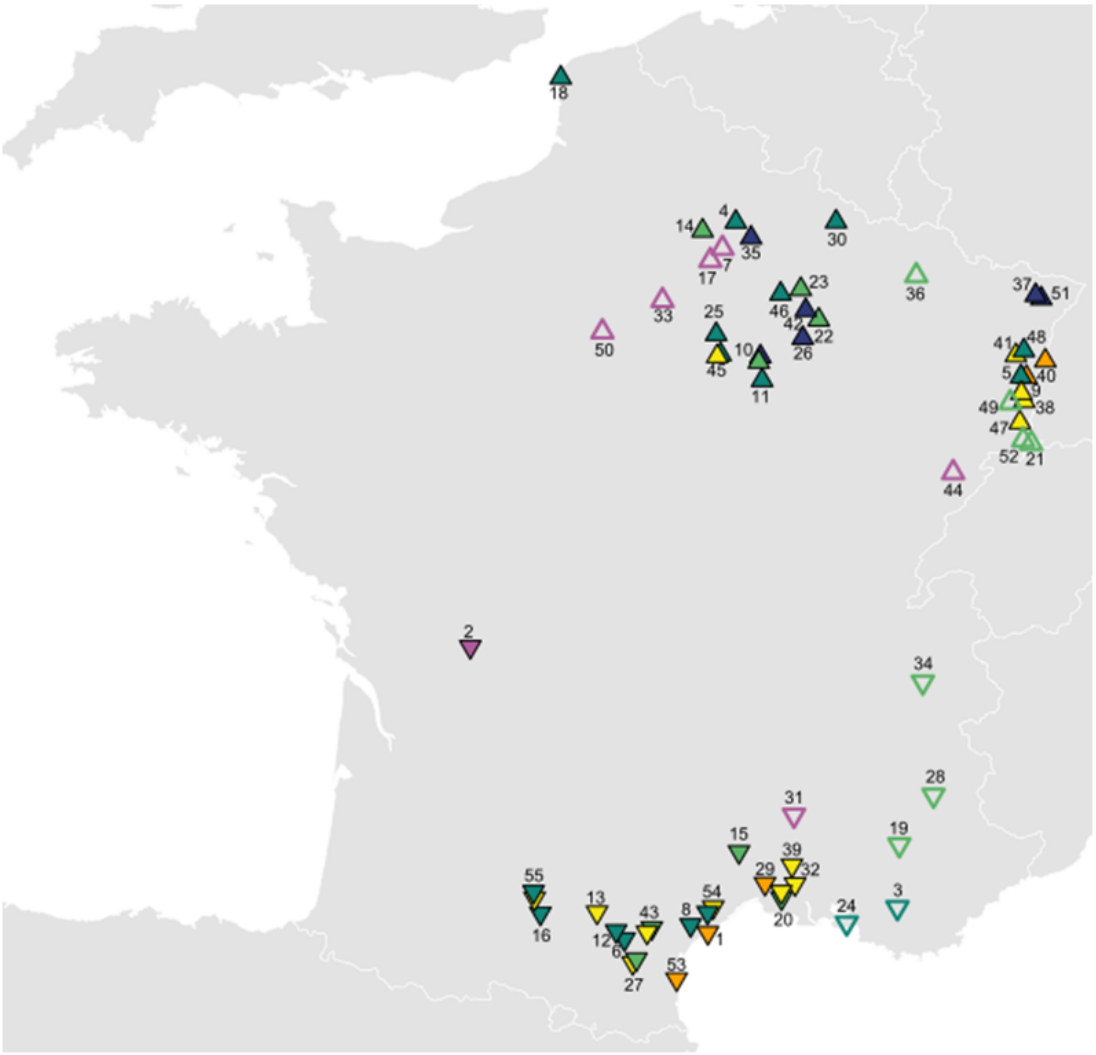
We sampled individuals in three key regions of present-day France: the Hauts-de-France and Grand-Est, spanning the Northern part of France from the Channel to the Franco-German border; and Occitanie, located between the Rhône River and the Pyreneans along the Mediterranean Sea (**Figure 27**). Since samples were collected from different geographical regions and periods, we started off with very little idea of the overall preservation in our dataset. As our enrichment method for both the mitochondrial genome and targeted loci across the nuclear genome had proven successful on a test sample **(Part III, Chapter II**), we decided to first undertake the study of ancient French populations through this approach to maximize the number of processed individuals, thus gaining statistical power.

This choice was also motivated by the fact that mitochondrial DNA is the best described marker in population genetics. Despite the increasing number of nuclear genomes available, it remains a solid choice to study large panels and compare groups of individuals. Mitochondrial haplogroups as direct evidence of an individual's ancestry have been rightfully criticized (Gerbault *et al.*, 2014), but used in the context of populations and not individuals, female lineages tell their own story and were shown to often summarize the history of the whole population (Brandt *et al.*, 2013; Szécsényi-Nagy *et al.*, 2017).

To integrate ancient French populations within the broader context of Western Eurasia, we gathered data from the literature that cover the region during the Mesolithic period and subsequent Neolithic and Bronze Age. Although mitochondrial was the preferred marker in the study of ancient individuals over the last 30 years, it was mostly studied through PCR-based approaches. Because of the afore-mentioned issues associated with this approach, we restricted our analyses to complete mitochondrial genomes, generated through targeted enrichment and shotgun sequencing.

Here we present the complete mitochondrial genomes, Y chromosome markers and genotypes on a number of nuclear loci of interest obtained through a DNA enrichment approach of 193 Mesolithic, Neolithic, Bronze Age and Iron Age individuals

(**Supplementary Table 4**). This study provides, for the first time, a high-resolution 5,000-year transect of the dynamics of maternal and paternal lineages in France as well as of autosomal genotypes associated with known phenotypes. This transect, that comprises three major cultural transitions of the late Prehistory of France (i.e. the advent of the Neolithic, the Bronze Age and the Iron Age), fills a large gap in the understanding of these events and their complexity, both locally and at the scale of Europe.



1	Agde (Hérault)	21	Hégenheim (Haut-Rhin)	41	Obernai (Bas-Rhin)
2	Agris (Charente)	22	Isle-sur-Marne (Marne)	42	Orcontes (Marne)
3	Saint Maximin-la-Sainte-Baume (Var)	23	La Cheppe (Marne)	43	Pépieux (Aude)
4	Beaurieux (Aisne)	24	La Couronne-Martigues (Bouches-du-Rhône)	44	Pont-de-Roide (Doubs)
5	Bergheim (Haut-Rhin)	25	La-Villeneuve-au-Chatelot (Aube)	45	Pont-sur-Seine (Aube)
6	Berriac (Aude)	26	Larzicourt (Marne)	46	Recy (Marne)
7	Berry-Au-Bac (Aisne)	27	Laure (Aude)	47	Rixheim (Haut-Rhin)
8	Béziers (Hérault)	28	Lauzet-Ubaye (Alpes-de-Haute-Provence)	48	Rosheim (Bas-Rhin)
9	Bischwihr (Haut-Rhin)	29	Le Cailar (Gard)	49	Rouffach (Haut-Rhin)
10	Bréviandes (Aube)	30	Mairy (Ardennes)	50	Rueil-Malmaison (Hauts-de-Seine)
11	Buchères (Aube)	31	Malataverne (Drôme)	51	Schwindratzheim (Bas-Rhin)
12	Carcassonne (Aude)	32	Manduel (Gard)	52	Sierentz (haut-Rhin)
13	Castelnaudary (Aude)	33	Mareuil-lès-Meaux (Seine-et-Marne)	53	Sigean (Aude)
14	Ciry Salsogne (Aisne)	34	Marlens (Haute-Savoie)	54	Valros (Hérault)
15	Corconnes (Gard)	35	Menneville (Aisne)	55	Villeneuve-Tolosane (Haute-Garonne)
16	Cugnaux (Haute-Garonne)	36	Mondelange (Moselle)		
17	Cuiry-Lès-Chaudardes (Aisne)	37	Morschwiller-le-Bas (Bas-Rhin)		
18	Escalles (Pas-de-Calais)	38	Niederergheim (Haut-Rhin)		
19	Forcalquier (Alpes-de-Hautes-Provence)	39	Nîmes (Gard)		
20	Garons (Gard)	40	Nordhouse (Bas-Rhin)		

Figure 27: Location of French archeological sites included in this study.
Symbols reflect their geographic area of origin: upper triangles: Northern France, lower triangles: Southern France. Filled symbols corresponds to newly reported sites. Color hues reflect the culture associated with each location: Ligth purple: Mesolithic hunter-gatherer, dark blue: Early Neolithic, dark green: Middle Neolithic, light green: Late Neolithic, yellow: Bronze Age, orange: Iron Age.

Chapter II. Results

I. Sample characterization

A total of 492 libraries covering 342 unique individuals were processed during this project, for the period spanning the Mesolithic to the Iron Age, among which 229 were screened through shotgun sequencing. Results for all studied samples can be found in **Supplementary Table 3**. Following read alignment and removal of PCR duplicates, we first determined to proportion of reads that aligned to the reference human genome in order to estimate the fraction of endogenous DNA within our samples (**Figure 28**). The fraction of endogenous DNA was high, with 57% of the samples that were shotgun sequenced being above 20% endogenous DNA. We found that the endogenous DNA content is dependent on the type of ancient sample, with the best preservation being found in temporal bones (38% +/- 20%), followed by teeth (3.8% +/- 6.2%). Despite their apparent density, long bones such as femurs and ulna yielded very low amounts of endogenous DNA, with an average content around 0.5% (+/- 3.2%).

Based on this initial screening, 246 libraries were captured for both the complete mitochondrial genome and a panel of 120 nuclear single nucleotide polymorphisms (SNPs) of interest including the Y chromosome (**See Supplementary Table 1** for the list of genotyped positions and **Supplementary Table 4** for the results). Among these, individuals who did not reach sufficient coverage or resolution (with more than 5% undetermined bases on the mitochondrial genome), or were assumed to be related to another individual in the dataset, sharing the same funerary structure/tomb and maternal lineage, were removed.

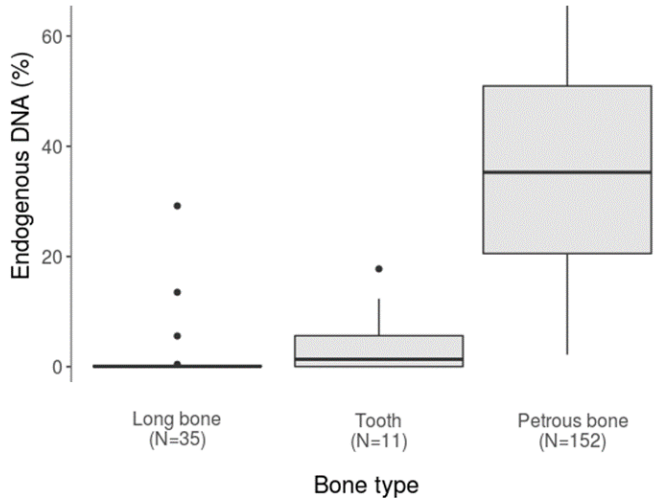


Figure 28: Boxplot showing the percentage of endogenous DNA per bone type across all sequenced samples

We then focused on authenticating whether our sequencing data were likely to be that of ancient samples. We based our estimations on two criteria: the size of DNA fragments, and the typical nucleotide misincorporation patterns that result from inflated cytosine deamination rates at overhangs, indicating severe post-mortem damage. Such misincorporation patterns are instrumental to authenticate ancient sequences versus modern contaminants (Briggs *et al.*, 2007). Since no USER treatment was used for the libraries destined to capture, we expected them to display moderate frequencies of C to T/G to A towards read termini. The distribution of damage patterns across DNA molecules for all captured individuals with more than 3000 mapped unique reads is shown **Figure 29**.

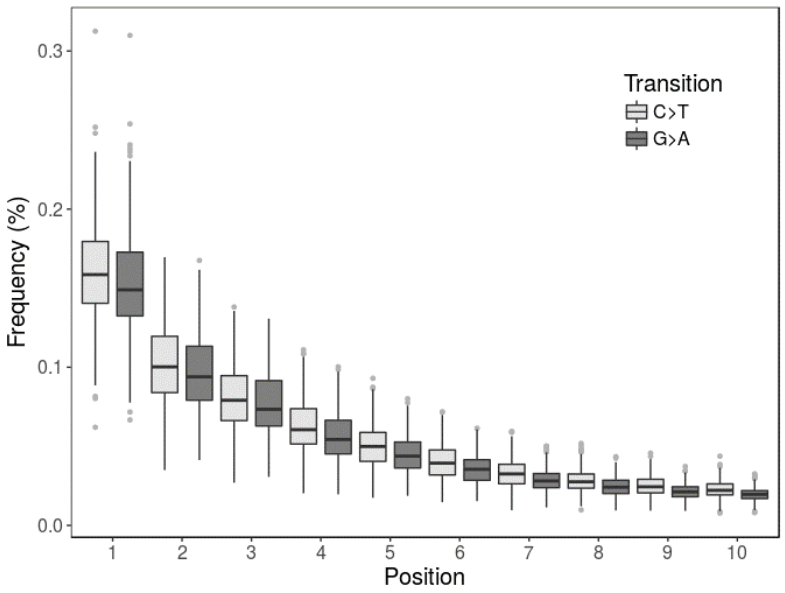


Figure 29: Frequency of transitions associated with aDNA damages on the 10 terminal positions of non-UNG treated DNA molecules

Here, we report the sequences of the mitochondrial genomes of 193 individuals following strict authenticity criteria, as well as the Y chromosome haplogroups of 38 individuals. Moreover, 149 individuals provided nuclear genotypes for at least one of the 120 targeted positions. After filtering, sequences were grouped together based on temporal and geographical proximity of the samples, often translating into cultural affinity. We combined our data with 638 previously reported complete mitochondrial genomes from contemporary individuals from across Europe and the Near East (**Figure 30**) (Gamba *et al.*, 2014; Haak *et al.*, 2015; Lazaridis *et al.*, 2014; Lipson *et al.*, 2017; Mathieson *et al.*, 2015; Olalde *et al.*, 2018; Posth *et al.*, 2016; Valdiosera *et al.*, 2018), and ran a set of descriptive statistical analyses on the present dataset at the haplogroup frequency level (using both PCA (**Figure 31b.**) and Ward Clustering (**Figure 32**)) and at the sequence level (F_{ST}, **Table 8**) to study the distribution of maternal lineages in various European populations through time. To follow the occurrence across time of each of our 20 core haplogroups and detect potential patterns, all French individuals were assembled into six groups to form a transect through time, ranging from the Mesolithic to the Bronze Age (**Figure 33**).

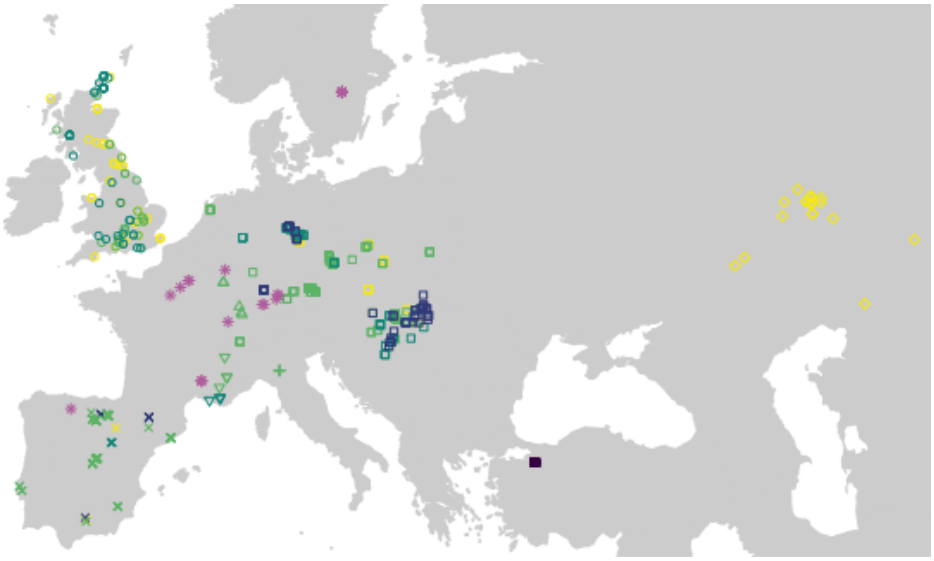


Figure 30: Location of published samples included in this study.
We gathered complete mitochondrial genomes from Western Eurasians from the Mesolithic to the Bronze Age. Symbols reflect their geographic area of origin: upper triangles: Northern France, lower triangles: Southern France, circles: British Isles, crosses: Iberian Peninsula, squares: Central Europe, diamonds: Pontic Steppe and Russia, asterisks: Western Europe Mesolithic hunter-gatherers. Similarly to Figure 30, color hues reflect the culture associated with each location: Ligth purple: Mesolithic hunter-gatherer, dark purple: Anatolian Neolithic, dark blue: Early Neolithic, dark green: Middle Neolithic, light green: Late Neolithic, yellow: Bronze Age.

The first two components of the PCA, describing 60% of the total variability, reveal that all included populations, to the exception of one outlier (Esc), fit within the edges of a triangle defined by Hunter-Gatherers (HG), Anatolian Neolithic farmers (ANE) and Bronze Age steppe pastoralists (SBA), with most Neolithic groups falling on a cline between ANE and SBA with little spread towards the HG edge. Since frequencies estimated on discrete core haplogroups do not account for shared lineages between groups, multidimensional scaling (MDS) was conducted on genetic distances computed from complete mitochondrial genomes (**Figure 31c**). We performed a Mantel-test to examine whether the increasing genetic distances, observed through pairwise F_{ST}, correlate with the temporal or the geographical distance between the ancient populations. The test revealed significant positive correlation (p-value: 0.02609739) between genetic and temporal distance, but no correlation with geographical distance (p-value: 0.8080192).

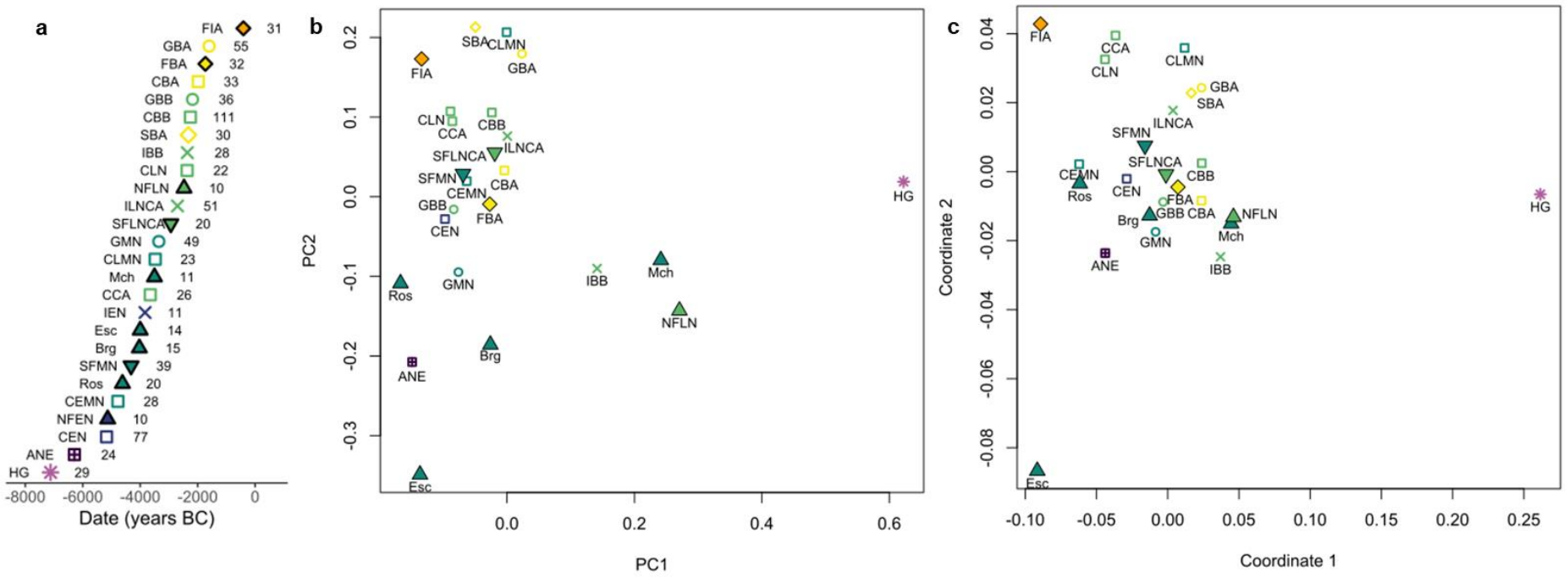


Figure 31: Clustering of the individuals included in the present study.

a. Dates (in years BCE) of the different populations and their respective number of individuals: HG: *Hunter-gatherers*, ANE: *Anatolia Neolithic*, CEN: *Central Europe Early Neolithic*, NFEN: *North France Early Neolithic*, CEMN: *Central Europe Early Middle Neolithic*, Ros: *Rosheim (Early Middle Neolithic,, Alsace)*, SFMN: *South France Middle Neolithic*, Brg*: Bergheim (Late Middle Neolithic, Alsace)*, Esc: *Escalles (Late Middle Neolithic, Pas-de-Calais)*, IEN*: Iberia Early Neolithic*, CCA: *Central Europe Chalcolithic*, Mch: *Michelsberg (Late Middle Neolithic, Northern France)*, CLMN: *Central Europe Late Middle Neolithic*, GMN: *Great Britain Middle Neolithic*, SFLNCA: *South France Late Neolithic/Chalcolithic*, ILNCA: *Iberia Late Neolithic/Chalcolithic*, NFLN: *North France Late Neolithic*, CLN: *Central Europe Late Neolithic* (Corded-Ware), IBB: *Iberia Bell Beaker*, SBA: *Steppe Bronze Age*, CBB: *Central Europe Bel Beaker*, GBB: *Great Britain Bell Beaker*, CBA: *Central Europe Bronze Age*, FBA: *France Bronze Age*, GBA: *Great Britain Bronze Age*, FIA: *France Iron Age*. More information about the different populations and their clustering can be found in **Table 2** and **Supplementary Table 6**. **b.** Principal component analysis conducted on mitochondrial haplogroup frequencies. The two first components represent 39.4 % and 26.15% respectively of the total variance. **c.** Multidimensional scaling performed on pairwise F_{ST} values**.** Values and associated p-values can be found in Table. Pairwise F_{ST} values and their associated p-values can be found in **Table 8.**

II. The Mesolithic substrate (~ 11,000-7000 BCE)

We report the complete mitochondrial sequences of six new French HGs. Their analysis together with 15 individuals previously published (Posth *et al.* 2016, Fu *et al.* 2016) shows that the French Mesolithic substrate between 11,000-7000 BCE, was characterized by mitochondrial haplogroups U5b (85%) and U5a (15%), comprising U5b1, U5b1b, U5b1h, U5b1a, U5b2a, U5b2b and U5a2, already observed in European Mesolithic hunter-gatherers (*e.g.* Lazaridis *et al*., 2014; Olalde *et al*., 2014; Posth *et al*., 2016). French Mesolithic hunter-gatherers were also characterized by a low nucleotide diversity (π 0.000923 +/- 0.000475), suggesting a rather homogeneous population across the territory.

III. Evolution of the populations during the Neolithic (~ 5500–2500 BCE)

The French Neolithic farmers (from ~5500 BCE onwards) display a wide diversity of haplogroups (U8, U3, N1a, K, J, T2, H, HV, X, V, W), with differences in frequencies depending on the period or even the archeological site. Reinforced by the Fisher's exact test p-values, showing significant differences between HGs and every other population (**Table 8**) this disparity confirms the Anatolian ancestry of early Neolithic farmers in France, as elsewhere in Europe. HG mitochondrial haplogroups (*e.g.* U5b), however, are carried by individuals from the Middle Neolithic onward, indicating that some autochthonous hunter-gatherer maternal lineages were incorporated into farming communities over the course of the Neolithic. (**Figure 33**).

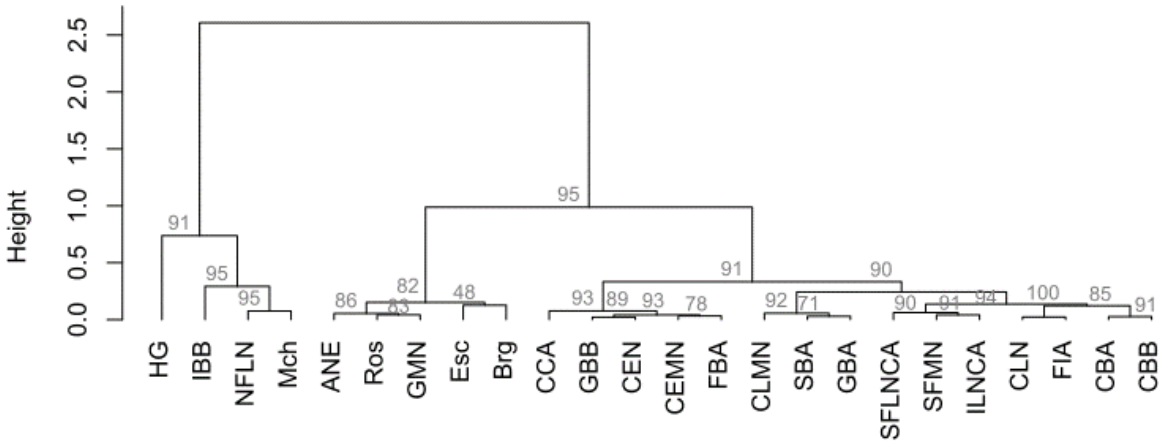


Figure 32: Ward's clustering of ancient populations based on haplogroup frequencies
Grey values on the edge of each cluster are approximately unbiased (au) p-values computed by 10,000 multiscale bootstrap resampling.

Data are very scarce for the earliest part of the Neolithic, corresponding respectively to the LBK culture that emerged from the Danubian LBK and became established in the Northern part of France, and to the Cardial culture in the Southeastern part of France at the Mediterranean coast and further north into the Rhone valley. In total, we obtained eight reliable mitochondrial genomes (i.e. Phy-Mer score above 0.95 (Navarro-Gomez *et al.* 2015)) from individuals from both Alsace and Champagne (Grand Est), and the Vallée de l'Aisne (Hauts-de-France) belonging to the LBK culture. We complemented these haplogroups with those obtained through the shotgun sequencing of two other individuals from Alsace (Mor6 and Schw432, **Table 7**). Interestingly, the currently rare haplogroup N1a that was previously described as a signature haplogroup among the LBK in Central Europe (Haak *et al.*, 2010), could not be detected in our samples. We estimated that, due to sample size, any haplogroup below 24% in the population would have 95% chance of being missed, which is more than the frequencies of 15% and 10%, respectively, observed for N1a in Early Neolithic individuals from Germany and Hungary.

Table 7: Mitochondrial haplogroups and location of individuals assigned to the LBK culture

Individual	Location	Department	Region	Dates	Haplogroup
BRE447	Bréviandes	Aube	Grand-Est (East)	5209-4912 cal BC	V
Mor5	Morschwiller-le-Bas	Bas-Rhin	Grand-Est (East)	5200-5000 BC	T2b23
Mor6	Morschwiller-le-Bas	Bas-Rhin	Grand-Est (East)		H3q1
Schw432	Schwindratzheim	Bas-Rhin	Grand-Est (East)		X2b
Schw72-15	Schwindratzheim	Bas-Rhin	Grand-Est (East)	5383-5220 cal BC	T2f
Schw72-16	Schwindratzheim	Bas-Rhin	Grand-Est (East)	5383-5220 cal BC	H1
Schw72-17	Schwindratzheim	Bas-Rhin	Grand-Est (East)		J1c2
LARZ4	Larzicourt	Marne	Grand-Est (East)	5200-4700 BC	J1c3
ORC1	Orcontes	Marne	Grand-Est (East)		J1c1b1
MDV248	Menneville	Aisne	Hauts-de-France (North)		K1a2

However, the Grossgartach (ca. 4700-4500 BCE) site of Rosheim (Ros) in Alsace, retains ties to Central Europe Early and Middle Neolithic (CEN, CEMN), as reflected by both their proximity on the PCA and their low pairwise F_{ST} values (**Figure 31b**, Table 7).

Although there are no data available for the early Neolithic in Southern France, both haplogroup frequency (**Figure 31b and 32**) and sequence-based approaches (**Figure 31c**) agree on a genetic continuity between the populations of the French Middle Neolithic (SFMN), belonging to the widespread Chasséen (~4400-3600 BCE) culture, and those of the Late Neolithic in both France (SFLNCA) and Iberia (ILNCA).

In Northern France, however, the maternal make-up of archeological sites attributed to the second part of the Middle Neolithic (~4500-3600 BCE) is heterogeneous. It is notable that none of the three groups, namely from the West to the East Esc (Escalles, Pas-de-Calais), Mch ("Michelsberg culture"), Brg (Bergheim, Alsace), are close to one another on either the PCA or the F_{ST}-based MDS (see **Figures 31b, 31c and 32**). Escalles holds an outlier position due to a skewed distribution of haplogroup frequencies where one maternal lineage (K1a) constitutes more than half of the haplogroups (8/14 individuals). In contrast, individuals from the Michelsberg culture sampled in northern France (Mch) hold a particular position on the PCA and Ward's Clustering analyses, where they are closer to HGs than any other population. This affinity is linked to the high proportion of the mitochondrial haplogroup U5b (36%), a major maternal lineage in European Mesolithic HGs (19/29), indicating a particularly high maternal contribution of HGs in this region. This pattern is also visible in Late Neolithic individuals from the sample regions, where 40% of mitochondrial lineages are from the U5b lineage. The site of Bergheim (~ 4300-3700 BCE, Alsace) is located 40 km from Rosheim, and is associated with the B.O.R.S (Bischeim occidental du Rhin Supérieur) and Michelsberg cultures. No significant difference is detected from both haplogroup frequency-based (Fisher's exact test p-value: 0.364) and sequence-based (F_{ST}=0.02, p-value =0.18424+-0.0039) approaches between the two populations. However, the maternal composition of the Bergheim group is responsible for its closeness to the population of Neolithic Great Britain (GBN).

As the Neolithic unfolds, we can observe in the French population a steady increase in the HG haplogroups, namely U5b and, to a lesser extent, U5a. Evident in particular from the second half of the Middle Neolithic (~4500 years BCE), this significant increase is not homogenous across the territory, which raises the question of the origin of these maternal lineages in the Neolithic background. In Champagne (Grand-Est) for example, the frequency of U5b haplogroups reaches 35% by the end of the IVth millennium BCE. On one archeological site in particular (Pont-sur-Seine), four out of five individuals sampled at random across the site belong to the U5 family and all fall within different haplogroups (U5b1h, U5b2c, U5b3b and U5a2d) (**Supplementary Table S4**).

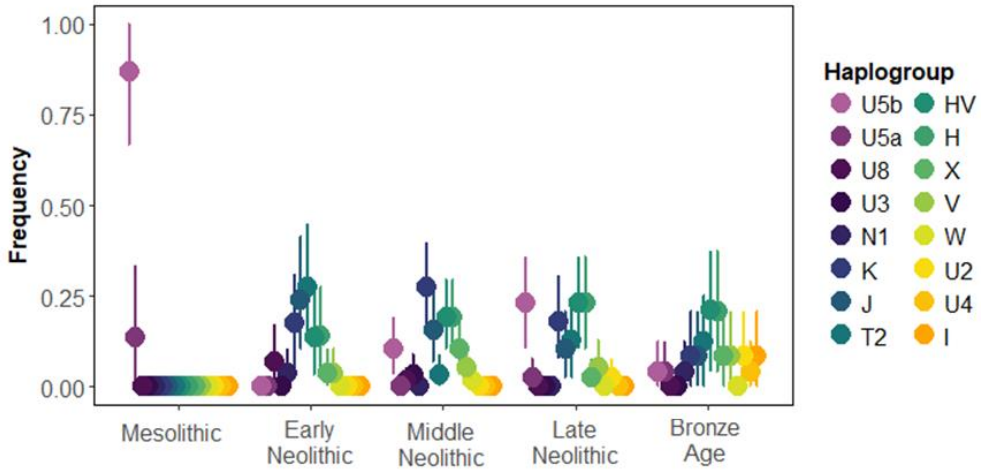


Figure 33: Changes in the average frequency of 16 core haplogroups in the ancient French population

Table 8: Linearized Slatkin's pairwise F_{ST} values between populations included in the present study

	HG	Early Neolithic				Middle Neolithic									Late Neolithic							Bronze Age				Iron Age
	HG	ANE	CEN	NFEN	IEN	CEMN	CLMN	Ros	Brg	Esc	Mch	SFMN	GMN	NFLN	SFLNCA	CLN	CCA	CBB	ILNCA	IBB	GBB	CBA	FBA	GBA	SBA	FIA
HG (25)	*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.013*	*0.000*	*0.000*	*0.018*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
ANE (24)	0.404	*	*0.410*	*0.213*	*0.431*	*0.291*	*0.171*	*0.363*	*0.707*	*0.428*	*0.225*	*0.124*	*0.611*	*0.216*	*0.218*	*0.156*	*0.048*	*0.023*	*0.031*	*0.159*	*0.339*	*0.218*	*0.278*	*0.004*	*0.030*	*0.219*
CEN (52)	0.362	0.005	*	*0.425*	*0.540*	*0.587*	*0.315*	*0.527*	*0.291*	*0.100*	*0.232*	*0.076*	*0.153*	*0.159*	*0.291*	*0.218*	*0.302*	*0.003*	*0.002*	*0.028*	*0.320*	*0.159*	*0.431*	*0.015*	*0.078*	*0.291*
NFEN (8)	0.659	0.064	0.012	*	*0.657*	*0.397*	*0.459*	*0.527*	*0.354*	*0.103*	*0.320*	*0.412*	*0.220*	*0.153*	*0.328*	*0.934*	*0.715*	*0.096*	*0.296*	*0.084*	*0.363*	*0.156*	*0.410*	*0.254*	*0.142*	*0.411*
IEN (9)	0.534	0.009	0.000	0.000	*	*0.218*	*0.359*	*0.254*	*0.641*	*0.142*	*0.220*	*0.502*	*0.264*	*0.191*	*0.302*	*0.628*	*0.235*	*0.125*	*0.339*	*0.142*	*0.319*	*0.291*	*0.425*	*0.142*	*0.162*	*0.502*
CEMN (21)	0.457	0.019	0.000	0.022	0.055	*	*0.404*	*0.882*	*0.226*	*0.198*	*0.291*	*0.217*	*0.396*	*0.256*	*0.628*	*0.218*	*0.459*	*0.156*	*0.071*	*0.147*	*0.855*	*0.254*	*0.489*	*0.142*	*0.156*	*0.414*
CLMN (8)	0.267	0.062	0.026	0.012	0.031	0.016	*	*0.329*	*0.431*	*0.104*	*0.754*	*0.752*	*0.332*	*0.587*	*0.843*	*0.459*	*0.587*	*0.762*	*0.762*	*0.587*	*0.625*	*0.860*	*0.757*	*0.955*	*0.763*	*0.507*
Ros (20)	0.454	0.011	0.000	0.000	0.043	0.000	0.028	*	*0.319*	*0.225*	*0.339*	*0.171*	*0.302*	*0.218*	*0.329*	*0.357*	*0.489*	*0.027*	*0.046*	*0.114*	*0.535*	*0.156*	*0.416*	*0.068*	*0.070*	*0.218*
Brg (13)	0.319	0.000	0.020	0.026	0.000	0.035	0.014	0.021	*	*0.251*	*0.489*	*0.554*	*0.710*	*0.459*	*0.403*	*0.737*	*0.144*	*0.155*	*0.485*	*0.457*	*0.431*	*0.540*	*0.459*	*0.103*	*0.153*	*0.411*
Esc (14)	0.551	0.005	0.063	0.174	0.111	0.055	0.160	0.046	0.043	*	*0.201*	*0.037*	*0.256*	*0.159*	*0.118*	*0.042*	*0.027*	*0.001*	*0.011*	*0.198*	*0.156*	*0.096*	*0.078*	*0.001*	*0.001*	*0.075*
Mch (11)	0.140	0.039	0.034	0.046	0.065	0.031	0.000	0.021	0.004	0.080	*	*0.235*	*0.360*	*0.763*	*0.497*	*0.254*	*0.315*	*0.291*	*0.218*	*0.763*	*0.519*	*0.628*	*0.456*	*0.456*	*0.300*	*0.137*
SFMN (39)	0.335	0.033	0.029	0.012	0.001	0.024	0.000	0.030	0.000	0.095	0.033	*	*0.302*	*0.254*	*0.708*	*0.378*	*0.125*	*0.068*	*0.860*	*0.156*	*0.428*	*0.410*	*0.395*	*0.075*	*0.075*	*0.814*
GMN (49)	0.304	0.000	0.019	0.048	0.030	0.006	0.021	0.013	0.000	0.023	0.014	0.009	*	*0.339*	*0.637*	*0.185*	*0.043*	*0.040*	*0.114*	*0.319*	*0.798*	*0.410*	*0.254*	*0.011*	*0.059*	*0.404*
NFLN (10)	0.146	0.046	0.051	0.109	0.076	0.042	0.000	0.052	0.006	0.096	0.000	0.031	0.016	*	*0.587*	*0.153*	*0.156*	*0.332*	*0.254*	*0.762*	*0.498*	*0.708*	*0.534*	*0.261*	*0.291*	*0.142*
SFLNCA (20)	0.289	0.027	0.015	0.031	0.029	0.000	0.000	0.017	0.012	0.080	0.003	0.000	0.000	0.000	*	*0.291*	*0.328*	*0.762*	*0.587*	*0.489*	*0.924*	*0.760*	*0.756*	*0.431*	*0.404*	*0.712*
CLN (16)	0.417	0.043	0.025	0.000	0.000	0.035	0.006	0.014	0.000	0.130	0.036	0.009	0.030	0.062	0.022	*	*0.331*	*0.057*	*0.254*	*0.046*	*0.344*	*0.191*	*0.372*	*0.206*	*0.219*	*0.604*
CCA (26)	0.402	0.072	0.012	0.000	0.044	0.001	0.000	0.001	0.053	0.143	0.022	0.031	0.051	0.062	0.014	0.015	*	*0.028*	*0.035*	*0.031*	*0.302*	*0.137*	*0.436*	*0.291*	*0.191*	*0.254*
CBB (108)	0.216	0.041	0.035	0.065	0.046	0.020	0.000	0.045	0.028	0.099	0.016	0.018	0.020	0.013	0.000	0.039	0.035	*	*0.027*	*0.166*	*0.456*	*0.955*	*0.432*	*0.431*	*0.715*	*0.419*
ILNCA (51)	0.287	0.054	0.050	0.030	0.018	0.044	0.000	0.053	0.002	0.126	0.034	0.000	0.022	0.030	0.000	0.020	0.046	0.020	*	*0.133*	*0.182*	*0.302*	*0.191*	*0.057*	*0.057*	*0.773*
IBB (28)	0.165	0.029	0.051	0.097	0.062	0.038	0.000	0.049	0.004	0.043	0.000	0.024	0.009	0.000	0.001	0.062	0.066	0.014	0.027	*	*0.319*	*0.762*	*0.226*	*0.096*	*0.114*	*0.159*
GBB (36)	0.289	0.009	0.008	0.017	0.023	0.000	0.000	0.000	0.005	0.046	0.000	0.003	0.000	0.001	0.000	0.012	0.012	0.001	0.016	0.009	*	*0.762*	*0.663*	*0.281*	*0.302*	*0.540*
CBA (24)	0.209	0.021	0.025	0.056	0.026	0.019	0.000	0.036	0.000	0.071	0.000	0.005	0.004	0.000	0.000	0.028	0.034	0.000	0.011	0.000	0.000	*	*0.798*	*0.628*	*0.799*	*0.373*
FBA (29)	0.260	0.014	0.003	0.012	0.009	0.001	0.000	0.005	0.004	0.069	0.005	0.006	0.013	0.000	0.000	0.009	0.003	0.002	0.016	0.017	0.000	0.000	*	*0.459*	*0.625*	*0.372*
GBA (55)	0.227	0.067	0.038	0.033	0.050	0.028	0.000	0.043	0.041	0.140	0.005	0.023	0.040	0.025	0.003	0.022	0.011	0.001	0.023	0.027	0.009	0.000	0.002	*	*0.838*	*0.288*
SBA (23)	0.250	0.060	0.036	0.055	0.042	0.032	0.000	0.051	0.034	0.135	0.020	0.033	0.043	0.022	0.007	0.023	0.024	0.000	0.035	0.030	0.011	0.000	0.000	0.000	*	*0.254*
FIA (14)	0.542	0.037	0.020	0.021	0.003	0.010	0.000	0.042	0.014	0.137	0.080	0.000	0.007	0.079	0.000	0.000	0.028	0.005	0.000	0.039	0.000	0.010	0.010	0.017	0.022	*

The lower triangle reports F_{ST} values, where significant values (i.e. supported by a p-value < 0.05) are highlighted in grey. The upper triangle reports p-values associated with each F_{ST} value after post-hoc adjustment following the Benjamini-Hochberg procedure. The number of individuals in each group is reported in brackets.

Part IV

Table 9: P-values of Fisher's exact test conducted on haplogroup frequencies

	HG	ANE	CEN	NFEN	IEN	CEMN	CLMN	Ros	Brg	Esc	Mch	SFMN	GMN	NFLN	SFLNCA	CLN	CCA	CBB	ILNCA	IBB	GBB	CBA	FBA	GBA	SBA	FIA
HG (29)	*																									
ANE (24)	0.000	*																								
CEN (77)	0.000	0.544	*																							
NFEN (10)	0.000	0.389	0.783	*																						
IEN (11)	0.000	0.681	0.959	0.908	*																					
CEMN (28)	0.000	0.196	0.631	0.581	0.572	*																				
CLMN (23)	0.000	0.025	0.139	0.544	0.323	0.636	*																			
Ros (20)	0.000	0.571	0.775	0.651	0.472	0.663	0.160	*																		
Brg (15)	0.000	0.648	0.250	0.473	0.458	0.173	0.104	0.370	*																	
Esc (14)	0.000	0.708	0.536	0.162	0.161	0.485	0.062	0.544	0.479	*																
Mch (11)	0.024	0.156	0.292	0.455	0.369	0.514	0.561	0.249	0.491	0.292	*															
SFMN (39)	0.000	0.160	0.051	0.901	0.581	0.181	0.292	0.250	0.479	0.250	0.277	*														
GMN (49)	0.000	0.337	0.265	0.577	0.423	0.499	0.338	0.683	0.770	0.937	0.627	0.447	*													
NFLN (10)	0.012	0.162	0.179	0.408	0.298	0.515	0.323	0.125	0.550	0.387	1.000	0.370	0.494	*												
SFLNCA (20)	0.000	0.107	0.417	0.576	0.514	0.869	0.827	0.301	0.300	0.353	0.596	0.677	0.885	0.550	*											
CLN (22)	0.000	0.186	0.179	0.821	0.677	0.088	0.373	0.277	0.515	0.125	0.323	0.239	0.402	0.112	0.277	*										
CCA (26)	0.000	0.065	0.694	0.708	0.550	0.514	0.594	0.515	0.046	0.124	0.469	0.124	0.301	0.100	0.417	0.370	*									
CBB (111)	0.000	0.002	0.000	0.359	0.157	0.204	0.515	0.113	0.093	0.366	0.541	0.017	0.301	0.301	0.774	0.437	0.292	*								
ILNCA (51)	0.000	0.003	0.000	0.473	0.216	0.016	0.107	0.020	0.292	0.043	0.162	0.783	0.116	0.323	0.524	0.030	0.008	0.001	*							
IBB (28)	0.000	0.030	0.013	0.075	0.043	0.195	0.216	0.047	0.233	0.544	0.734	0.088	0.482	0.860	0.530	0.026	0.039	0.177	0.060	*						
GBB (36)	0.000	0.268	0.346	0.787	0.613	0.536	0.369	0.669	0.338	0.764	0.594	0.300	0.888	0.369	0.683	0.750	0.745	0.814	0.019	0.195	*					
CBA (33)	0.000	0.079	0.012	0.338	0.303	0.219	0.369	0.121	0.368	0.578	0.631	0.107	0.499	0.578	0.677	0.515	0.120	0.985	0.024	0.447	0.775	*				
FBA (32)	0.000	0.323	0.242	0.945	0.783	0.694	0.300	0.658	0.403	0.728	0.804	0.193	0.323	0.798	0.775	0.324	0.594	0.417	0.014	0.199	0.783	0.594	*			
GBA (55)	0.000	0.000	0.000	0.468	0.160	0.088	0.624	0.050	0.017	0.044	0.627	0.009	0.034	0.233	0.499	0.301	0.516	0.843	0.001	0.044	0.353	0.628	0.369	*		
SBA (30)	0.000	0.004	0.002	0.233	0.150	0.044	0.356	0.024	0.016	0.030	0.117	0.001	0.025	0.030	0.185	0.594	0.111	0.526	0.000	0.004	0.233	0.613	0.177	0.594	*	
FIA (31)	0.000	0.014	0.041	0.449	0.472	0.075	0.300	0.063	0.160	0.014	0.034	0.219	0.245	0.011	0.277	0.805	0.095	0.160	0.044	0.002	0.245	0.107	0.022	0.027	0.153	*

P-values are reported in the lower triangle after post-hoc adjustment following the Benjamini-Hochberg procedure. The number of individuals in each group is reported in brackets.

IV. Transitioning into Bronze Age (~2300-1000 BCE)

We performed capture on 30 Bronze Age individuals across the French territory, to whom we added 1 individual who was sequenced at low coverage (RIX2) (**Table 10**, **Supplementary Table 4**). Since no geographical or temporal stratification was detected from the analysis of partial genomes from this period (see **Part V**), the 31 individuals were grouped together under the label FBA (**Figure 31a, b and c**, filled yellow diamond). During the French Bronze Age, we observed the appearance of new mitochondrial haplogroups such as U2, U4 and I. These haplogroups were described as part of the «Steppe component» in Central Europe (Haak *et al.*, 2015). Thus, the situation on the French territory seemingly differs from the one in Iberia, where no major maternal impact from the population of the Pontic steppes has been observed (Olalde *et al.*, 2018; Szécsényi-Nagy *et al.*, 2017; Valdiosera *et al.*, 2018).

Table 10: Mitochondrial haplogroups and location of individuals assigned to the Bronze Age.

Individual	Location	Department	Region	Dates	Haplogroup
BIS130	Bischwihr	Haut-Rhin	Grand-Est (East)	2000-1800 BC	K2b1b
BIS159	Bischwihr	Haut-Rhin	Grand-Est (East)	2000-1800 BC	U5a1a1
BIS382	Bischwihr	Haut-Rhin	Grand-Est (East)	2000-1800 BC	U2e2a4
BIS385	Bischwihr	Haut-Rhin	Grand-Est (East)	2000-1800 BC	T2c1d1
BIS388-1	Bischwihr	Haut-Rhin	Grand-Est (East)	2000-1800 BC	U2e1
BIS388-2	Bischwihr	Haut-Rhin	Grand-Est (East)	2000-1800 BC	X2b-226
EUG11	Laure	Aude	Occitanie (South)	2028-1878 cal BC	V
FAD9	Pépieux	Aude	Occitanie (South)	1893-1700 cal BC	K1b1a1
MAND1175	Manduel	Gard	Occitanie (South)		X2b-226
MIT1030B	Garons	Gard	Occitanie (South)	1200-1000 BC	J2a1a1
MIT1031	Garons	Gard	Occitanie (South)		J1b1a1
MIT11	Garons	Gard	Occitanie (South)	1772-1608 BC cal	K1c
MIT1167A	Garons	Gard	Occitanie (South)		T2e
MIT1167B	Garons	Gard	Occitanie (South)	1200-1000 BC	T2e
MIT1167C	Garons	Gard	Occitanie (South)	1200-1000 BC	H1e1a
NIED	Niederergheim	Haut-Rhin	Grand-Est (East)		U5b2b
OBE3626-2	Obernai	Bas-Rhin	Grand-Est (East)	2000-1700 BC	V
OBE3722	Obernai	Bas-Rhin	Grand-Est (East)	2000-1700 BC	I4a
PIR3037AB	Valros	Hérault	Occitanie (South)		U4c1a
PIR3116B	Valros	Hérault	Occitanie (South)	2336-2135 cal BC	H2a1
PIR3117	Valros	Hérault	Occitanie (South)		K2a5
PSS4170	Pont-sur-Seine	Aube	Grand-Est (East)	2196-1977 cal BC	H
QUIN234	Castelnaudary	Aude	Occitanie (South)		H3
QUIN58	Castelnaudary	Aude	Occitanie (South)		H1ah
QUIN59	Castelnaudary	Aude	Occitanie (South)		N1a1a1a
RIX15	Rixheim	Haut-Rhin	Grand-Est (East)	1700-1600 BC	T2b19
RIX4	Rixheim	Haut-Rhin	Grand-Est (East)	1750-1630 cal BC	K1a4a1a
VIGN203	Nîmes	Gard	Occitanie (South)		U5b1b1-@16192
VIGN3052	Nîmes	Gard	Occitanie (South)		T2e
VTQ14	Villeneuve-Tolosane	Haute-Garonne	Occitanie (South)		I2
RIX2	Rixheim	Haut-Rhin	Grand-Est (East)	1750-1630 cal BC	T2c1d-152

V. Iron Age France (~ 800 – 100 BCE)

For the same reason I mentioned in the previous paragraph, all French Iron Age individuals were grouped together a FIA, independently of their geographical origin (**Figure 31a, b and c,** filled orange diamond). We captured 14 individuals to whom we added 17 mitochondrial genomes obtained from the newly generated low coverage genomes (**Table 11**).

Table 11: Mitochondrial haplogroups and location of individuals assigned to the Iron Age.

Individual	Culture	Location	Department	Region	Haplogroup
Jeb8	Hallstatt	Sainte-Croix-en-Plaine	Haut-Rhin	Grand-Est (East)	J1c8a
BES1248	Hallstatt	Bessan	Hérault	Occitanie (South)	J1c8a2
NOR2B2	Halstatt C, early La Tène	Nordhouse	Bas-Rhin	Grand-Est (East)	J1c-16261
NOR2B6	Halstatt C, early La Tène	Nordhouse	Bas-Rhin	Grand-Est (East)	K1a2a
NOR3-15	Halstatt C, early La Tène	Nordhouse	Bas-Rhin	Grand-Est (East)	J1c2o
NOR3-6	Halstatt C, early La Tène	Nordhouse	Bas-Rhin	Grand-Est (East)	H7d
NOR4	Halstatt C, early La Tène	Nordhouse	Bas-Rhin	Grand-Est (East)	U5a1a1
COL11	La Tène B	Colmar	Haut-Rhin	Grand-Est (East)	H2b
BERG351-2	La Tène	Bergheim	Haut-Rhin	Grand-Est (East)	K1a-195
BERG643-2	La Tène	Bergheim	Haut-Rhin	Grand-Est (East)	K1c1
CLR05	La Tène	Le Cailar	Gard	Occitanie (South)	H-16291
CLR1	La Tène	Le Cailar	Gard	Occitanie (South)	H3-16311
CLR11	La Tène	Le Cailar	Gard	Occitanie (South)	U8
CLR12	La Tène	Le Cailar	Gard	Occitanie (South)	H1
CLR13	La Tène	Le Cailar	Gard	Occitanie (South)	W5b
PECH10	La Tène	Sigean	Aude	Occitanie (South)	H1t
PECH3	La Tène	Sigean	Aude	Occitanie (South)	V
PECH5	La Tène	Sigean	Aude	Occitanie (South)	H1ab
PECH8	La Tène	Sigean	Aude	Occitanie (South)	J1c3j
PECH9	La Tène	Sigean	Aude	Occitanie (South)	J1c5
Pey74	La Tène	Agde	Hérault	Occitanie (South)	H-152
ATT26	La Tène	Attichy-Bitry	Oise	Hauts-de-France (North)	H8
BFM265	La Tène	Bucy-le-long	Aisne	Hauts-de-France (North)	U2e1b2
COL153A	La Tène	Colmar	Haut-Rhin	Grand-Est (East)	K1a26
COL153i	La Tène	Colmar	Haut-Rhin	Grand-Est (East)	H1q
ERS1164	La Tène	Erstein	Bas-Rhin	Grand-Est (East)	H2a2a1
ERS86	La Tène	Erstein	Bas-Rhin	Grand-Est (East)	HV0
ERS88	La Tène	Erstein	Bas-Rhin	Grand-Est (East)	U5a1g
PEY163	La Tène	Agde	Hérault	Occitanie (South)	W1g
PEY53	La Tène	Agde	Hérault	Occitanie (South)	J1c11a
PT2	La Tène	Gailhan	Gard	Occitanie (South)	J1c1b1

FIA is located at the edge of both the haplogroup frequency-based PCA (**Figure 31b.**) and MDS plot based on pairwise F_{ST} values (**Figure 31c**.). This position stems from an elevated frequency of maternal lineages H and J, reaching respectively 32.3% and 25.8% of the total haplogroups in this population. A comparatively high proportion of haplogroup H is only found in Bronze Age populations from the Pontic Steppe (SBA) where it reaches 30%, while J is frequent in Late Neolithic populations from Central Europe (CLN) and Iberia (ILNCA), where it reaches 22%. While Fisher's exact test indicates that FIA is significantly different from Late Neolithic/Bell Beaker Iberia (pvalues: 0.044 and 0.0002), North France Late Neolithic (0.011), French Bronze Age (0.022) and British Bronze Age (0.027) (**Table 9**), pairwise F_{ST} values between FIA and these populations are low and not statistically supported (**Table 8**). It is noteworthy though that fixation indexes were estimated from fewer sequences than haplogroup frequencies (14 instead of 31), and that F_{ST} appears to be sensitive to sample size.

VI. Y chromosome haplogroup analysis

Among the 73 males who provided coverage of Y chromosome SNPs, 38 allowed haplogroup assignment. Among these, 13 were supported by shotgun sequencing (**Supplementary Table 8**). We further used shotgun sequencing data to assign haplogroups to 10 more individuals: MDV248 (NFEN), BUCH2 (Buc), Es97-7 (Esc); OBE3626-1, PIR3037AB, RIX2, and RIX15 (FBA) as well as BES1248, PEY163 and WET370-1 (FIA) (**Supplementary Table 8**). We reported in Tables 12 to 15 informative derived alleles obtained from either capture or shotgun sequencing (in italics), neglecting internal nodes of the Y chromosome haplogroup tree for clarity. The total number of covered positions and derived alleles can be found in **Supplementary Table 5**.

Among the successfully genotyped individuals, the only male individual from the Mesolithic was assigned to haplogroup I. We successfully retrieved Y chromosome genotypes for four male individuals from the Early Neolithic in Northern France (NFEN). While the two individuals from Schwindratzheim (Alsace, Grand-Est) belonged two haplogroups C1a2 and G2a, both the individuals from Larzicourt (Marne, Grand-Est) and Menneville (Aisne, Hauts-de-France) were assigned to haplogroup H2. Haplogroup C1a2 was described in a Mesolithic hunter-gatherer from Spain and likely introduced into the Neolithic pool through admixture with hunter-gatherers, while G2a and H2 were described in various European Early Neolithic contexts in both Central Europe and Iberia (Olalde *et al.*, 2014).

Table 12: Y chromosome haplogroup assignments for Mesolithic and Early Neolithic individuals

Individual	Position	Marker	Mutation	Coverage	Haplogroup
PER3123	15023364	M258	T->C	1	I
	7173143	L16	G->A	1	IJK
Schw72-15	6845955	V20	G->A	3	C1a2
Scwh72-17	23973594	PF3141	T->G	3	G2a
LARZ4	19360748	Z18961	A->G	1	H2
MDV248	*21940316*	*M3035*	*C->T*	*1*	*H~*
	17419493	*Y19955*	*G->T*	*1*	*H2~*
	7551850	*Y12585*	*G->A*	*1*	*H2~*
	16448046	*Z18895*	*A->G*	*1*	*H2*
	21441419	*Z18976*	*A->G*	*1*	*H2*
	7413297	*Z18841*	*T->C*	*1*	*H2*
	13986260	*Z18845*	*C->G*	*1*	*H2a1~*
	14160776	*Z18847*	*C->T*	*1*	*H2a1~*
	14268133	*Z18850*	*G->A*	*1*	*H2a1~*
	14622170	*Z18858*	*G->A*	*1*	*H2a1~*
	14685957	*Z18861*	*C->T*	*1*	*H2a1~*
	16673568	*Z18904*	*C->A*	*1*	*H2a1~*
	17547480	*Z18922*	*G->T*	*1*	*H2a1~*

Middle Neolithic in north and northeastern France is characterized by haplogroups I/I2 (13 individuals) and E1b1 (4 individuals), the latter being restricted to the site of Bergheim (Table 11). The I/I2 lineage is also found in CRE3, a Middle Neolithic individual from Southern France, and FAD3, a Late Neolithic individual.

Table 13: Y chromosome haplogroup assignments for Middle Neolithic individuals

Individual	Position	Marker	Mutation	Coverage	Haplogroup
ROS45	15023364	M258	T->C	2	I
	16638804	M438	A->G	2	I2
	7173143	L16	G->A	4	IJK
ROS102	15023364	M258	T->C	6	I
	7173143	L16	G->A	8	IJK
	8536868	L758	C->G	1	I
ROS47	15023364	M258	T->C	6	I
	16638804	M438	A->G	3	I2
	18257568	CTS8876	G->A	1	I
	7173143	L16	G->A	10	IJK
ROS78	15023364	M258	T->C	8	I
	16638804	M438	A->G	1	I2
	7173143	L16	G->A	10	IJK
BUCH2	17800761	M2936	T->C	1	H
	18182848	M2955	G->T	1	H
	19535440	M2992	A->T	1	H
	6855809	M2713	G->A	1	H
	8781791	Z13964	G->A	2	H
	14372352	Z18855	C->A	1	H2
	14921398	Z18865	G->A	1	H2
	15280036	Z18870	A->G	1	H2
	15545891	Z18878	A->C	1	H2
	15604862	L284	C->T	1	H2
	15680713	Z18879	C->T	1	H2
	15906801	Z18885	G->A	1	H2
	17706964	Z18925	A->G	1	H2
	17903963	Z18933	C->T	2	H2
	17906394	Z18934	A->T	2	H2
	19521258	Z18964	T->C	1	H2
	21387746	Z18973	A->G	1	H2
	21869856	L285	C->T	1	H2
	22590375	Z18996	T->C	1	H2
	22888894	SK1181	G->C	1	H2
	23129781	Z19011	T->C	1	H2
	23425961	Z19021	T->C	1	H2
	28789140	Z19029	G->A	1	H2
	16337179	M9893	C->T	1	H2~
	16549932	Y20188	G->A	1	H2~
	16581180	Z18898	G->A	1	H2~
	16593646	Z18900	G->A	1	H2~
	16627833	Y21631	G->A	2	H2~
	18588951	Z18942	G->A	1	H2~
	18741066	Y20189	G->A	3	H2~
	22846177	Y20193	G->A	1	H2~
	7318171	Y20230	G->A	1	H2~
	7550038	Y21584	G->A	1	H2~
	8521719	Y21662	C->G	1	H2~
	14313144	Z18852	C->T	2	H2a
	15612011	Z19078	G->C	1	H2a
	16348351	Z19082	A->T	1	H2a
	16434953	Z19083	C->T	1	H2a
	19447028	Z19107	A->T	1	H2a
	21345286	Z19109	T->C	1	H2a
	2716568	Z19065	G->T	1	H2a
	14607890	Z19075	G->T	1	H2a~
	21586014	Z19112	C->T	1	H2a~
	14622170	Z18858	G->A	1	H2a1~
	16972874	Z18909	C->T	1	H2a1~
	17518771	Z18921	G->C	1	H2a1~
	17623670	Z18924	G->A	1	H2a1~

Individual	Position	Marker	Mutation	Coverage	Haplogroup
	21274985	Z18969	T->G	1	H2a1~
	28798163	Z19030	G->C	1	H2a1~
Berg157-4	21610831	P2	G->A	1	E1b1
	21903853	L336	G->A	1	E1b1b
BERG157-1	16638804	M438	A->G	2	I2
BERG79	8430640	M5415	G->A	1	E1b1~
	15467824	M215	A->G	1	E1b1b
	21610831	P2	G->A	1	E1b1
	21893303	M78	C->T	1	E1b1b1a1
	21903853	L336	G->A	1	E1b1b
BERG157-3	15467824	M215	A->G	1	E1b1b
	21610831	P2	G->A	2	E1b1
	21778998	M96	C->G	1	E
	21893303	M78	C->T	2	E1b1b1a1
BERG157-5	7173143	L16	G->A	5	IJK
	15023364	M258	T->C	4	I
	16638804	M438	A->G	5	I2
BERG157-9	15467824	M215	A->G	1	E1b1b
	21893303	M78	C->T	1	E1b1b1a1
BERG157-7	7173143	L16	G->A	10	IJK
	9108252	P187	G->T	4	F
	15023364	M258	T->C	3	I
	16638804	M438	A->G	5	I2
Es48b		L16	G->A	7	IJK
		M258	T->C	4	I
Es42-3		L16	G->A	8	IJK
		M258	T->C	4	I
Es97-1	7629583	F3689	A->G	1	IJK
	21220741	FGC1560	T->C	1	IJ
	7694266	FGC1564	G->A	2	IJ
	8590752	P127	C->T	1	IJ
	9921164	FGC1578	G->A	1	IJ
	10051801	FGC2417	G->A	1	I
	14073053	CTS1800	G->A	1	I
	14646409	CTS3076	C->T	1	I
	14974451	L1197	C->T	1	I
	15089989	CTS3641	T->C	1	I
	15377802	CTS4077	G->A	1	I
	15536870	CTS4273	C->T	1	I
	15862842	CTS4848	C->T	1	I
	15937959	CTS4982	C->T	1	I
	16171560	CTS5263	G->A	1	I
	16836548	CTS6344	G->A	1	I
	17467526	PF3759	G->A	1	I
	18172947	CTS8742	A->G	1	I
	18582617	CTS8963	C->T	1	I
	19233673	CTS10058	A->G	1	I
	21839183	PF3814	A->G	1	I
	21939618	PF3817	G->A	1	I
	23154034	L847	C->T	1	I
	2688442	CTS48	T->A	1	I
	2974782	PF3574	A->C	1	I
	7712917	PF3642	A->T	2	I
	8262092	FGC2413	C->T	1	I
	8484606	PF3661	C->A	1	I
	8485677	FI3	C->A	1	I
	8643763	PF3665	A->G	1	I
PSS3072	15023364	M258	T->C	1	I
	7173143	L16	G->A	1	IJK
BLP10	15023364	M258	T->C	2	I
	7173143	L16	G->A	4	IJK
CRE3	16638804	M438	A->G	2	I2
	7173143	L16	G->A	2	IJK
FAD3	15023364	M258	T->C	1	I
	7173143	L16	G->A	4	IJK

Our capture dataset comprised one Bell Beaker associated individual from Northern France: CBV95. While it could not be included in the analytical framework based on mitochondrial lineages, its Y haplogroup provides the earliest clear evidence of the presence of haplogroup R1b in France around 2500 BCE (**Table 12**). This lineage was associated with the arrival of Steppe migrants in Central Europe during the Late Neolithic. Signatures of this massive migration were described in other parts of Europe and in Bell Beaker individuals from Southern France, but was almost absent in Iberia (Olalde *et al.*, 2018).

Table 14: Y chromosome haplogroup assignments for Bell Beaker and Bronze Age individuals

Individual	Position	Marker	Mutation	Coverage	Haplogroup
CBV95	17400785	L388	G->A	1	R1b1a1
	23631629	F652	C->A	1	R
PSS4170	9170545	M415	C->A	2	R1b1
BIS3882	18617596	CTS9018	C->T	1	R1b1a1a
OBE3626-1	18017528	F459	G->T	1	R
	21219443	M760	A->G	1	R
	24360964	F765	G->A	1	R
	7548900	F82	G->A	1	R
	15377120	CTS4075	A->G	1	R1
	7771131	P238	G->A	1	R1
	9989615	P231	A->G	1	R1
	16426937	CTS5676	C->G	1	R1b1
	23845409	PF6270	T->C	1	R1b1
	7081561	CTS910	C->T	1	R1b1
	14637352	CTS3063	T->C	1	R1b1a
	15803415	CTS4764	G->A	2	R1b1a
	15239181	CTS3876	G->C	1	R1b1a1a
	17986687	PF6475	C->A	1	R1b1a1a
	18394634	L752	T->C	1	R1b1a1a
	14042701	CTS1738	C->T	1	R1b1a1a2
	14317555	CTS2466	G->A	1	R1b1a1a2
	16971648	CTS6532	T->G	1	R1b1a1a2
	17594966	CTS7659	C->G	2	R1b1a1a2
	17813541	CTS8052	C->T	1	R1b1a1a2
	18047475	L749	A->C	1	R1b1a1a2
	18117193	CTS8627	C->T	1	R1b1a1a2
	23476936	PF6525	G->T	1	R1b1a1a2
	24444622	L1351	C->T	1	R1b1a1a2
	28771116	CTS12972	C->G	1	R1b1a1a2
	2897433	CTS329	C->G	1	R1b1a1a2
	6912992	CTS623	T->G	1	R1b1a1a2
	8411202	PF6434	A->G	1	R1b1a1a2
	8502236	L51	G->A	1	R1b1a1a2a1
	9084870	PF6540	G->T	1	R1b1a1a2a1a
	8469661	S359	G->A	1	R1b1a1a2a1a2a5~
	9395960	BY116	G->A	1	R1b1a1a2a1a2c1a1d2
RIX2	9889199	M651	G->A	1	R
	22722580	L1349	T->C	1	R1b1
	23992762	PF6272	C->A	1	R1b1
	8110520	PF6248	T->A	1	R1b1
	17545608	CTS7585	G->T	1	R1b1a
	15239181	CTS3876	G->C	1	R1b1a1a
	15286480	PF6459	G->C	1	R1b1a1a
	18656508	P297	G->C	1	R1b1a1a
	14079811	L762	T->A	1	R1b1a1a2
	14317555	CTS2466	G->A	1	R1b1a1a2
	17813541	CTS8052	C->T	1	R1b1a1a2
	23379254	CTS11948	G->A	1	R1b1a1a2
	28771116	CTS12972	C->G	1	R1b1a1a2
	9464078	PF6438	C->T	1	R1b1a1a2
	6753511	L23	G->A	1	R1b1a1a2a

Individual	Position	Marker	Mutation	Coverage	Haplogroup
	14641193	L52	C->T	1	R1b1a1a2a1a
	18907236	P310	A->C	1	R1b1a1a2a1a
	17951958	S20434	C->T	1	R1b1a1a2a1a1g
	22157311	P312	C->A	1	R1b1a1a2a1a2
	21666996	FGC31961	G->A	1	R1b1a1a2b2
	14300457	CTS2426	G->A	1	R
	15594523	F295	A->G	1	R
	17930099	CTS8311	C->A	1	R
	19267344	P285	C->A	1	R
	21409706	P227	G->C	1	R
	22934109	CTS11075	A->G	1	R
	23202551	CTS11647	C->G	1	R
	8027859	M628	G->C	1	R
	14424045	CTS2680	C->T	1	R1
	16426937	CTS5676	C->G	1	R1b1
	22722580	L1349	T->C	1	R1b1
	2686555	CTS46	G->A	1	R1b1
	17400785	L388	G->A	1	R1b1a1
	16005138	CTS5082	A->C	1	R1b1a1a
	17732408	CTS7904	T->C	1	R1b1a1a
RIX15	17755905	CTS7941	G->A	1	R1b1a1a
	21447844	PF6501	A->T	1	R1b1a1a
	10062719	PF6441	C->G	1	R1b1a1a2
	14317555	CTS2466	G->A	1	R1b1a1a2
	15740440	PF6462	A->G	1	R1b1a1a2
	17281258	E101	T->G	1	R1b1a1a2
	19179540	L1353	G->A	1	R1b1a1a2
	19417394	CTS10349	A->C	1	R1b1a1a2
	21993844	PF6508	G->A	1	R1b1a1a2
	23085375	CTS11371	C->G	1	R1b1a1a2
	23124367	CTS11468	G->T	1	R1b1a1a2
	24394612	PF6527	G->A	1	R1b1a1a2
	2897433	CTS329	C->G	1	R1b1a1a2
	5166408	PF6411	A->G	2	R1b1a1a2
	17589518	CTS7650	C->T	2	R1b1a1a2a1a
	9084870	PF6540	G->T	1	R1b1a1a2a1a
	17285993	P224	C->T	1	R
	17723850	CTS7880	C->T	1	R
	21409706	P227	G->C	1	R
	22687547	CTS10663	A->T	1	R
	22934109	CTS11075	A->G	1	R
	23134896	M799	C->T	1	R
	23631629	F652	C->A	1	R
	2810583	CTS207	A->G	1	R
	7647357	P242	G->A	1	R1
EUG11	8633545	P245	T->C	1	R1
	19504659	L820	T->A	1	R1b1a
	22889018	L754	G->A	1	R1b1a
	17594966	CTS7659	C->G	1	R1b1a1a2
	20838224	PF6496	A->G	1	R1b1a1a2
	21993844	PF6508	G->A	2	R1b1a1a2
	23242935	L1348	G->C	1	R1b1a1a2
	8233186	L483	C->T	1	R1b1a1a2
	8826595	PF6436	C->T	1	R1b1a1a2
	23268859	PH5393	G->A	1	R1b1a1a2a1a2c1a5d2
MAND1175	15590342	P225	G->T	1	R1
	9170545	M415	C->A	1	R1b1
PIR3037AB					R1b1a1a2
PIR3116B	17839981	CTS8116	G->A	1	R1
	9170545	M415	C->A	1	R1b1
FAD9	9170545	M415	C->A	1	R1b1
FAD2	9170545	M415	C->A	1	R1b1
QUIN58	17461478	CTS7400	T->C	1	R1b1a1a2
	21804435	M1261	C->T	1	P
	2803717	CTS196	C->T	1	P

Individual	Position	Marker	Mutation	Coverage	Haplogroup
	9170545	M415	C->A	1	R1b1

The R1b paternal lineage then becomes predominant during Bronze (12/12 individuals) and Iron Age (7/10 individuals), in northern and southern France alike. The only exceptions in our dataset are the Iron Age individuals NOR2B2, BES1248 and WET370-1, who carry respectively the Y haplogroups G2a, I1 and H2a1, though the latter yet needs to be accurately dated (**Table 12 and 13**).

Table 15: Y chromosome haplogroup assignments for Iron Age individuals

Individual	Position	Marker	Mutation	Coverage	Haplogroup
	14028148	L31	C->A	1	G2a
NOR2B2	15027529	M201	G->T	1	G
	16325291	CTS5504	T->C	1	G
	18182848	M2955	G->T	1	H
	23153863	M3070	T->A	1	H
	21940316	M3035	C->T	1	H~
	15260651	Z18869	A->G	2	H2
	19244108	Z18958	A->G	1	H2
	22202792	Z18992	G->T	1	H2
	14328924	Z18854	G->A	1	H2
	19244108	Z18958	A->G	1	H2
	21715967	Z18980	T->C	1	H2
WET370-1	21890111	L286	T->C	1	H2
	22613270	Z18997	T->A	1	H2
	23043740	Z19009	A->G	1	H2
	14009773	Z18846	T->C	1	H2~
	16581180	Z18898	G->A	1	H2~
	8441583	SK1189	A->G	1	H2~
	8676565	Y21659	A->G	1	H2~
	9105625	Z41268	C->G	1	H2~
	13678185	FGC17422	C->T	1	H2~
	17796410	Z18929	G->A	1	H2a1~
Pey74	15581983	M207	A->G	2	R
	15667208	FGC1168	G->C	1	R
	22687547	CTS10663	A->T	1	R
	23134896	M799	C->T	1	R
	24360964	F765	G->A	1	R
	6868118	F47	G->A	1	R
	9889199	M651	G->A	1	R
	14424045	CTS2680	C->T	1	R1
PEY163	7671535	F93	C->T	1	R1
	14273103	PF6255	T->G	1	R1b1
	17755905	CTS7941	G->A	1	R1b1a1a
	23403749	CTS11985	G->A	1	R1b1a1a
	14005779	PF6448	G->A	1	R1b1a1a2
	18381735	PF6482	A->G	1	R1b1a1a2
	24444622	L1351	C->T	1	R1b1a1a2
	9464078	PF6438	C->T	2	R1b1a1a2
	9452025	Z17558	G->A	1	R1b1a1a2a1a2c1a5c1b1a1
	10051801	FGC2417	G->A	1	I
	13804066	Z16985	G->C	1	I
	14337364	CTS2514	T->C	1	I
	15595624	CTS4340	G->A	1	I
	16548548	CTS5908	G->A	1	I
BES1248	16567253	CTS5946	A->G	1	I
	17548890	CTS7593	G->A	1	I
	17818847	CTS8064	G->A	1	I
	19048602	L41	G->A	1	I
	19233673	CTS10058	A->G	1	I
	22845794	CTS10941	A->G	1	I

Individual	Position	Marker	Mutation	Coverage	Haplogroup
	6926038	CTS646	T->A	1	I
	7321418	CTS1301	C->T	1	I
	7681156	PF3640	T->A	1	I
	8484606	PF3661	C->A	1	I
	9950316	FGC9452	C->A	1	I1~
	14071004	L841	T->A	1	I1
	16392707	L842	G->T	1	I1
	16677210	CTS6109	C->T	1	I1
	17226095	CTS6992	C->T	1	I1
	19048605	L75	T->A	1	I1
	23269094	CTS11783	T->C	1	I1
	23303290	L843	A->G	1	I1
	2683242	CTS40	C->T	1	I1
	7066305	CTS883	C->T	1	I1
	7742116	FGC2425	T->A	1	I1
	7784699	Z2718	C->T	1	I1
	8554437	Z2727	C->T	1	I1
	8629665	Z2729	C->T	1	I1
PECH10	9170545	M415	C->A	1	R1b1
PECH3	6766034	PF6418	C->T	1	R1b1a1a
CLR1	13887941	L407	G->A	1	R1b1a1a2
	21867787	M45	G->A	1	P1~
	9170545	M415	C->A	1	R1b1
PECH8	17013730	PF6466	G->A	1	R1b1a1a2
	17930099	CTS8311	C->A	1	R
	22475524	BY194	A->C	1	R1b1a1a2a1a2c1a1a1a1a1~
PECH5	10062719	PF6441	C->G	1	R1b1a1a2
	17839981	CTS8116	G->A	1	R1
	19054889	L757	C->T	1	R1b1a1a2
	9093383	M1195	G->C	1	P

VII. Phenotypic changes from the Mesolithic to the Bronze Age

To unravel the timelines of the evolution of genetic adaptation to a changing lifestyle, we analyzed 73 autosomal loci genotyped across 149 individuals producing the first nuclear data for a prehistoric population on the territory of present-day France (**Figure 34**). These loci encompass variants involved in both physical and physiological traits associated with the Neolithic transition, often showing signatures of positive selection in present-day Europeans (Details on the targeted positions are given in **Supplementary Table 1**).

Among the phenotypic traits of interest, we focused our attention on genetic variants involved in eye and skin pigmentation (such as SLC24A5, SLC45A2, GRM5, HERC2, IRF4, TYR) (Beleza *et al.*, 2013; Walsh *et al.*, 2013). It has been stated previously that Mesolithic HGs and Anatolian Neolithic farmers differ in their pigmentation, with the latter carrying derived alleles responsible for a light skin that are near fixation in Europe today (Mathieson *et al.*, 2015), although this might correspond to an over-simplification of the complexity of the genetic determinism of these traits (Hysi *et al.*, 2018). HGs on the other hand, were described as having a rather dark skin and light eye color (Brace *et al.*,

2018; Olalde *et al.*, 2014). The new dataset that we assembled reveals interesting patterns for the transition between the Mesolithic and the Neolithic. The Mesolithic individuals from whom we retrieved reliable genotype information after stringent filtering all possess the ancestral pigmentation variant for loci SLC45A2 and GRM5, which had been associated with a dark skin tone. In contrast, the derived alleles of these two loci are found at significantly higher frequency in the French Neolithic population (SLC45A2: 24.9%, GRM5: 38.4%), although they remain far from the frequency reached by these alleles in present—day Europe, where they are near fixation in the population (93.8 and 68.9%, respectively). In contrast, SLC24A5, described as the principal mutation associated with light skin pigmentation, is found at 96% in our Neolithic population, which is close to its current frequency in Europe (99%). These results suggest different evolutionary timelines of these mutations in France.

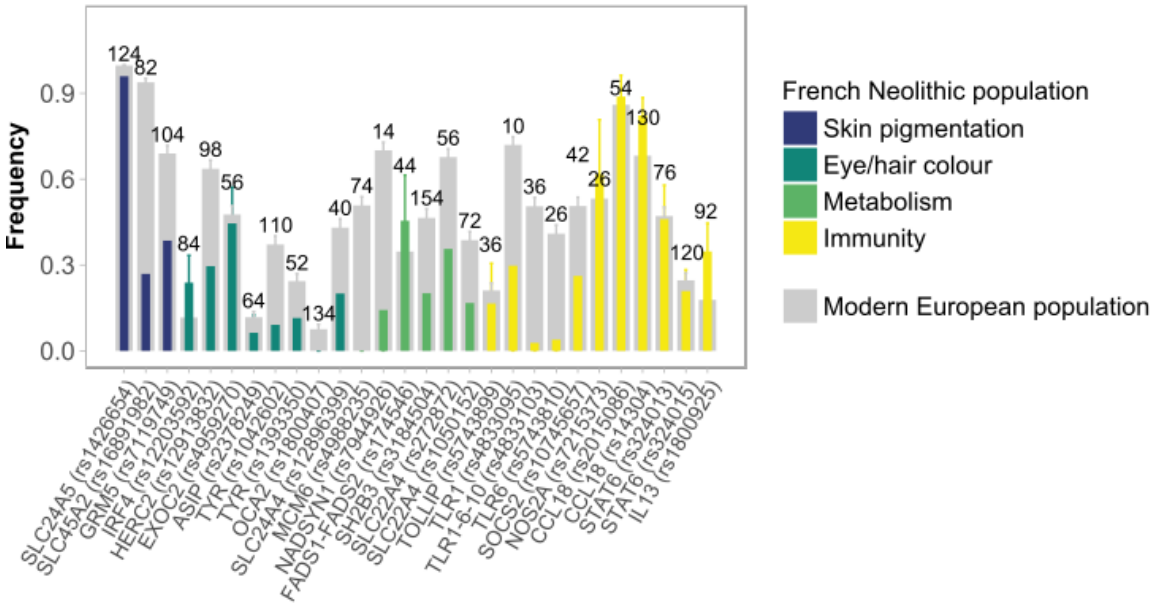


Figure 34: Frequency of the derived allele for several types of genetic markers in Neolithic French and present-day Europeans.
Numbers indicated above the bar chart correspond to the total number of observations of the derived allele for each variant in the Neolithic population. The color code reflects the function associated with the considered locus.

The analysis of the alleles involved in the pigmentation of the eye color shows that all Mesolithic individuals harbor derived alleles for variants in HERC2 and IRF4, associated with the blue eye color. At the time of the Neolithic, the frequency of these two variants drops significantly, reaching respectively 29.6% and 24.4% (95% CI after 10000 resampling: 20.5-38.8% and 15.6-33.3%). Interestingly, during the Bronze Age and in

Europe today, these two variants show different frequencies. While HERC2 rises in frequency again from 37.5% during the Bronze Age to finally reach more than 63% in present-day Europe, IRF4 further decreases from 23.81% during the Bronze Age to reach a frequency of 11% at present. It is noteworthy that no derived alleles could be detected, from the Mesolithic to the Bronze Age, among the different targeted loci from the MC1R gene, involved in the characters associated with red hair color in modern Europeans. In present-day Europe, the frequency of the derived alleles for these various loci ranges from 1 to 7 % (Gerstenblith *et al.*, 2007).

None of the genotyped Neolithic individuals carried the mutation responsible for the persistency of lactase in Europeans, which is expected for the Neolithic (Allentoft *et al.*, 2015; Haak *et al.*, 2015)(Haak *et al.*, 2015; Allentoft *et al.*, 2015). Since we retained the corresponding data for a single Bronze Age individual only after variant filtering, we cannot conclude on the presence of this allele in Bronze Age individuals in France.

We analyzed several innate immunity-related variants that show signs of recent positive selection in modern European populations (Hunt *et al.*, 2008; Olalde *et al.*, 2014; Yasuda *et al.*, 2007; Zhou *et al.*, 2012). Indeed, pathogens are a major driving force of our evolution, therefore populations of different provenance and lifestyle are expected to show different profiles of allele frequencies for such loci. For a number of genes involved in the immune response (NOS2A, TOLLIP, CCL18, STAT6, IL3), the frequency of derived alleles was comparable to that found in present-day Europeans, whereas frequencies of variants in the TLR1-6-10 locus are out of the range of present-day Europeans (Barreiro *et al.*, 2009).

Chapter III. Discussion

I. Continuity with both the Danubian and the Mediterranean Neolithic routes

The various haplotypes of mitogenomes we obtained from both autochthonous Mesolithic HGs who inhabited Western Europe since the end of the Ice Age (Fu *et al.*, 2016) as well as from early farmer individuals preserved in Neolithic contexts in southern, northwestern and northern France argue in favor of demographic processes accompanying the transition between the Mesolithic and the Neolithic. Indeed, the mitochondrial haplotypes of the French Mesolithic HGs and of the early French Neolithic farmers are distinct indicating that the corresponding individuals belonged to two different populations. Since the French Neolithic mitochondrial haplotypes had also been found in both Neolithic farmers from Central Europe and Iberia, our results are consistent with the arrival of two distinct Neolithic migration waves into France, in support of archeological results (Demoule *et al.* 1986) and paleogenomic data from other geographic regions (*e.g.*, Olalde *et al.*, 2015; Lipson *et al.*, 2017). The first wave coincides with the expansion of the LBK culture. Indeed, mitochondrial variability observed on the Grossgartach site of Rosheim (4750-4600 BCE) in eastern France appears to be close to the one found in Central Europe, except for the scarcity of haplogroup N1a1a1, compared to what was found previously for Central Europe (Haak, 2005). This particular haplogroup occurred only once on this site (Ros43) and nowhere else in our entire Neolithic dataset. Despite being found in one individual from the early Neolithic site of Els Trocs in Spain (I0412, Haak/Lazaridis *et al.* 2015), haplogroup N1a was preponderant in Central Europe where it reached 23% of the total lineages identified for the LBK, making it a signature of the Danubian migration wave, along with haplogroup T2. This local continuity, suggested in another study based on HVR variation (Rivollat *et al.* 2015), fits archaeological data for the region, where cultures such as Grossgartach and later Rössen are directly derived from the LBK culture that immigrated in eastern France around 5300 BCE (Jeunesse, 1998).

Another wave must have travelled along the Mediterranean coastline, to arrive in southern France and ultimately reach the Iberian Peninsula, suggested by the maternal proximity between both populations throughout the Neolithic. This migration wave

coincides with the expansion of the Cardial culture that expanded northward along the Rhône valley and evolved during the Middle Neolithic into the Chasséen cultures (*e.g.*, Vaquer 2010). While we lack samples for the Cardial period, our Middle and Late Neolithic results reveal genetic homogeneity between both Spain and southern France that persists until the Late Neolithic, matching the archaeological record for these regions where similar cultural and technological advances could be observed on both sides of the Pyrenees. On top of cultural exchanges, this genetic continuity implies a high mobility of people along with goods along the Mediterranean Sea during the Neolithic.

II. A contrasted picture across the French territory as the Neolithic progresses

In contrast to the homogeneity observed in Southern France, the second part of the Middle Neolithic in Northern France shows geographic stratification. Our data reveal a genetic discontinuity between Early and Middle Neolithic populations in Alsace, with an increased affinity of individuals from Middle Neolithic Bergheim with Neolithic individuals from Great Britain (**Figure 31b and 31c**). We do not observe any maternal affinity between the Middle Neolithic from Northern France and its Southern France counterpart. As individuals sampled in the Paris Basin around this period are lacking in our dataset, the hypothesis that places the origin of the Michelsberg into the contact of the Chasséen culture, originating from southern France, and groups from the Paris Basin such as the Cerny culture (Jeunesse *et al.* 1998), supported by the conclusions of Beau *et al.* 2017 on the neighboring site of Gougenheim (Beau *et al.* 2017), remains unsolved.

The site of Escalles, located atop the Mont d'Hubert, dominates a vast area including the Cap Blanc Nez and the Pas-de-Calais detroit, thus offering a clear view on the British coast. This proximity with Great Britain incited archaeologists to speculate about possible interactions between populations from both sides of the Channel. The mitochondrial haplotypes from the Escalles individuals, however, are different from the Neolithic individuals of Britain (Olalde *et al.*, 2018). This genetic discontinuity does not lend support of to the hypothesis of genetic continuity between the populations on mainland Europe and the British Isles. In particular, the site of Escalles, under the influence of both the Michelsberg and the Chasséen culture, holds an outlier position in our clustering analyses (**Figure 31b and 31c**). The site itself is unusual for the period and

the human remains show the same signs of butchering and burning as the faunal remains. As human remains were scattered, the chance of analyzing a combination of bones (i.e. left and right temporal, or femur and temporal) from the same individual was significant, which led us to consider only results that could be attributed unambiguously to a single individual. Nevertheless, one particular maternal lineage, namely K1a4a1, is overrepresented in this population. Assuming individual unearthed on the site were locals, this might result from endogamy, consistent with the remote location of the site, or a preferred particular matrimonial regime in this particular group. We caution that the number of individuals from this site is small (14), therefore biases in haplogroup frequencies may be exacerbated. Therefore, this site alone cannot be used as a proxy for the Middle Neolithic in the region. In this respect, the contemporaneous site of Bergheim in Alsace appears to be more promising to solve this question, as analyses revealed that its maternal composition resembles that of Neolithic Great Britain. Thus, a Northern French population related to Middle Neolithic Alsatians could be the source of the peopling of Great Britain during the IVth millennium BCE.

The site-specific variability in the composition and frequency of the various haplotypes result highlights the potential heterogeneity across archaeological sites that belong to the same culture, but where communities of various sizes possibly embraced different matrimonial traditions.

III. Interactions with hunter-gatherers

Another variable parameter across sites is the fraction of maternal lineages inherited from the autochthonous Mesolithic HGs who inhabited Europe before the migration of early farmers from both the Danubian and the Mediterranean currents. Hunter-gatherer maternal lineages such as U5b/U5a are absent from the individuals preserved in Early Neolithic samples contexts from northeast France, and are only detected in the skeletons dating from the second half of the Middle Neolithic onwards to further increases in frequency by the Late Neolithic. This late appearance in Neolithic contexts in France of the HG mitochondrial lineages relative to the lineages of immigrant farmers originating from Anatolia is similar to what is observed in Central Europe, where the process of assimilation of autochthonous populations was at times delayed for centuries (Brandt *et al.*, 2013; Haak *et al.*, 2015; Lipson *et al.*, 2017).

The French Mesolithic population is characterized mainly by mitochondrial haplogroups U5b. U5b1h is the second most common haplogroup among Mesolithic HGs from Western Europe, and is the most frequent U5b sub-haplogroup among individuals from the Middle to the Late Neolithic in Northern France. U5b2c was found among HGs (La Braña, Kotias, Olalde *et al.* 2015, Jones *et al.* 2015), but individuals from Early Neolithic Hungary (I1506, *Gamba et al. 2014*), Late Neolithic and Bronze Age Czech Republic (I7198, I5025, I7275, Olalde *et al.* 2018) and Germany (I6591, Olalde *et al.* 2018) also share this haplogroup. Other occurrences of U5b3 are Chalcolithic individuals from Spain (ATP2, I6475, I0262, Gunther *et al.* 2015, Olalde *et al.* 2018), Late Neolithic individuals from Central Europe (I7211, I3600, Olalde *et al.* 2018) and Bronze Age individuals from Great Britain (I2458, I2654, Olalde *et al.* 2018). This haplogroup, identified for the first time in the modern Sardinian population (ca. 5% in modern Sardinians and 0/21 ancient Sardinians, Olivieri *et al.* 2017), is thought to have occurred along the Mediterranean coast after spreading from an Italian refuge during the Holocene (Pala *et al.* 2009). Even more puzzling is the presence of U5a2d in the site of Pont-sur-Seine in northern France, which elsewhere only occurs in two HGs from Motala, Sweden (I0015, I0017, Lazaridis *et al.* 2015) and one individual from Neolithic Great Britain (I4950, Olalde *et al.* 2018). Another U5a sub-haplogroup (U5a2), however, has been found among French HGs (LesCloseaux3, Mareuil LesMeaux1, Posth *et al.* 2016), so that a wider distribution of this maternal lineage across Europe cannot be excluded.

The difference in the modality of assimilation of hunter-gatherers into farming communities across Europe reveals complex interactions, potentially dictated by local factors such as climate and access to resources (Gonzàles-Fortes *et al.* 2017, Valdiosera *et al.*, 2018). Consistent with this idea, the HG maternal contribution into Neolithic populations was larger in northern France than in the eastern France, with 35% of sampled individuals from Champagne between 4300 and 2200 BCE belonging to haplogroup U5, for only two individuals out of 13 in Alsace. This observation, suggesting close interactions between farmers and hunter-gatherers, matches the archaeological data for this region, where a tradition of hunting was maintained during the Neolithic (Achard-Corompt *et al.*, 2010).

IV. Influence of the population from Eurasian steppe during the onset of Bronze Age and Iron Age

During the Bronze Age, both our mitochondrial and Y chromosome data testify of the arrival on the French territory of lineages that have been identified in the Pontic-Caspian steppe region, although the signature is more apparent through paternal lineages (Haak *et al.*, 2015). This indicates the influence of pastoralists of the Pontic steppes associated with the Kurgan (Yamna) culture, the so-called Yamnaya. Individuals who derived part of their ancestry from these pastoralists must have reached France on its westernmost front during the Bronze Age. This influence persists through the Iron Age and to this day, where the paternal lineage R1b remains preponderant in the population.

V. Adaptation over time to new environments and lifestyles

Major cultural transitions and demographic events such as the Neolithic and later Bronze Age had a significant impact on ancient populations, through the massive migration of people and their adaptation to new environments and new ways of subsistence. Traces of these events still linger in the genomes of present-day Europeans, as a variety of phenotypes and signatures of positive selection. Our results provide a picture of a Neolithic population from the territory of present-day France, that reveals selection on loci involved in pigmentation, diet and immunity (as described in Mathieson *et al.* 2015) that are consistent with adaptation to high latitudes and a new diet. For a number of loci, especially innate immunity related variants (TOLLIP, NOS2A, CCL18, IL13, STAT6) and pigmentation (SLC24A5), the allelic frequencies in the Neolithic population are comparable to those observed in modern European populations, indicating that selection on these loci predates the Neolithic. Some of these genes are notably involved in the Th-2 immune signaling, an innate immunity pathway elicited by allergic reactions and parasitic worm infections. Interleukin-13 (IL13) along with other cytokines is secreted by T helper 2 (Th2) cells (McKenzie *et al.*, 1998) and was shown to upregulate the T cell attracting chemokine CCL18, involved in the pathogenesis of a variety of diseases (Baay *et al.*, 2011; van Lieshout *et al.*, 2006; Van Lieshout and MacQueen, 2008). Another upregulator of the Th-2 immune signaling pathway is the STAT6 gene, that harbors common asthma-associated genetic variants found to enhance resistance to *Ascaris* infection (Peisong *et al.*, 2004). Helminths were present in the environment of our

ancestors, even before the emergence of humans as a species (Fumagalli *et al.*, 2011), and are likely to have represented a stable threat to human populations throughout time up to the present. Therefore, they potentially exerted a selective pressure on specific immunity pathways, translating into strong genetic signatures in present-day populations. This genetic background combined with the lack of exposure to parasites in industrialized societies resulted in a susceptibility to the development of autoimmune and allergic conditions in the corresponding populations (Dunne and Cooke, 2005; Fumagalli *et al.*, 2009). Candidates for positive selection surrounding the onset of farming are loci such as TLR1-6-10, possibly associated with resistance to mycobacteria (Barreiro *et al.*, 2009), or variants associated with coeliac disease in SH2B3 and SLC22A4 that could have been the target for positive selection owing to the role they play in inflammations and the elimination of a wide array of environmental toxins (Huff *et al.*, 2012). We also report moderate frequencies in NADSYN1 and FADS1-2, respectively involved in fatty acid metabolism and circulating vitamin D levels, which also suggest selection related to variation in diet and environments with lower insolation, respectively (Mathieson *et al.*, 2015).

Finally, neither Mesolithic HGs nor Neolithic farmers on the French territory carried the allele allowing for lactase persistence during adulthood, in agreement with studies of Neolithic populations further East that reported the presence of this allele only by the end of the Bronze Age (Allentoft *et al.*, 2015; Haak *et al.*, 2015). Thus, French Neolithic farmers were not able to digest fresh cow milk.

Chapter IV. Conclusion

We report for the first time a combination of mitochondrial and nuclear genomic data for 193 individuals sampled across the territory of present-day France, spanning the Mesolithic, Neolithic, Bronze and Iron Age. Between 5000 and 4500 BCE, the maternal composition of populations from Eastern and Southern France reflects the population they originated from: Danubian LBK in the east as Northern France was described as associated with the continental LBK cultural complex, and the Mediterranean Cardial in the south since Southern France was part of the Cardial complex that spread along the Mediterranean Sea. As the Neolithic progresses, farmers start harboring a variety of maternal lineages and Mesolithic hunter-gatherer haplogroups are found alongside Near Eastern lineages on almost every sampled archeological site. The extent of this assimilation process is variable, contributing to differentiate populations from one another. While the population from Southern France remains homogeneous for the duration of the Neolithic, the maternal makeup of Northern and Eastern France go through a transition as a widespread cultural complex establishes, namely the Michelsberg. While the place of origin of this complex remains unclear, populations that are associated with it are heterogeneous in their maternal composition, as contemporaneous groups sampled in various locations show a very distinct proportions of the different maternal lineages. With the advent of Bronze Age, new maternal lineages enter the population but remain at low frequency, while the appearance of Y chromosome haplogroup R1b testifies alone of the migration into France of individuals harboring the genetic signature of the Pontic Steppe. Nuclear genotypes obtained in our dataset indicate undergoing selection on a variety of loci associated with skin pigmentation, metabolism and immunity, but also older episodes where the allelic frequency observed in Neolithic French is comparable to that of present-day Europeans. While we acknowledge that mitochondrial DNA results alone need to be interpreted with caution when attempting at constructing narratives about past populations, the combination of mitochondrial, Y chromosome and autosome data we report here offer an unprecedented glimpse into prehistoric French populations.

Part V. Paleogenomic overview of the peopling of France from the Mesolithic to the Iron Age

Chapter I. Introduction

Ancient genomes have proved to be valuable tools to describe past populations, either through their complete or partial sequence. (Fu *et al.*, 2016; Keller *et al.*, 2012; Lazaridis *et al.*, 2014; Olalde *et al.*, 2014; Sánchez-Quinto *et al.*, 2012; Skoglund *et al.*, 2012). Diachronic series have investigated temporal genome-wide dynamics within defined European regions such as Central Europe (Gamba *et al.*, 2014), Great Britain (Olalde *et al.*, 2018) or Iberia (Valdiosera *et al.*, 2018), unravelling the complexity of large scale cultural transitions such as the advent of both the Neolithic and the Bronze Age. On the one hand, autochthonous hunter-gatherers were shown to be falling outside the range of modern European genetic variation, while Neolithic farmers were closer to Southern Europeans, particularly present-day Sardinians (Keller *et al.*, 2012; Skoglund *et al.*, 2012). Then, by the onset of Bronze Age, the affinity of populations from Central and Northern Europe shifted towards their present-day counterparts through a massive gene flow from a population originating in the Pontic-Caspian steppe (Haak *et al.*, 2015).

In **Part IV**. we explored the first diachronic series of ancient genomic data available across key regions of present-day France, aiming at tracking the impact of these large-scale genetic shifts in the ancient French. While we gained valuable insight into the timing of positive selection on specific phenotypes and successfully retrieved Y chromosome genotypes for a number of individuals, the core of this study consisted in thoroughly characterizing the changes of maternal lineages through time. This study recapitulates previous findings with the increase in frequency of Mesolithic maternal lineages throughout the Neolithic, signature of the incorporation of hunter-gatherer women in farming communities. Late Neolithic and Bronze Age France were also characterized by the appearance of new maternal and paternal lineages into the Neolithic pool. While Y chromosome haplogroups are available only for a fraction of studied individuals, and the new maternal lineages were found at low frequency in the population, both attested a degree of gene flow between ancient French and individuals deriving part of their ancestry from the Bronze Age Steppe herders.

But as we mentioned in **Part I**., parental lineages sometimes fail to recapitulate the extent of a gene flow, especially in a case where it is biased towards one sex. Indeed, the contribution of hunter-gatherers in Neolithic communities could have occurred primarily through the incorporation of women, a scenario observed in Central Africa where maternal gene flow from Pygmy hunter-gatherers to the ancestors of present-day Bantu-speaking farmers was identified as accounting for shared autochthonous maternal lineages between the two groups (Quintana-Murci *et al.*, 2008). On the other hand, the massive migration that originated from the Steppe and modified the genetic composition of European populations at the onset of the Bronze Age was described as male-driven (Goldberg *et al.*, 2017). Moreover, as we grouped together individuals according to their cultural and temporal affinity in order to gain statistical power, we could not explore much of the diversity within groups.

To make up for this, we present here the first genome-wide data on a series of individuals sampled across the French territory, covering the periods ranging from the Mesolithic to the Iron Age.

Chapter II. Results

I. Sample characterization

To investigate temporal genome-wide dynamics we sequenced 58 ancient individuals to 0.05 - 0.5 X coverage: 4 Mesolithic Hunter-Gatherers, 18 Neolithic, 15 Bronze Age and 21 Iron Age library were sequenced, resulting in 5,000 years genomic transect from prior the onset of agriculture during the Mesolithic period to the Iron Age. (**Supplementary Table 8**).

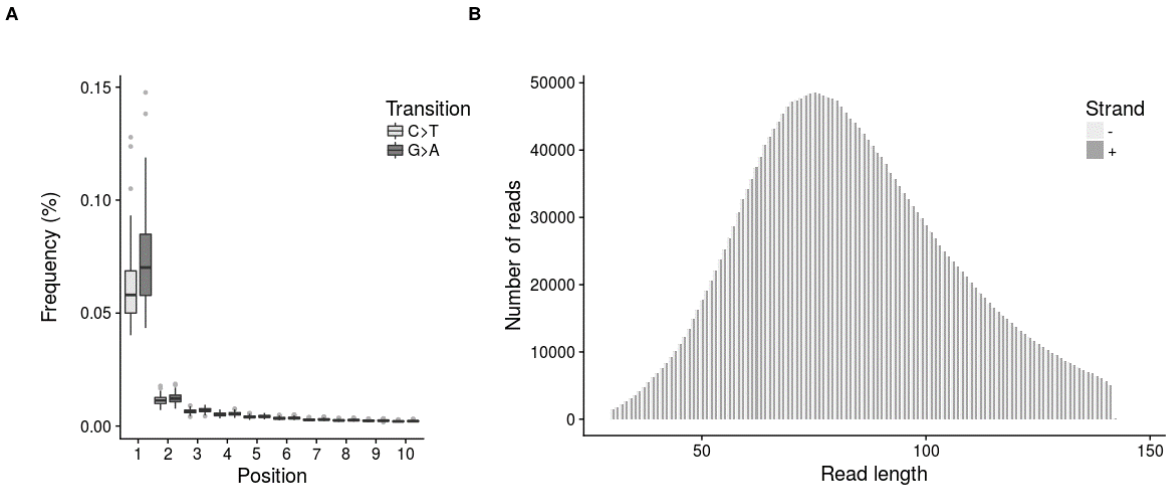


Figure 35: Authenticity of DNA across all ancient French
Pooled data for the 60 sequenced individuals. A. Distribution of the frequency of transitions associated with aDNA damages across DNA molecules. B. Size distribution of DNA fragments.

Mitochondrial and Y chromosome haplogroups obtained from our 58 individuals are consistent with those obtained through targeted capture on separately built libraries (**Supplementary Table 4 and 8**). All sequenced individuals showed an excess of both C to T and G to A transitions on the two terminal bases of each read (**Figure 35**), and the average read length across samples was around 80bp, consistent with aDNA. Further, contamination estimates derived from X chromosome polymorphisms in male individuals were consistently low (0.68%±1.26% and 3.55%±0.62% depending on the method considered), with only one individual displaying more than 5% of contamination of the X chromosome (PIR3037AB).

All analyzed individuals belong to well defined archaeological contexts from the Mesolithic period through to the Iron Age and some were directly radiocarbon dated over

the course of this study (**Supplementary Table 8**). We combined our observed genotype data with either complete genomes or capture of 1240k genome-wide polymorphisms in ancient individuals, and genotypes of modern individuals from Europe, Caucasus and the Near East genotyped on the Affymetrix Human Origins array (Fu *et al.*, 2016; Gamba *et al.*, 2014; Haak *et al.*, 2015; Lazaridis *et al.*, 2014; Lipson *et al.*, 2017; Mathieson *et al.*, 2015; Olalde *et al.*, 2018; Valdiosera *et al.*, 2018). We limited our selection of published ancient individuals to the Mesolithic to Iron Age period in Eurasia, excluding individuals that were reported as contaminated or related. We processed all published samples along with ours from raw sequencing fastq files, and chose not to co-analyse non-UDG treated along with UDG-treated individuals.

To determine how these ancient genomes relate to each other, to other ancient European genomes and to modern-day human populations, we used a combination of principal components analysis (PCA), where ancient individuals were projected onto the PCA defined by modern individuals (lsqproject option from smartpca, **Figure 36**), and various f-statistics and D-statistics.

The PCA performed over the whole ancient dataset replicates previous findings: (*i*) a clear genetic distinction between early farmers and resident hunter-gatherers, (*ii*) affinity of the former with the southwestern modern-day European variation, (*iii*) an increased affinity of prehistoric farmers to western hunter-gatherers over time due to increased admixture between the two populations and (iv) a shift towards the genetic variation of the Bronze Age herders of the Pontic Steppe around the onset of Bronze Age for populations of Western Europe, to the exception of most Iberian individuals.

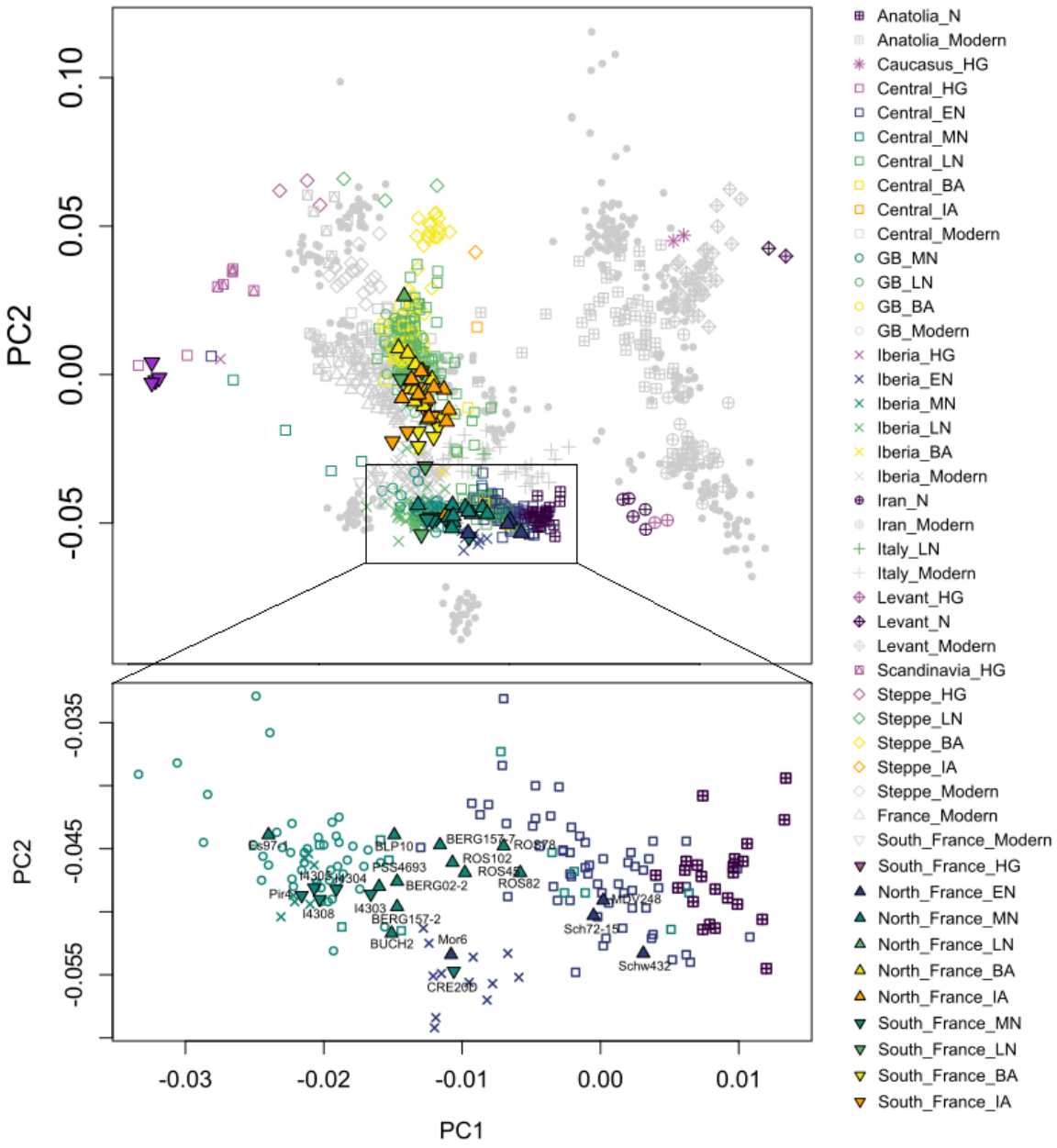


Figure 36: Principal component analysis of ancient Western Eurasians projected onto the present-day variation of genotypes

Zooming into the genetic variation among the ancient individuals from France on the PCA (upper and lower triangles with solid black contour), three main divisions are visible. These clusters are located in different regions of the plot, and don't reflect any geographically pattern between Northern and Southern France. Instead the stratification is associated with chronology and reveals clear shifts in the genomic affinities of the ancient genotypes throughout the time frame covered by our sampling. The first cluster comprises Mesolithic hunter-gatherers (**Figure 36**, magenta lower triangles), located near other Western hunter-gatherers towards the edge of the plot. The second cluster comprises Early and Middle Neolithic French (**Figure 36**, dark blue and dark green triangles) that fall within the modern-day Southern European genomic variation. The third and broader group includes most Late Neolithic individuals (**Figure 36**, light green triangles) along with both Bronze and Iron Age individuals (**Figure 36**, yellow and orange triangles) populations falling within the modern-day Central European, thus differentiating from Early Neolithic French, while closer to but remaining outside of present-day France genetic variation.

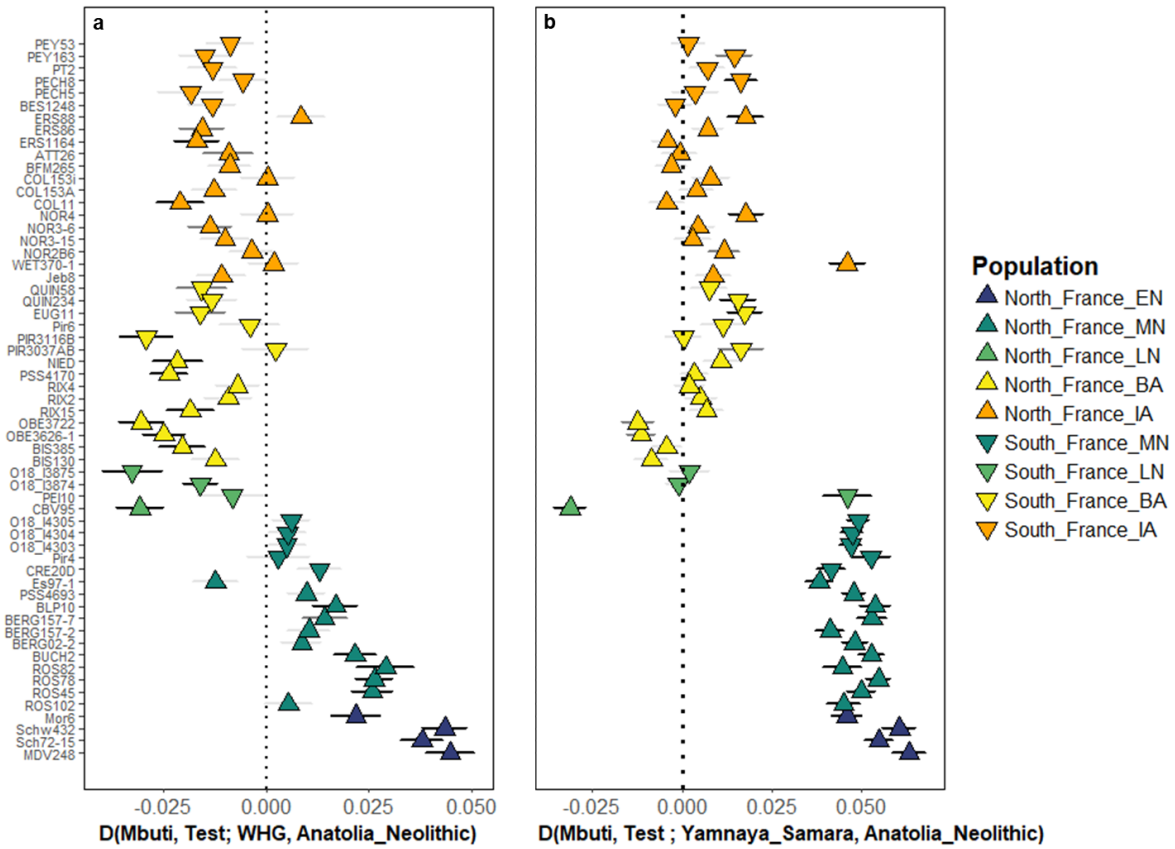


Figure 37: Genetic affinity between ancient French and either Western hunter-gatherers or Steppe herders.

a. D-statistics of the form D(Mbuti, Test: WHG, Anatolia_Neolithic). **b**. D-statistics of the form D(Mbuti, Test; Yamnaya_Samara, Anatolia_Neolithic). Error bars represent one standard error and the grey scale used reflect the value of the associated Z score: |Z|>3: black; 3>|Z|>2.58 (p=0.01): dark grey; |Z|<2.58: light grey. Values of D(Mbuti, Test: WHG, Anatolia_Neolithic) for Mesolithic hunter-gatherers from Southern France ranged from -0.1991 to -0.2075 (Z: -34.899 to -38.708), outside the range of any other ancient individual, and were not displayed on the figures.

To explain patterns of genetic shift and estimate the genetic contributions of different prehistoric groups (hunter-gatherers, early Anatolian farmers, and Steppe herders) to other ancient populations, we inferred admixture fractions using both unsupervised and supervised ADMIXTURE (Alexander *et al.* 2009). We also computed a set of f-statistics using ADMIXTOOLS (Patterson *et al.* 2012). To increase the statistical power for our low coverage individuals and perform a set of statistics, we first identified criteria to split them into groups: their respective geographical position and dates/cultures, and the values of D-statistics of the form D(Mbuti, Test; WHG, Anatolia_Neolithic) and D(Mbuti, Test, Yamnaya_Samara, Anatolia_Neolithic). The first reflects the proportion of admixture with Western hunter-gatherers (**Figure 37a**), and

the latter is sensitive to the contribution of the Yamnayas to the allelic frequencies of our individuals (**Figure 37b**). We also used outgroup f3-statistics of the form f3(Ancient French population/individual, modern/ancient Western Eurasian population; Mbuti) to ascertain the affinity of our ancient French to other ancient and modern populations.

II. Mesolithic France

Our oldest samples, corresponding to the Mesolithic, were excavated from a cave near Agris (Charentes) in Southern France. Dated from 7177 to 7057 BCE, these individuals fall towards the Western hunter-gatherer vicinity of the PCA plot, located near both Mesolithic individual Loschbour and relatively close to both the Iberian Mesolithic individual from La Braña and the Hungarian KO1 (Gamba *et al.* 2014) (**Figure 36**). This affinity was confirmed by both outgroup f3-statistics and D-statistics of the form D(Mbuti, French HG; WHG, WHG) with combinations of Loschbour (Luxembourg), La Braña (Spain) and KO1 (Hungary) as WHG, where French hunter-gatherers were found to share significantly more alleles with Loschbour than with any of the two others (**Table 16 and Supplementary Table 9** for values of this statistics on all other ancient French). With respect to modern populations, the highest values of outgroup f3-statistics for French HG were found for two Baltic populations (Lithuanian and Estonian), followed by the Basque, Icelandic and Norwegian populations (**Figure 38**).

Table 16: Values of D(Mbuti, French HG; WHG, WHG) and corresponding standard errors and Z-scores

Outgroup	French HG	WHG	WHG	D value	SE	Z score
Mbuti	PER1150C	Loschbour	KO1	-0.1043	0.011460	-9.105
Mbuti	PER1150C	La Braña	KO1	0.0001	0.011313	0.012
Mbuti	PER1150C	La Braña	Loschbour	0.0961	0.010953	8.777
Mbuti	PER3023	Loschbour	KO1	-0.0974	0.012135	-8.027
Mbuti	PER3023	La Braña	KO1	-0.0011	0.011867	-0.095
Mbuti	PER3023	La Braña	Loschbour	0.0906	0.011391	7.950
Mbuti	PER3123	Loschbour	KO1	-0.0952	0.012169	-7.827
Mbuti	PER3123	La Braña	KO1	-0.0083	0.011896	-0.698
Mbuti	PER3123	La Braña	Loschbour	0.0904	0.011307	7.998
Mbuti	PER503	Loschbour	KO1	-0.0850	0.010898	-7.801
Mbuti	PER503	La Braña	KO1	-0.0019	0.011249	-0.170
Mbuti	PER503	La Braña	Loschbour	0.0835	0.010921	7.644

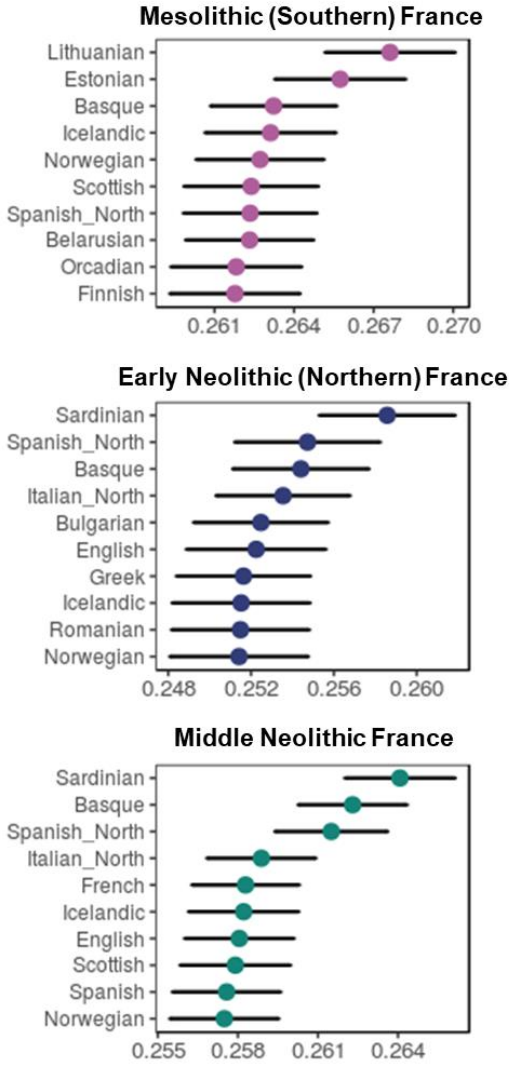


Figure 38: f3(Ancient French, Modern population; Mbuti) for Mesolithic, Early and Middle Neolithic French

Only the ten highest values of the statistics are shown for each population. See **Supplementary Figure 6** to access the results for all individuals.

III. Neolithic France

As mentioned previously, our Neolithic genomes all cluster with modern Southern Mediterranean individuals, and contemporaneous European Neolithic individuals, echoing the results of previous direct analyses of European Neolithic and post-Neolithic genomes (**Figure 36**) (Keller *et al.*, 2012; Skoglund *et al.*, 2012). It is also highlighted by outgroup f3-statistics of the form f3(Early/Neolithic France, modern population; Mbuti), as Early and Middle Neolithic, which include a diversity of Neolithic cultures such as LBK, Grossgartach and Michelsberg in the North-East, and the widespread Chasséen in the South display the highest values of this statistics with present-day Sardinians, Basque and Northern Spanish. (**Figure 38, Supplementary Figure 4**). In contrast to this apparent genomic affinity, we observed a high mtDNA diversity during this period, as described in **Part IV**, and previously observed in other parts of Europe (Brandt *et al.*, 2013; Haak *et al.*, 2010; Szécsényi-Nagy *et al.*, 2017).

A structure exists within this cluster. Early Neolithic individuals from Northern and Northeastern France are located within the genetic diversity of Central European Early farmers, associated with the LBK cultural complex (**Figure 36**, dark blue diamonds). We report the exception of Mor6, an individual from the LBK site of Morschwiller in Alsace. Located within the diversity of Iberian Early Neolithic on the PCA (**Figure 36**), it also shares more drift with Iberian Early Neolithic than with population associated with the Danubian migration wave, reflected by higher value of the statistics f3(Mbuti, Early Neolithic French individual, ancient population) (**Supplementary Figure 5**). While unearthed from an unambiguous LBK context, this individual has been sent for C^{14} dating in order to confirm its cultural assignment, since the archeological site of Morschwiller was shown to also yield Late Neolithic individuals. It was therefore not included in the "Early Neolithic (Northern) France" cluster for analyses.

Transitioning towards the Middle Neolithic, a noticeable shift towards the hunter-gatherer edge results in a separate subcluster that overlaps with contemporaneous individuals from both Great Britain and Iberia **(Figure 36**, dark green circles and crosses). This change of affinity between Middle Neolithic French and other ancient populations is also visible through f3(Middle Neolithic France, ancient populations; Mbuti), as either one of Neolithic Iberia, England and Wales can be found among the five highest values of the statistics across individuals, along with Neolithic populations from both Iberia and, to a lesser extent, Central Europe (**Figure 39**).

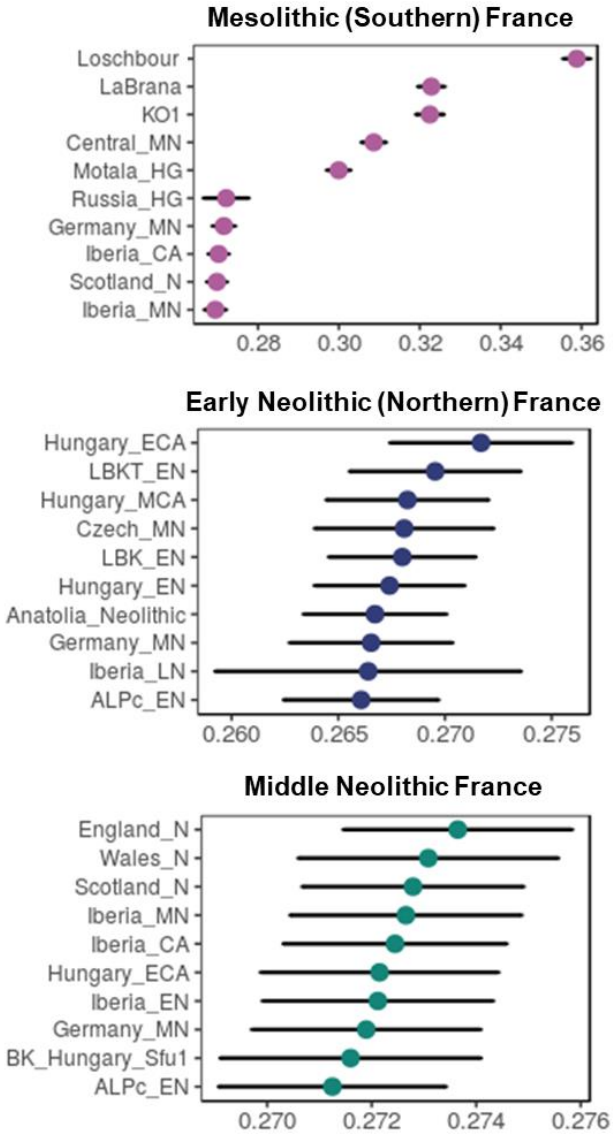


Figure 39: f3(Ancient French population, Ancient population; Mbuti) for Mesolithic, Early and Middle Neolithic French
Only the ten highest values of the statistics are shown for each population. See **Supplementary Figure 5** to access the results for all individuals.

The increase in genetic affinity with hunter-gatherers is confirmed by both unsupervised and supervised ADMIXTURE analysis, where from the Middle Neolithic onward an increase in the component associated with Western hunter-gatherer can be observed. (**Figure 40** and **Supplementary Figures 4** for the unsupervised analysis)

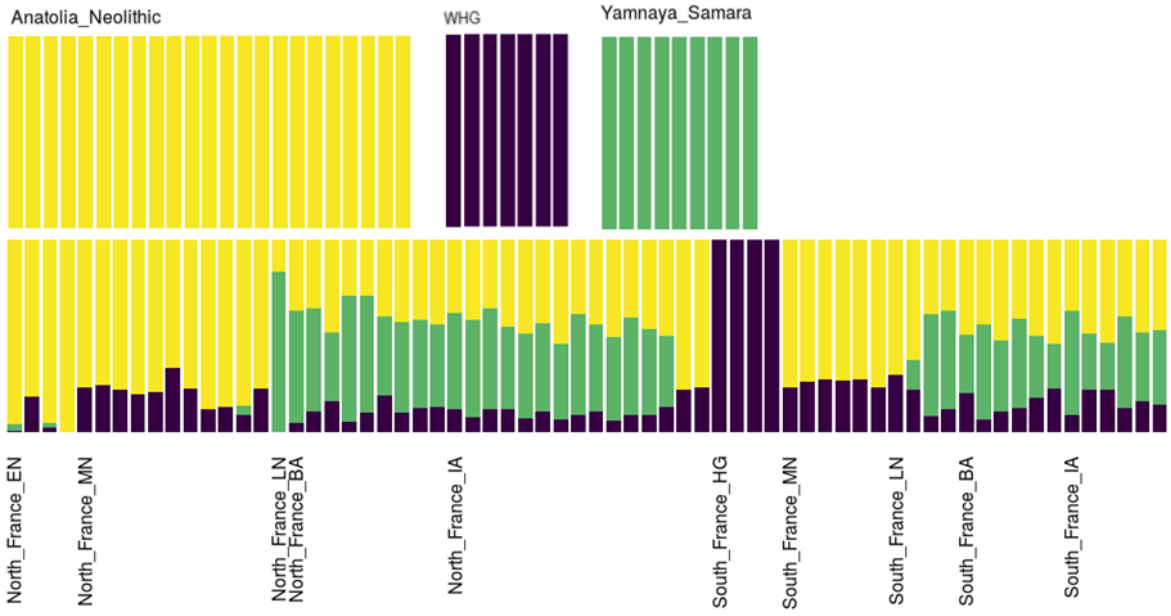


Figure 40: Supervised ADMIXTURE on ancient French where Western hunter-gatherers, Anatolian farmers and Steppe herders are used as source populations

The distribution of Middle Neolithic individuals on the PCA hints at slight differences between Northern and Southern France, possibily reflecting different affinities with other ancient populations (**Figure 36,** upper and lower dark green triangles). Individuals from Northern France seemingly share more genetic affinity with Central Europeans than their Southern France counterparts, who fit within the distribution of Iberian and British Neolithic individuals. A notable exception is Es97-1, an individual from Escalles (4200-3800 BCE, Pas-de-Calais), located at the opposite edge of the distribution of Middle Neolithic French on PC1 compared to other Northern France individuals. The case of CRE20D (4400-4200 BCE, Hérault) is also interesting, as it appears among individuals associated with the Early Neolithic Cardial culture (**Figure 36**, dark blue crosses), despite being possibly contemporaneous and geographically close to Pir4 (4450-4173 cal BCE, Hérault).

Since the Neolithic farmers who migrated along the two different routes displayed slightly different gene pools, we investigated the origin of the Neolithic component in our individuals in order to determine from which place of Europe it originated. To assess the affinity of our Neolithic, Bronze and Iron Age groups to different Early Neolithic populations, we computed D-statistics of the form D(Mbuti, Ancient French individual; Iberia_EN, LBK_EN) where Iberia_EN and LBK_EN are Early Neolithic groups from Iberia and Germany respectively. For most groups, values that are different from zero are not supported, as they don't pass the significance threshold Z > 3. This is a known bias of this statistics when the relationship between population is close, as it is then difficult to reject the null hypothesis (i.e. the tree topology is correct) for low coverage ancient genomes. For a few particular cases, affinity was significant for either one of the two early Neolithic population. Middle Neolithic individuals from Southern France were drawn towards Iberia Early Neolithic (Z=-3.01, -3.05 and -3.35 for CRE20D, I4305 and I4303 respectively), indicating that Middle Neolithic populations from Southern France share more alleles with Early migrants from the Mediterranean current. French individuals associated with the LBK culture showed positive values of this statistics that were not supported by a high enough Z score, with values only comprised between 1.57 (Schw432) and 2.61 (Sch72-15). Regrouping these individuals together in order to increase coverage did not increase this value (D=0.0078, Z=1.782). We obtained similar results when performing the same statistics replacing LBK_EN with other Danubian Early Neolithic populations such as Hungary_EN, LBKT_EN (**Supplementary Table 10**). This statistic however does not account for more complex scenarios were one of these populations might share alleles with another group, such as hunter-gatherers. To test this potential correlation, we used Spearman's rank correlation test, a measure of the relationship between variables when it can be described as a non-linear but monotonous function.

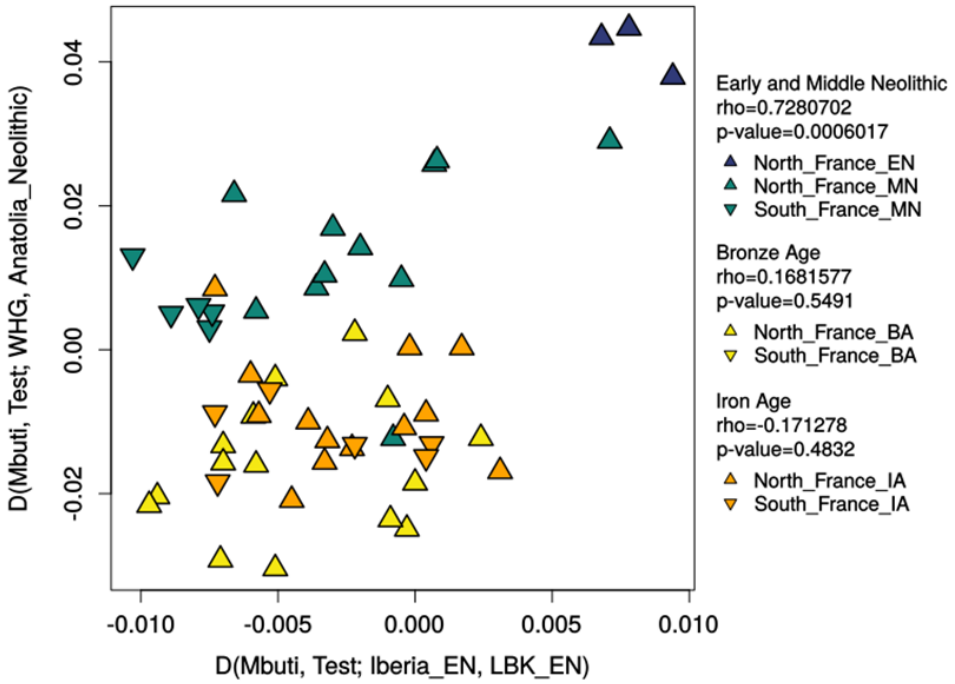


Figure 41: Result from Spearman's rank correlation test.

Correlation was tested between the values of D-statistics of the form D(Mbuti, Test; WHG, Anatolia_Neolithic) and D(Mbuti, Test: Iberia_EN, LBK_EN) on individuals from Early and Middle Neolithic (dark blue and dark green triangles), Late Neolithic and Bronze Age (light green and yellow triangles) and Iron Age individuals (orange triangles). Rho values and p-values are indicated on the figure.

Values of this statistics on the combination of D(Mbuti, Ancient French individual; Iberia_EN, LBK_EN) and D(Mbuti, Ancient French individual, WHG, Anatolian_Neolithic) reveal a correlation between the number of shared alleles with Iberia_EN and WHG during Early and Middle Neolithic (rho=0.728, p-value=0.0006), indicating that the signal conveyed by D(Mbuti, Ancient French individual; Iberia_EN, LBK_EN) likely reflects the shared proportion of hunter-gatherer ancestry between Middle Neolithic individuals from Southern France and Early Neolithic Iberians. This pattern disappears over Bronze and Iron Age. (**Figure 41**).

IV. Late Neolithic and the onset of Bronze Age

The apparent genomic stasis of the Neolithic is interrupted during the IIIrd millennium BCE coinciding with the Late Neolithic and the onset of the Bronze Age in Central Europe. Interestingly, our two Late Neolithic samples, PEI10 and CBV95, are found at different positions on the PCA plot. Where PEI10, a Bell Beaker individual from Southern France (**Figure 36**, light green lower triangle), remains close to earlier Neolithic

individuals, Northern France CBV95 and other reported French Bell Beakers (Olalde *et al.* 2018) falls among other Late Neolithic and Bronze Age individuals from Central Europe, shifting towards modern Central European genotypes (**Figure 36**, light green upper triangle). ADMIXTURE analysis reveals that our Late Neolithic individual from Northern France (CBV95) harbors the highest proportion of the Yamnaya associated component, while the latter is variable in Late Neolithic Southern France (**Figure 40**). Notably, PEI10 retains a similar profile to Middle Neolithic individuals from Southern France, as was hinted by its position on the PCA (**Figure 36**) Another individual from Haute-Savoie (I1388, Olalde *et al.* 2018) also remains close to the Neolithic cluster while both I3874 and I3875, also from Southern France, are found within the diversity of other Late Neolithic and Bronze Age individuals. This difference in their affinity is also highlighted by f3(French Bell Beaker, ancient population; Mbuti) as shown in **Figure 43** where individuals are labeled after their department of origin: CBV as Bell Beaker Aisne (Hauts-de-France, Northern France), PEI 10 as Bell Beaker Aude (Occitanie, Southern France) and the combination of I3874 and I3875 as Bell Beaker ADHP ("Alpes de Haute-Provence", Provence-Alpes-Côte d'Azur, Southern France). I1338 was excluded from the different statistics after filtering for genotyping rate. Nevertheless, unsupervised ADMIXTURE profiles can be found for this individual in **Supplementary Figure 4** (Beaker_Savoie).

To better appreciate the affinity of each of the Bell Beaker individuals to various Steppe populations, we computed D-statistics of the form D(Mbuti, Test, Middle Neolithic Population, Bell Beaker individual) with Middle Neolithic populations form the same region. For both Northern and Southern France, the Steppe population shared more alleles with Bell Beakers than with Middle Neolithic individuals, suggesting that the arrival of the Bell beaker Complex in Southern France was mediated by migrants with steppe genetic affinities. Nevertheless, steppe populations share more alleles with I3874, I3875 than with PEI10, the latter displaying no significant genetic affinity to any of the Steppe populations, suggesting that Southern France Bell Beaker did not form a homogenous group (**Figure 42**).

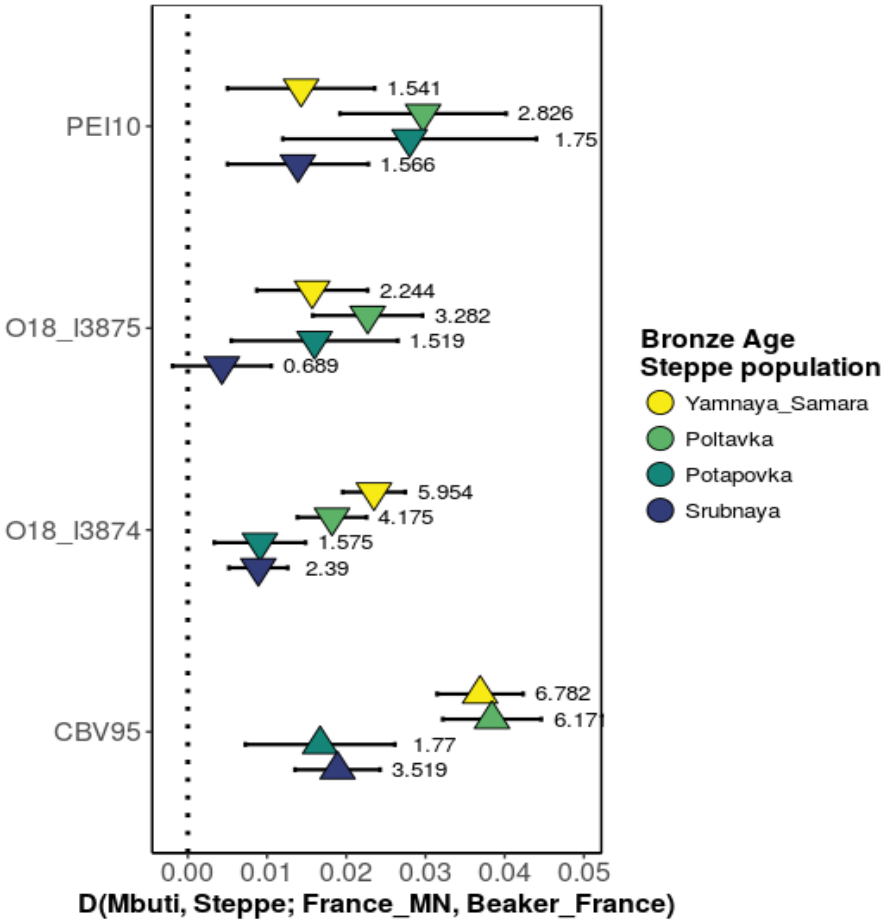


Figure 42: Shared alleles between populations from the Steppe and French Bell Beakers. D-statistics of the form D(Mbuti, Steppe population; France_MN, Beaker_France) where France_MN corresponds to either Southern or Norther France MN depending on the geographical origin of the Beaker individual tested. Steppe populations are Bronze Age Yamnaya_Samara, Poltavka, Potapovka and Srubnaya. Z score associated with the value of the statistics is displayed on the figure.

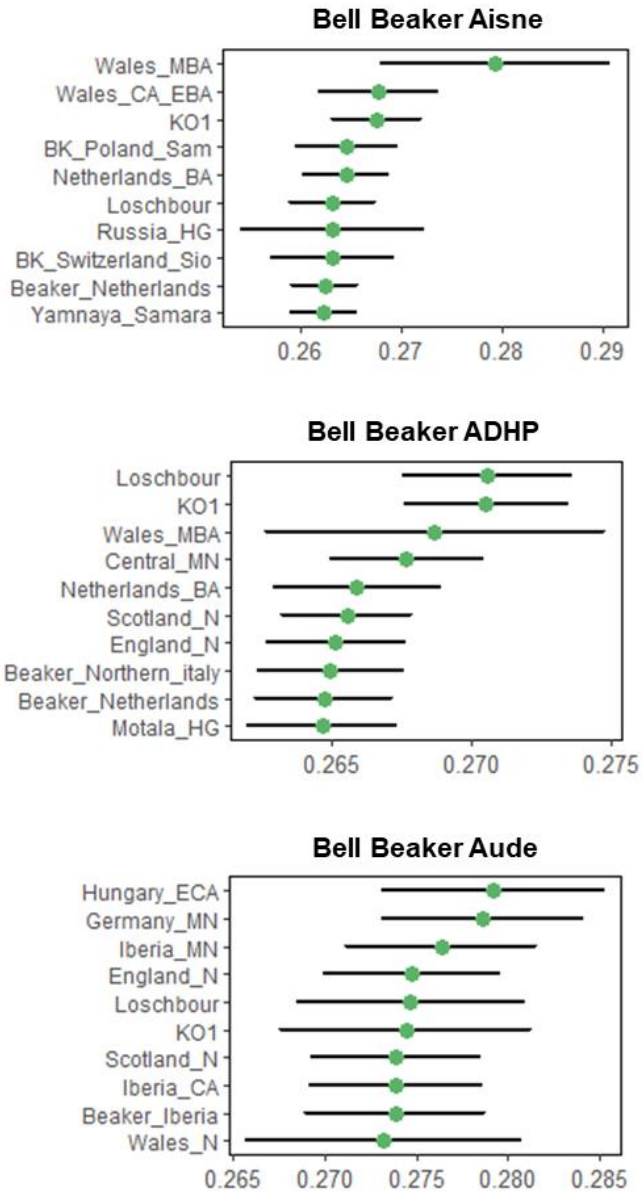


Figure 43:f3(Bell Beaker, Ancient population; Mbuti) performed on Bell Beaker from Northern and Southern France

The different labels correspond to the following individuals: CBV95 (Bell Beaker Aisne, Northern France), I3874 and I3875 (Alpes de Haute-Provence, Southern France), PEI10 (Bell Beaker Aude, Southern France). Only the ten highest values of the statistics are shown for each population.

V. Bronze and Iron Age France

Bronze and Iron Age France cannot be distinguished from the PCA, as both are shifted towards modern Central Europe, falling within Bronze Age Britain and Central Europe genetic diversity (**Figure 36**). Both Bronze and Iron Age populations from Northern and Southern France alike share the most drift with present-day Lithuanians, Scottish and Basque, as well as Icelandic and Northern Spanish (**Figure 44**). With respect to ancient populations, Bronze and Iron Age France share the most drift with Western hunter-gatherers such as Loschbour and KO1, as well as an Early Bronze Age population from Wales (**Figure 45**).

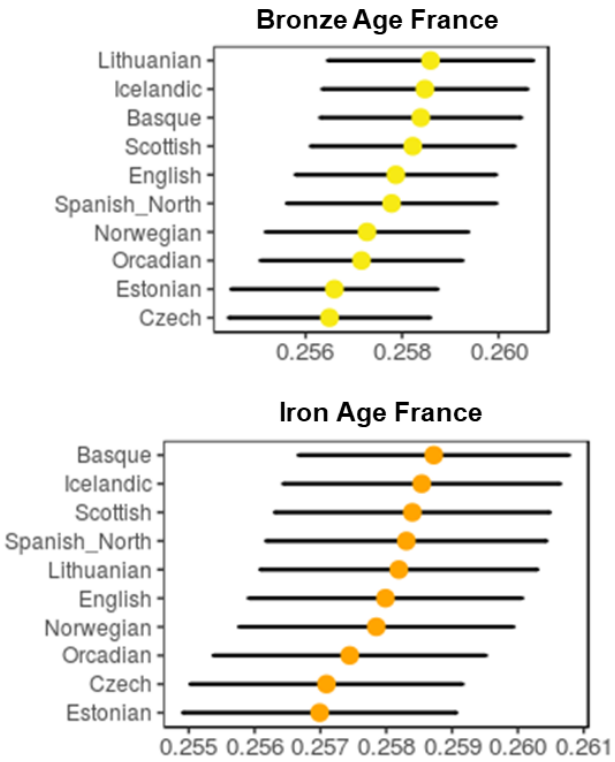


Figure 44: f3(Ancient French, Modern population; Mbuti) for Bronze Age and Iron Age French
Only the ten highest values of the statistics are shown for each population. See **Supplementary Figure 6** to access the results for all individuals.

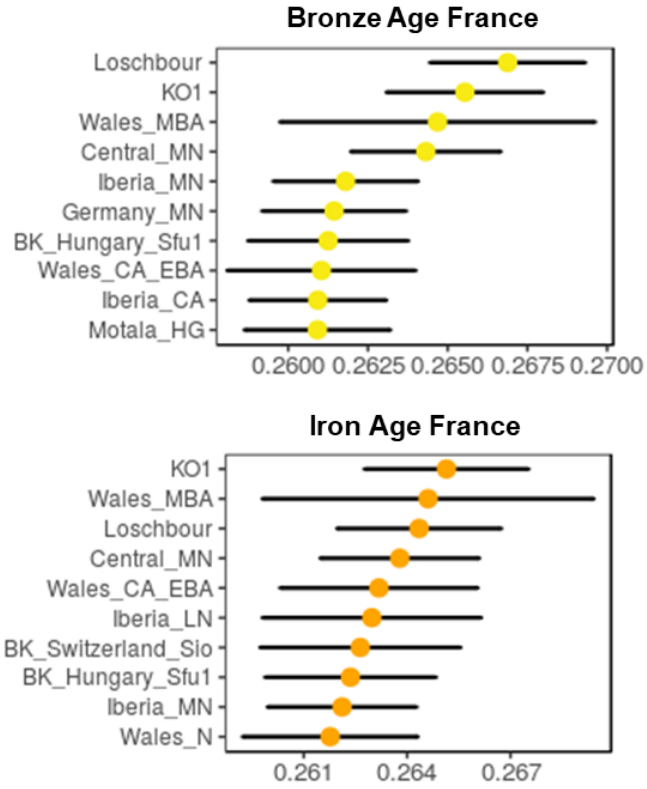


Figure 45: f3(Ancient French, Ancient population; Mbuti) for Bronze Age and Iron Age French

Only the ten highest values of the statistics are shown for each population. See **Supplementary Figure 5** to access the results for all individuals.

Chapter III. Discussion

I. Structure among Western hunter-gatherers

The affinity of Mesolithic individuals from Southwestern France with population from Northeastern Europe echoes previous findings (Skoglund *et al.*, 2012). Since populations from Scandinavia and around the Baltic Sea were shown to have retained a higher proportion of their ancestry from Western hunter-gatherers, these results were expected (Skoglund *et al.*, 2012). More surprising is the shared affinity with the individual from Loschbour (Luxembourg) at the expense of both La Braña and KO1. This reveals some degree of structure among Western hunter-gatherers, where Loschbour appears as the best proxy for French hunter-gatherers.

II. Neolithic continuity and admixture with hunter-gatherers

The genetic variation of Neolithic France is encompassed within that of other contemporaneous European populations. While we lack data for the earliest part of the Neolithic in Southern France, the affinity of one Middle Neolithic individual (CRE20D) with Iberian Early and Middle Neolithic individuals supports a Cardial origin for the Neolithic population of this region (**Figure 36**, **Supplementary Figure 5**). Three of our individuals associated with the LBK culture that stems from the Danubian migration wave are directly found among other Central European Early Neolithic individuals. As the Neolithic progresses, the genetic contribution of Western hunter-gatherers increases, resulting in a progressive shift of genotypes away from Early Neolithic individuals on the PCA plot (**Figure 36**). A first cluster of individuals forms the first part of Middle Neolithic in Eastern France, here represented by individuals from the Grossgartach culture, retaining ties with Central European contemporaneous individuals. This is consistent with the Grossgartach deriving from the Danubian sphere, and matches the results obtained from mitochondrial DNA. When reaching the second part of Middle Neolithic, only a slight distinction remains between individuals from Northern and Southern France, as they all fit within the diversity of Middle Neolithic Iberia and Great Britain (**Figure 36**). This affinity, confirmed by outgroup f3-statistics (**Figure 39**, **Supplementary Figure 5**), would support the idea of population continuity on both sides of the Channel by the end of the Middle Neolithic. In parallel to these observations on autosomal genotypes,

observed Neolithic Y-chromosome lineages and mtDNA haplogroups are also consistent with either a Mesolithic or Neolithic background (Lazaridis *et al.*, 2014; Olalde *et al.*, 2014), which is further support to a migration of farmers into France with substantial admixture with local hunter-gatherers as the Neolithic progresses.

III. A shift in affinities with the advent of Late Neolithic and Bronze Age

By the time the Bronze Age starts, the situation is contrasted among Late Neolithic individuals. Bell Beaker associated CBV95 in Northern France shows the highest proportion of steppe related ancestry amongst our dataset, and belongs to the Y chromosome haplogroup R1b. PEI10, a woman associated with the same cultural complex in Villedubert, near Carcassonne in South-western France, derives her ancestry only from Neolithic farmers and Western hunter-gatherers, similar to Middle Neolithic individuals from the sample region. Our results here suggest that Southern France shares affinities with Iberia, whereas the Northern part of France is more connected to Central Europe, and was ultimately reached by the massive westward migration from the Pontic Steppe that occurred at the onset of Bronze Age. This observation is consistent with the recent publication from Olalde *et al.* 2018 where they show that Bell Beaker populations from Iberia and Central Europe show little genetic affinity, as the first displays a genetic signature from the Pontic Steppe when the second remains in the continuity of the Middle Neolithic (**Figure 36**). These results suggest that the westward migration wave that originated in the Pontic Steppe reached the French territory and contributed to the increase in the frequency of the R1b lineage, that represents 58.5% of Y chromosome lineages in the present-day French population (Myres *et al.*, 2011). Nevertheless, other Bell Beaker individuals reported by Olalde and colleagues point to a more complex picture in the south of France. While individuals from Southwestern France (I3874 and I3875) display genetic affinity with a number of populations from the Steppe (**Figure 42**), I1388, another Bell Beaker individual located in Haute-Savoie, harbors a profile similar to that of PEI10 (**Supplementary Figures 4**). While the number of individuals reported here is not sufficient to draw conclusions regarding a geographic stratification of the population in this region, our hypothesis would be that the diffusion of Steppe-associated ancestry in Southern France was successful in the South-east and along the Mediterranean coast, but did not permeate inland from there and westwards into the Pyreneans foothills. More

genomes will be necessary to disentangle local and regional patterns of genetic variation during the spread of the Bell Beaker cultural complex in Southern France.

IV. No apparent changes as France transitions to the Iron Age

In contrast to what was described by Gamba *et al.* 2014, there is no further shift towards Eastern Eurasian genotypes during Iron Age as it was the case in Central Europe, with increased genetic affinities along with technological and typological affinities with Steppe cultures at this time. Individuals from Hallstatt and La Tène culture on the French territory display similar affinities towards both modern and ancient populations. This could indicate that the transition from Bronze Age to Iron Age in France was driven by cultural diffusion from Central Europe without further migration of an external population. This is consistent with the archeological and linguistic hypothesis that states that the Celts from the 2nd Iron Age descended from populations already established in Western Europe (Brun, 2008). Indeed, first described as a people from Germany (Hubert, 1950), the Celts from the second Iron Age expanded towards the south and the east of Europe. Yet, no migration was detected in the archeological records during the Iron Age to explain their presence in the British Isles. This discrepancy led the archeologists to posit that the population from which originated the Celtic culture and language was established prior to the Iron Age (Hawkes and Hawkes, 1973). In particular, J.P. Demoule observed that a sense of symbolic and cultural continuity was maintained throughout the Bronze Age and the Iron Age, within boundaries that correspond to that of the Bell Beaker cultural complex of the IIIrd millennium BCE (Demoule, 1998, 2014). Since by the Bronze Age, populations of Western and Central Europe, to the exclusion of Iberia, appear to be genetically homogenous, it is possible that further migrations between different parts of Europe could go unnoticed. Iron Age genomes from other parts of Europe, such as Central Europe and Great Britain would be crucial to appreciate the extent of migration versus cultural diffusion in the spread of Iron Age across Europe.

Interestingly the genetic diversity of Iron Age and present-day France, albeit close, remain mutually exclusive, indicating that one or several events of later demographic events are still missing from the picture.

Chapter IV. Conclusion

We present the first comprehensive genomic data for a substantial dataset spanning over five millennia of the pre- and protohistory on the territory of present-day France. The preliminary results obtained from 58 individuals reveal that the period ranging from the Mesolithic to the Iron Age was shaped by major events that deeply altered the composition of the population across the territory. The difference observed between Early and Middle Neolithic individuals in Northern and Southern France is consistent with two independent migrations waves, originating from slightly different populations. This echoes the results we obtained through uniparentally-inherited markers. Upon their arrival, these individuals started mixing with local hunter-gatherers, who left an increasing signature in farmers' genomes as the Neolithic unfolded, blurring the geographical stratification that existed until then. Steppe-related ancestry, named after a Bronze Age population of herders living in the Pontic Steppe, is detected in individuals associated with the Bell Beaker cultural complex and late Bronze and Iron Age individuals both Northern and Southern France, mirroring the situation observed in Central, Northern Europe and the British Isles. This suggests that populations harboring a genetic component descending from the Steppe herders migrated into France at the onset of Bronze Age, even though their contribution was not homogeneous across the territory by the Late Neolithic. Surprisingly, Bronze and Iron Age French cannot be distinguished using their genotypes, suggesting that the spread of the Iron Age culture across the territory might not have been driven by migration of non-European populations. Europe-wide genomic information is still scarce for this period, thus limiting potential comparisons. While bringing an unprecedented resolution to the study of prehistoric French populations, our results emphasize the necessity of more in-depth diachronic studies across Europe to better integrate newly reported genomes and ultimately unravel all the complexity of the peopling of the European continent over thousands of years.

Part VI. Conclusion and perspectives

Ancient DNA research has transformed our understanding of human history, while revealing its complexity. It led us to look for evidence from aDNA to interpret present-day genetic variation, rather than focusing on the latter to explain our past. If paleogenomics successfully delved into the distant past, the record being held by an ancient horse preserved in permafrost and dated approximately 560,000 to 780,000 years (Orlando *et al.*, 2013), very ancient samples and some geographical regions of great interest remain challenging. Moreover, the minute amount and degradation state in which DNA subsists make it a methodological challenge to work with. Technical improvements, both in the recovery of DNA through extraction and in library construction can circumvent these limitations to some extent, by allowing the recovery of the very short DNA fragments that characterize the least preserved samples. But as the length of DNA fragments decreases, so does their potential informativity, as short molecules are less likely to align to unique locations across the genome, and are thus discarded from post-processing. In but a few years, aDNA studies have moved from the amplification and sequencing of a handful of targeted regions through PCR, to the generation of gigabases of DNA sequences with the advent of high throughput sequencing. Not only increasing the amount of data available, it also dramatically improved the quality and reliability of the generated data. Indeed, high throughput sequencing allows both to track, through individual barcoding, crossed contamination between samples after DNA library construction, but also enables to assess the authenticity of the sequences through their characteristic damage patterns, where PCR failed to pass both criteria. Automatization of various pre-processing steps, such as library construction or capture, also contributed to reduce the human handling of aDNA extract, and to some extent the risk of contamination.

With this project, we set out to provide a comprehensive overview of the peopling of present-day France through the end of its pre- and protohistory, covering major social and cultural transformations such as the Neolithization and the Metal Ages. Our goal was to fill a gap, both in terms of geography and genomic data availability, in the study of the peopling of the European continent. Our results integrate in a broader frame initiated with the extensive study of the same periods in other European countries such as Germany, Hungary, Great Britain or Spain (Brandt *et al.*, 2013; Gamba *et al.*, 2014; Lipson *et al.*,

2017; Olalde *et al.*, 2018; Valdiosera *et al.*, 2018). In that respect, we reached our initial objective with the combined study of about 246 individuals across the French territory, that formed a diverse dataset comprising 193 complete mitochondrial genomes, genotyping over 70 nuclear markers and Y chromosome diagnostic SNPs, and a deeper characterization of 58 individuals through low coverage genomes. This thesis summarized the analyses conducted on these individuals. These results are part of a publication to be submitted before the end of the year.

The success rate of the project relies upon an unprecedented sampling and an overall good DNA preservation in the remains included in the studied dataset. Built upon preliminary studies in the laboratory, an ANR[1] funded project coordinated by Melanie Pruvost started in 2015, establishing a wide network of archeologists and anthropologists and dedicating funding to the present study. Complementing already established collaborations, this new network provided a substantial sampling, with bone samples for periods ranging from the Mesolithic to the Middle Ages. The discovery of the exceptional preservation of DNA in the petrosal part of the temporal bone (Gamba *et al.*, 2014) prior to the onset of the project led us to reorient the sampling towards this particular part of the skeleton, which represent 79 % of the total of processed samples. With an average endogenous content of 38 % +/- 20% overall in screened individuals, this was a major asset in the achievement of this project.

We also benefited from an optimized enrichment method where biotinylated RNA baits were transcribed from either the complete mitochondrial or genomic target regions that were amplified by PCR beforehand, which provides the economy of multiple oligonucleotide synthesis for each bait. While the mitochondrial capture was adapted from an already established protocol within the laboratory (Massilani *et al.*, 2016), the procedure for nuclear targets was performed with a two-step PCR, using a pair of primer per genomic target, and a universal pair of primers to incorporate the T3 promoter upstream of either the forward or the reverse strand, thus allowing the synthesis of baits complementary to either DNA strands. This strategy reduced by half the costs of the oligonucleotides during the capture set up and minimizes the risk of the formation of

[1] Agence Nationale de la Recherche

primer dimer artefacts that is increased with long 5'tags. Altogether, this strategy represented a valuable economic alternative to sequence capture using commercial microarray-derived oligonucleotides for a medium size number of markers captured. Capturing nuclear SNPs alongside the mitochondrial genome revealed that the baits were efficient at recovering their target region, but the initial complexity of the ancient DNA library is a limiting factor. This motivated the choice of only best-preserved individuals among the 246 initially screened to perform the enrichment.

We successfully retrieved complete mitochondrial genomes for 193 individuals (< 5% missing sites). Nuclear data were retrieved from 149 individuals, allowing us to characterize allelic frequencies for 70 nuclear markers associated with a wide range of phenotypes displaying signatures of positive selection in present-day Europeans. The degree of preservation of endogenous DNA in our samples opened unanticipated options, putting more in-depth population genomic studies within our reach. We seized this opportunity and generated genomes for 58 individuals in our panel encompassing three main cultural transitions: the transition from the Mesolithic to the Neolithic, that of Neolithic to Bronze Age and finally to the Iron Age. These genomes, with a coverage ranging between 0.05X and 0.5X, were compared with both modern genotyping data and contemporaneous genomic data available for these periods, as well as for their relationship with modern French.

With respect to the Neolithic period, the results we obtained from both mitochondrial DNA and low coverage genomes were consistent with published results across Europe, while complementing previous findings for Neolithic France (Beau *et al.*, 2017; Lacan *et al.*, 2011; Rivollat *et al.*, 2015, 2016). We showed that the Neolithic people who inhabited the territory of present-day France descended from early farmers in Anatolia, therefore forming a group distinct from the autochthonous hunter-gatherers who lived there. As the Neolithic unfolded across the territory and gave rise to various different cultures, both datasets confirmed the introgression of hunter-gatherer DNA into the genomes of Neolithic farmers, supporting their progressive integration within farming communities. Although widespread, the process of admixture with hunter-gatherers shows regional variability, and maternal lineages suggest that it was preponderant in North-western France. By the Middle Neolithic in France, the population had become homogenous as we

could no longer trace back its ancestry to one of the two early Neolithic migration waves, associated with the LBK and the Cardial culture. This contrasts with the archeology of the period, that describes a regionalization of the Neolithic where various cultures emerged from the breakdown of the widespread LBK and Cardial cultures. While Y chromosome haplogroups are less diverse, the paternal lineages we report for the Neolithic period are consistent with previous findings (Gamba *et al.*, 2014; Haak *et al.*, 2015; Lazaridis *et al.*, 2014; Mathieson *et al.*, 2015).

The study of a selection of nuclear markers revealed that Neolithic people were very different from present-day Europeans. Derived allele frequencies reveal selection on loci involved in pigmentation, diet and immunity that are consistent with adaptation to high latitudes and a new diet. For a number of loci, especially innate immunity related variants and pigmentation, the allelic frequencies in the Neolithic population are comparable to those observed in modern European populations, indicating that selection on these loci predates the Neolithic. Finally, neither Mesolithic HGs nor Neolithic farmers on the French territory carried the allele allowing for lactase persistence during adulthood, in agreement with studies of Neolithic populations further East that reported the presence of this allele only by the end of the Bronze Age (Allentoft *et al.*, 2015; Haak *et al.*, 2015).

Moving on towards the Bronze Age, a signature of the Yamnaya-associated ancestry is visible in most individuals belonging to the Late Neolithic Bell Beaker culture, confirming some extent of gene flow between ancient French and individuals who derived part of their ancestry from the Steppe around 2500 BCE. When looking at uniparentally-inherited markers alone, the situation is more contrasted: the Yamnaya signature is detectable during the Late Neolithic and the Bronze Age as Y chromosome haplogroup R1b appears and becomes exclusive in our dataset, while the new maternal lineages introduced during this period (I, U5, U2) remain at a low frequency. This discrepancy between both markers would argue in favor of a male-driven geneflow, but has yet to be properly investigated through genome-wide data as it was done for Central Europe ((Goldberg *et al*., 2017; Lazaridis and Reich, 2017). A striking feature of our results is the impossibility to disentangle Bronze and Iron Age French. While this could mean that the Iron Age culture developed in the French Bronze Age population substrate through the diffusion of techniques, this could also result from the apparent homogeneity of Late Neolithic and

Bronze Age Western Europeans, to the exclusion of Iberians. In such context, tracing migrations from one part of Europe to the other becomes more challenging and will require a deeper temporal and geographical resolution.

As this study relied on a sampling from many different sites and regions, it revealed the heterogeneity that exists between contemporaneous sites sometimes assigned to the same cultural complex. The site-specific variability in the composition and frequency of the various haplotype results highlight the potential heterogeneity across archaeological sites that belong to the same culture, but where communities of various sizes possibly had different demographic histories or embraced different matrimonial traditions. This point is crucial in ancient DNA studies, where a handful of archeological sites are often used to infer population-wide histories. Another often overlooked factor is the level of intra-site genetic variation. Here, we chose to generate the partial genomic sequence of individuals sampled from various locations and periods in order to cover both a wide temporal transect and geographical area. While this dataset offers an unprecedented glimpse into the composition of the ancient French population, we acknowledge that it does not allow for a direct comparison between populations from different locations and periods. Indeed, the pseudo-haploid low coverage genome of a single individual provides little information regarding the genetic diversity of the group he was drawn from. A thorough study of the genetic makeup of whole ancient communities would therefore complement the results and hypotheses presented in this thesis.

Characterization of the French population will continue beyond the periods covered in the present study in the context of the ANR project I mentioned above and within the laboratory of the Institut Jacques Monod in Paris. While partial genomes were sufficient to disentangle population history over several millennia in a cost-efficient way, the limited coverage also represents a drawback. Indeed, the proportion of overlapping loci across the genome is reduced when comparing pairs of low coverage genomes. This hampered our ability to compare, within geographical regions, later periods with the preceding genetic subsrate. To integrate the evolution of the human population on the territory of France within the global processes that shaped the population of Europe, it would be relevant to generate high-resolution genomes from key periods and regions, as our experiments revealed an excellent DNA preservation in many of our samples. This

would prepare the integration of our samples with the high coverage genomes that are about to be generated throughout Europe.

Bibliography

Achard-Corompt, G., Achard-Corompt, N., Desbrosse, V., Auxiette, G., Fechner, K., Moreau, C., Peltier, V., Riquier, V., and Vanmoerkerke, J. (2010). Chasse, culte ou artisanat? Premiers résultats du projet de recherche relatif aux fosses à profil « en V, Y, W ». Bull. Société Préhistorique Fr. *107*, 588–591.

Alberti, F., Gonzalez, J., Paijmans, J.L.A., Basler, N., Preick, M., Henneberger, K., Trinks, A., Rabeder, G., Conard, N.J., Münzel, S.C., et al. (2018). Optimized DNA sampling of ancient bones using Computed Tomography scans. Mol. Ecol. Resour. *18*, 1196–1208.

Allentoft, M.E., Collins, M., Harker, D., Haile, J., Oskam, C.L., Hale, M.L., Campos, P.F., Samaniego, J.A., Gilbert, M.T.P., Willerslev, E., et al. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc. R. Soc. B Biol. Sci. *279*, 4724.

Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Ammerman, A.J., and Cavalli-Sforza, L.L. (1984). The Neolithic Transition and the Genetics of Populations in Europe (Princeton University Press).

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. *23*, 147.

Armbruster, B., and Pernot, M. (2006). La technique du tournage utilisée à l'Âge du Bronze final pour la fabrication d'épingles de bronze trouvées en Bourgogne. Bull. Société Préhistorique Fr. *103*, 305–311.

Arruda, S.R., Pereira, D.G., Silva-Castro, M.M., Brito, M.G., and Waldschmidt, A.M. (2017). An optimized protocol for DNA extraction in plants with a high content of secondary metabolites, based on leaves of Mimosa tenuiflora (Willd.) Poir. (Leguminosae). Genet. Mol. Res. GMR *16*.

Audouze, F., and Buchsenschutz, O. (1989). Villes, villages et campagnes de l'Europe celtique : du début du IIe millénaire à la fin du Ier siècle avant J.-C. (Paris : Hachette).

Ávila-Arcos, M.C., Cappellini, E., Romero-Navarro, J.A., Wales, N., Moreno-Mayar, J.V., Rasmussen, M., Fordyce, S.L., Montiel, R., Vielle-Calzada, J.-P., Willerslev, E., et al. (2011).

Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. Sci. Rep. *1*.

Baay, M., Brouwer, A., Pauwels, P., Peeters, M., and Lardon, F. (2011). Tumor Cells and Tumor-Associated Macrophages: Secreted Proteins as Potential Targets for Therapy. Clin. Dev. Immunol. *2011*.

Barbujani, G., and Chikhi, L. (2006). DNAs from the European Neolithic. Heredity *97*, 84–85.

Barbujani, G., Bertorelle, G., Capitani, G., and Scozzari, R. (1995). Geographical structuring in the mtDNA of Italian. Proc. Natl. Acad. Sci. U. S. A. *92*, 9171–9175.

Barbujani, G., Bertorelle, G., and Chikhi, L. (1998). Evidence for Paleolithic and Neolithic gene flow in Europe. Am. J. Hum. Genet. *62*, 488–492.

Barreiro, L.B., and Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat. Rev. Genet. *11*, 17–30.

Barreiro, L.B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J.K., Bouchier, C., Tichit, M., Neyrolles, O., Gicquel, B., et al. (2009). Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. PLoS Genet. *5*, e1000562.

Batini, C., Hallast, P., Zadik, D., Delser, P.M., Benazzo, A., Ghirotto, S., Arroyo-Pardo, E., Cavalleri, G.L., Knijff, P. de, Dupuy, B.M., et al. (2015). Large-scale recent expansion of European patrilineages shown by population resequencing. Nat. Commun. *6*, 7152.

Battaglia, V., Fornarino, S., Al-Zahery, N., Olivieri, A., Pala, M., Myres, N.M., King, R.J., Rootsi, S., Marjanovic, D., Primorac, D., et al. (2009). Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe. Eur. J. Hum. Genet. *17*, 820–830.

Beau, A., Rivollat, M., Réveillas, H., Pemonge, M.-H., Mendisco, F., Thomas, Y., Lefranc, P., and Deguilloux, M.-F. (2017). Multi-scale ancient DNA analyses confirm the western origin of Michelsberg farmers and document probable practices of human sacrifice. PLoS ONE *12*.

Beleza, S., Johnson, N.A., Candille, S.I., Absher, D.M., Coram, M.A., Lopes, J., Campos, J., Araújo, I.I., Anderson, T.M., Vilhjálmsson, B.J., et al. (2013). Genetic architecture of skin and eye color in an African-European admixed population. PLoS Genet. *9*, e1003372.

Bell, L.S., Skinner, M.F., and Jones, S.J. (1996). The speed of post mortem change to the human skeleton and its taphonomic significance. Forensic Sci. Int. *82*, 129–140.

Bellwood, P., Gamble, C., Le Blanc, S.A., Pluciennik, M., Richards, M., and Terrell, J.E. (2007). First Farmers: the Origins of Agricultural Societies, by Peter Bellwood. Malden

(MA): Blackwell, 2005; ISBN 0-631-20565-9 hardback £60; ISBN 0-631-20566-7 paperback £17.99, xix+360 pp., 59 figs., 3 tables. Camb. Archaeol. J. *17*, 87.

Benazzi, S., Douka, K., Fornai, C., Bauer, C.C., Kullmer, O., Svoboda, J., Pap, I., Mallegni, F., Bayle, P., Coquerelle, M., et al. (2011). Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. Nature *479*, 525–528.

Bennett, E.A., Massilani, D., Lizzo, G., Daligault, J., Geigl, E.-M., and Grange, T. (2014). Library construction for ancient genomics: single strand or double strand? BioTechniques *56*, 289–290, 292–296, 298, passim.

Beutler, E. (1994). G6PD deficiency. Blood *84*, 3613–3636.

Bocquet-Appel, J.-P., Naji, S., Linden, M.V., and Kozlowski, J.K. (2009). Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14C dates. J. Archaeol. Sci. *36*, 807–820.

Bogard, M., Lamoril, J., and Ameziane, N. (2005). Principes de biologie moléculaire en biologie clinique.

Boyd, B. (2006). On 'sedentism' in the Later Epipalaeolithic (Natufian) Levant. World Archaeol. *38*, 164–178.

Brace, S., Diekmann, Y., Booth, T.J., Faltyskova, Z., Rohland, N., Mallick, S., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., et al. (2018). Population Replacement in Early Neolithic Britain. BioRxiv 267443.

Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.-J., et al. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science *326*, 137–140.

Brandt, G., Haak, W., Adler, C.J., Roth, C., Szécsényi-Nagy, A., Karimnia, S., Möller-Rieker, S., Meller, H., Ganslmeier, R., Friederich, S., et al. (2013). Ancient DNA reveals key stages in the formation of Central European mitochondrial genetic diversity. Science *342*, 257.

Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., et al. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. *104*, 14616–14621.

Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajković, D., Kućan, Ž., et al. (2009). Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. Science *325*, 318–321.

Brun, P. (1987). Princes et princesses de la celtique: le premier âge du fer en Europe, 850-450 av. J.-C. (Errance).

Brun, P. (2008). L'origine des Celtes. Communautés linguistiques et réseaux sociaux.

Brun, P., and Ruby, P. (2008). L'âge du fer en France: premières villes, premiers états celtiques (Découverte).

Buffon, G.L.L. comte de, Daubenton (Louis-Jean-Marie), M., Montbéliard, P.G. de, and Illon), L.C. (M le comte de, Bernard Germain Etienne de La Ville sur (1767). Histoire naturelle, generale, et particuliere, avec la description du Cabinet du roy.. (Imprimerie royale).

Cann, R.L., Stoneking, M., and Wilson, A.C. (1987). Mitochondrial DNA and human evolution. Nature *325*, 31–36.

Cardoso, J.L. (2014). Cronología absoluta del fenómeno campaniforme al Norte del estuario del Tajo: implicaciones demográficas y sociales. Trab. Prehist. *71*, 56–75.

Carozza, L., and Ambert, P. Origine et développement de la première métallurgie française: état de la question. 15.

Carozza, L., Marcigny, C., Burens-Carozza, A., and Ghesquière, E. (2007). Les travaux et les jours : la lente transformation des sociétés paysannes de l'âge du Bronze. In L'archéologie préventive dans le monde, (La Découverte), pp. 42–56.

Carpenter, M.L., Buenrostro, J.D., Valdiosera, C., Schroeder, H., Allentoft, M.E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., 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.

Casals, F., Sikora, M., Laayouni, H., Montanucci, L., Muntasell, A., Lazarus, R., Calafell, F., Awadalla, P., Netea, M.G., and Bertranpetit, J. (2011). Genetic adaptation of the antibacterial human innate immunity network. BMC Evol. Biol. *11*, 202.

Castellano, S., Parra, G., Sánchez-Quinto, F.A., Racimo, F., Kuhlwilm, M., Kircher, M., Sawyer, S., Fu, Q., Heinze, A., Nickel, B., et al. (2014). Patterns of coding variation in the complete exomes of three Neandertals. Proc. Natl. Acad. Sci. U. S. A. *111*, 6666–6671.

Ceppellini, R., Curtoni, E., Mattiuz, P., Miggiano, V., Seudelder, G., and Serra, A. (1967). Genetics of leukocyte antigens: a family study of segregation and linkage. Curtoni E Mattiuz P Tosi R Eds Histocompat. Test. 149–185.

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E.A., Grange, T., and Geigl, E.-M. (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. PloS One *5*.

Chang, J.T. RECENT COMMON ANCESTORS OF ALL PRESENT-DAY INDIVIDUALS. 25.

Chargaff, E. (1950). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia *6*, 201–209.

Childe, G. (1950). What happened in history.

Childe, V.G. (1929). The Danube in prehistory (Clarendon Press).

Childe, V.G. (1936). Man makes himself. Watts Lond.

Clarke, L.J., Czechowski, P., Soubrier, J., Stevens, M.I., and Cooper, A. (2014). Modular tagging of amplicons using a single PCR for high-throughput sequencing. Mol. Ecol. Resour. *14*, 117–121.

Collins, M.J., Nielsen–Marsh, C.M., Hiller, J., Smith, C.I., Roberts, J.P., Prigodich, R.V., Wess, T.J., Csapò, J., Millard, A.R., and Turner–Walker, G. (2002). The survival of organic matter in bone: a review. Archaeometry *44*, 383–394.

Collis, J. (1984). Oppida: earliest towns north of the Alps (Department of Prehistory and Archaeology, University of Sheffield).

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., et al. (2013). 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.

Damgaard, P.B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., and Allentoft, M.E. (2015). Improving access to endogenous DNA in ancient bones and teeth. Sci. Rep. *5*.

Darwin, C. (1859). On the Origin of Species, 1859 (Routledge).

Darwin, C. (1868). The Variation of Animals and Plants Under Domestication (O. Judd).

Demoule, J.-P. (1998). Les indo-européens, un mythe sur mesure. La Recherche 40–47.

Demoule, J.-P. (2007). La révolution néolithique en France (La Découverte).

Demoule, J.-P. (2014). Mais où sont passés les Indo-Européens?: le mythe d'origine de l'Occident (Éditions du Seuil).

Dunne, D.W., and Cooke, A. (2005). A worm's eye view of the immune system: consequences for evolution of human autoimmune disease. Nat. Rev. Immunol. *5*, 420–426.

Eckhart, L., Bach, J., Ban, J., and Tschachler, E. (2000). Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. Biochem. Biophys. Res. Commun. *271*, 726–730.

Fortes, G.G., Speller, C.F., Hofreiter, M., and King, T.E. (2013). Phenotypes from ancient DNA: approaches, insights and prospects. BioEssays News Rev. Mol. Cell. Dev. Biol. *35*, 690–695.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Petri, A.A., Prüfer, K., de Filippo, C., et al. (2014). The genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.

Fumagalli, M., Pozzoli, U., Cagliani, R., Comi, G.P., Riva, S., Clerici, M., Bresolin, N., and Sironi, M. (2009). Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. J. Exp. Med. *206*, 1395–1408.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L., and Nielsen, R. (2011). Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. PLOS Genet. *7*, e1002355.

Furtwängler, A., Reiter, E., Neumann, G.U., Siebke, I., Steuri, N., Hafner, A., Lösch, S., Anthes, N., Schuenemann, V.J., and Krause, J. (2018). Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. Sci. Rep. *8*, 14075.

Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. *5*, 5257.

Gates, K.S. (2009). An Overview of Chemical Processes That Damage Cellular DNA: Spontaneous Hydrolysis, Alkylation, and Reactions with Radicals. Chem. Res. Toxicol. *22*, 1747–1760.

Geigl, E.-M. (2002). On the circumstances surrounding the preservation and analysis of very old DNA. Archaeometry *44*, 337–342.

Gerbault, P., Allaby, R.G., Boivin, N., Rudzinski, A., Grimaldi, I.M., Pires, J.C., Vigueira, C.C., Dobney, K., Gremillion, K.J., Barton, L., et al. (2014). Storytelling and story testing in domestication. Proc. Natl. Acad. Sci. *111*, 6159–6164.

Gerstenblith, M.R., Goldstein, A.M., Fargnoli, M.C., Peris, K., and Landi, M.T. (2007). Comprehensive evaluation of allele frequency differences of MC1R variants across populations. Hum. Mutat. *28*, 495–505.

Gettings, K.B., Lai, R., Johnson, J.L., Peck, M.A., Hart, J.A., Gordish-Dressman, H., Schanfield, M.S., and Podini, D.S. (2014). A 50-SNP assay for biogeographic ancestry and phenotype prediction in the U.S. population. Forensic Sci. Int. Genet. *8*, 101–108.

Gilbert, M.T.P., Bandelt, H.-J., Hofreiter, M., and Barnes, I. (2005). Assessing ancient DNA studies. Trends Ecol. Evol. *20*, 541–544.

Ginolhac, A., Rasmussen, M., Gilbert, M.T.P., Willerslev, E., and Orlando, L. (2011). mapDamage: testing for damage patterns in ancient DNA sequences. Bioinforma. Oxf. Engl. *27*, 2153–2155.

Glocke, I., and Meyer, M. (2017). Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. Genome Res. *27*, 1230–1237.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., et al. (2009). Solution Hybrid Selection with Ultra-long Oligonucleotides for Massively Parallel Targeted Sequencing. Nat. Biotechnol. *27*, 182–189.

Goldberg, A., Günther, T., Rosenberg, N.A., and Jakobsson, M. (2017). Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. Proc. Natl. Acad. Sci. 201616392.

Green, R.E., Briggs, A.W., Krause, J., Prüfer, K., Burbano, H.A., Siebauer, M., Lachmann, M., and Pääbo, S. (2009). The Neandertal genome and ancient DNA authenticity. EMBO J. *28*, 2494–2502.

Gronenborn, D. (2014). The Persistence of Hunting and Gathering. Oxf. Handb. Archaeol. Anthropol. Hunt.-Gatherers.

Grunenwald, A., Keyser, C., Sautereau, A.-M., Crubézy, E., Ludes, B., and Drouet, C. (2014). Adsorption of DNA on biomimetic apatites: Toward the understanding of the role of bone and tooth mineral on the preservation of ancient DNA. Appl. Surf. Sci. *292*, 867–875.

Guilaine, J. (1972). L'Âge du bronze en Languedoc occidental, Roussillon, Ariège (Paris, France: Éditions Klincksieck, DL 1972).

Guilaine, J., and Manen, C. (2004). From Mesolithic to Early Neolithic in the western Mediterranean. (Oxford University press), pp. 21–51.

Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W., et al. (2005). Ancient DNA from the First European Farmers in 7500-Year-Old Neolithic Sites. Science *310*, 1016–1018.

Haak, W., Brandt, G., Jong, H.N. de, Meyer, C., Ganslmeier, R., Heyd, V., Hawkesworth, C., Pike, A.W.G., Meller, H., and Alt, K.W. (2008). Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. Proc. Natl. Acad. Sci. *105*, 18226–18231.

Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Sarkissian, C.S.I.D., Brandt, G., Schwarz, C., Nicklisch, N., et al. (2010). Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities. PLOS Biol. *8*, e1000536.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature.

Hagelberg, E., and Clegg, J.B. (1991). Isolation and characterization of DNA from archaeological bone. Proc R Soc Lond B *244*, 45–50.

Haile, J. (2012). Ancient DNA extraction from soils and sediments. Methods Mol. Biol. Clifton NJ *840*, 57–63.

Hancock, A.M., Alkorta-Aranburu, G., Witonsky, D.B., and Di Rienzo, A. (2010). Adaptations to new environments in humans: the role of subtle allele frequency shifts. Philos. Trans. R. Soc. B Biol. Sci. *365*, 2459–2468.

Harding, A.F. (2000). European Societies in the Bronze Age (Cambridge University Press).

Hawkes, C.F.C., and Hawkes, S.C. (1973). Greeks, Celts, and Romans: studies in venture and resistance (Rowman and Littlefield).

Hervella, M., Rotea, M., Izagirre, N., Constantinescu, M., Alonso, S., Ioana, M., Lazăr, C., Ridiche, F., Soficaru, A.D., Netea, M.G., et al. (2015). Ancient DNA from South-East Europe Reveals Different Events during Early and Middle Neolithic Influencing the European Genetic Heritage. PLOS ONE *10*, e0128810.

Higham, T., Compton, T., Stringer, C., Jacobi, R., Shapiro, B., Trinkaus, E., Chandler, B., Gröning, F., Collins, C., Hillson, S., et al. (2011). The earliest evidence for anatomically modern humans in northwestern Europe. Nature *479*, 521–524.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., and Wilson, A.C. (1984). DNA sequences from the quagga, an extinct member of the horse family. Nature *312*, 282–284.

Hirszfeld, L., and Hirszfeldowa, H. (1919). Essai d'application des méthodes sérologiques au problème des races.

Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-del-Molino, D., Dorp, L. van, López, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. Proc. Natl. Acad. Sci. *113*, 6886–6891.

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., and Pääbo, S. (2001). Ancient DNA. Nat. Rev. Genet. *2*, 353–359.

Hofreiter, M., Rabeder, G., Jaenicke-Després, V., Withalm, G., Nagel, D., Paunovic, M., Jambrĕsić, G., and Pääbo, S. (2004). Evidence for Reproductive Isolation between Cave Bear Populations. Curr. Biol. *14*, 40–43.

Hubert, H. (1950). Les Celtes depuis l'époque de La Tène et la civilisation celtique (A. Michel).

Huff, C.D., Witherspoon, D.J., Zhang, Y., Gatenbee, C., Denson, L.A., Kugathasan, S., Hakonarson, H., Whiting, A., Davis, C.T., Wu, W., et al. (2012). Crohn's disease and genetic hitchhiking at IBD5. Mol. Biol. Evol. *29*, 101–111.

Hunt, K.A., Zhernakova, A., Turner, G., Heap, G.A.R., Franke, L., Bruinenberg, M., Romanos, J., Dinesen, L.C., Ryan, A.W., Panesar, D., et al. (2008). Newly identified genetic risk variants for celiac disease related to the immune response. Nat. Genet. *40*, 395–402.

Hysi, P.G., Valdes, A.M., Liu, F., Furlotte, N.A., Evans, D.M., Bataille, V., Visconti, A., Hemani, G., McMahon, G., Ring, S.M., et al. (2018). Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. Nat. Genet. *50*, 652–656.

Jacquard, A. (1978). Eloge de la différence. La génétique et les hommes (Le Seuil).

Jeunesse, C. (2015). The dogma of the Iberian origin of the Bell Beaker: attempting its deconstruction. J. Neolit. Archaeol. 158–166.

Joffroy, R. (1979). Vix et ses trésors (Tallandier).

Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. *6*, 8912.

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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., et al. (2012). New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat. Commun. *3*, 698.

Kimura, M. (1983). The Neutral Theory of Molecular Evolution (Cambridge University Press).

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*.

Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., and Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature *464*, 894–897.

Kristiansen, K., and Larsson, T.B. (2005). The Rise of Bronze Age Society: Travels, Transmissions and Transformations (Cambridge University Press).

Krüttli, A., Bouwman, A., Akgül, G., Casa, P.D., Rühli, F., and Warinner, C. (2014). Ancient DNA Analysis Reveals High Frequency of European Lactase Persistence Allele (T-13910) in Medieval Central Europe. PLOS ONE *9*, e86251.

Kurosaki, K., Matsushita, T., and Ueda, S. (1993). Individual DNA identification from ancient human remains. Am. J. Hum. Genet. *53*, 638–643.

Lacan, M., Keyser, C., Ricaut, F.-X., Brucato, N., Tarrús, J., Bosch, A., Guilaine, J., Crubézy, E., and Ludes, B. (2011). Ancient DNA suggests the leading role played by men in the Neolithic dissemination. Proc. Natl. Acad. Sci. U. S. A. *108*, 18255–18259.

Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D., et al. (2008). Correlation between genetic and geographic structure in Europe. Curr. Biol. CB *18*, 1241–1248.

Larsen, B.B., Miller, E.C., Rhodes, M.K., and Wiens, J.J. (2017). Inordinate Fondness Multiplied and Redistributed: the Number of Species on Earth and the New Pie of Life. Q. Rev. Biol. *92*, 229–265.

Lazaridis, I., and Reich, D. (2017). Failure to Replicate a Genetic Signal for Sex Bias in the Steppe Migration into Central Europe. BioRxiv 114124.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.

Leonard, J.A., Wayne, R.K., and Cooper, A. (2000). Population genetics of Ice Age brown bears. Proc. Natl. Acad. Sci. *97*, 1651–1654.

Levene, P.A. (1919). The Structure of Yeast Nucleic Acid Iv. Ammonia Hydrolysis. J. Biol. Chem. *40*, 415–424.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. *25*, 1754–1760.

Lichardus, J., and Lichardus-Itten, M. (1985). La protohistoire de l'Europe.

van Lieshout, A.W., van der Voort, R., le Blanc, L.M., Roelofs, M.F., Schreurs, B.W., van Riel, P.L., Adema, G.J., and Radstake, T.R. (2006). Novel insights in the regulation of CCL18 secretion by monocytes and dendritic cells via cytokines, Toll-like receptors and rheumatoid synovial fluid. BMC Immunol. *7*, 23.

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

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.

Loeb, L.A. (1989). Endogenous Carcinogenesis: Molecular Oncology into the Twenty-first Century- Presidential Address. Mol. Oncol. 9.

Loreille, O., Roumat, E., Verneau, O., Bouchet, F., and Hänni, C. (2001). Ancient DNA from Ascaris: extraction amplification and sequences from eggs collected in coprolites. Int. J. Parasitol. *31*, 1101–1106.

Luca Cavalli-Sforza, L., Menozzi, P., and Piazza, A. (1995). The History and Geography of Human Gene. In Princeton, New Jersey, USA, p.

Malmström, H., Gilbert, M.T.P., Thomas, M.G., Brandström, M., Storå, J., Molnar, P., Andersen, P.K., Bendixen, C., Holmlund, G., Götherström, A., et al. (2009). Ancient DNA

reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. Curr. Biol. CB *19*, 1758–1762.

Maricic, T., Whitten, M., and 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.Journal *17*, 10–12.

Massilani, D., Guimaraes, S., Brugal, J.-P., Bennett, E.A., Tokarska, M., Arbogast, R.-M., Baryshnikov, G., Boeskorov, G., Castel, J.-C., Davydov, S., et al. (2016). Past climate changes, population dynamics and the origin of Bison in Europe. BMC Biol. *14*, 93.

Matheson, C.D., Gurney, C., Esau, N., and Lehto, R. (2010). Assessing PCR Inhibition from Humic Substances. 8.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

Mayr, E. (1942). Systematics and the Origin of Species, from the Viewpoint of a Zoologist (Harvard University Press).

Mazurié de Keroualin, K. (2003). Génèse et diffusion de l'agriculture en Europe: agriculteurs-chasseurs-pasteurs (Editions Errance).

McKenzie, G.J., Bancroft, A., Grencis, R.K., and McKenzie, A.N.J. (1998). A distinct role for interleukin-13 in Th2-cell-mediated immune responses. Curr. Biol. *8*, 339–342.

Mellars, P. (2006). Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. Science *313*, 796–800.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., Filippo, C. de, et al. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science *338*, 222–226.

Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.-L., Martínez, I., Gracia, A., Castro, J.M.B. de, Carbonell, E., et al. (2014). A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature *505*, 403–406.

Mille, B., and Bouquet, L. (2004). Le métal au 3e millénaire avant notre ère dans le Centre-Nord de la France. Anthropol. Praehist. *115*, 197–215.

Motulsky, A.G. (1984). Environmental Mutagenesis and Disease in Human Populations. In Mutation, Cancer, and Malformation, E.H.Y. Chu, and W.M. Generoso, eds. (Boston, MA: Springer US), pp. 1–11.

Mourant, A.E. (1954). The Distribution of the Human Blood Groups. Distrib. Hum. Blood Groups.

Myres, N.M., Rootsi, S., Lin, A.A., Järve, M., King, R.J., Kutuev, I., Cabrera, V.M., Khusnutdinova, E.K., Pshenichnov, A., Yunusbayev, B., et al. (2011). A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur. J. Hum. Genet. *19*, 95–101.

Navarro-Gomez, D., Leipzig, J., Shen, L., Lott, M., Stassen, A.P.M., Wallace, D.C., Wiggs, J.L., Falk, M.J., van Oven, M., and Gai, X. (2015). Phy-Mer: a novel alignment-free and reference-independent mitochondrial haplogroup classifier. Bioinforma. Oxf. Engl. *31*, 1310–1312.

Nei, M., Maruyama, T., and Chakraborty, R. (1975). The Bottleneck Effect and Genetic Variability in Populations. Evolution *29*, 1–10.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., et al. (2008). Genes mirror geography within Europe. Nature *456*, 98–101.

O'Dushlaine, C.T., Morris, D., Moskvina, V., Kirov, G., Consortium, I.S., Gill, M., Corvin, A., Wilson, J.F., and Cavalleri, G.L. (2010). Population structure and genome-wide patterns of variation in Ireland and Britain. Eur. J. Hum. Genet. *18*, 1248–1254.

Olalde, I., Allentoft, M.E., Sánchez-Quinto, F., Santpere, G., Chiang, C.W.K., DeGiorgio, M., Prado-Martinez, J., Rodríguez, J.A., Rasmussen, S., Quilez, J., et al. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature *507*, 225–228.

Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., et al. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature *499*, 74–78.

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

Pääbo, S., Gifford, J.A., and Wilson, A.C. (1988). Mitochondrial DNA sequences from a 7000-year old brain. Nucleic Acids Res. *16*, 9775–9787.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M. (2004). Genetic analyses from ancient DNA. Annu. Rev. Genet. *38*, 645–679.

Patterson, N., Richter, D.J., Gnerre, S., Lander, E.S., and Reich, D. (2006). Genetic evidence for complex speciation of humans and chimpanzees. Nature *441*, 1103–1108.

Peisong, G., Yamasaki, A., Mao, X.-Q., Enomoto, T., Feng, Z., Gloria-Bottini, F., Bottini, E., Shirakawa, T., Sun, D., and Hopkin, J.M. (2004). An asthma-associated genetic variant of STAT6 predicts low burden of ascaris worm infestation. Genes Immun. *5*, 58–62.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., et al. (2015). Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. PloS One *10*, e0129102.

Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., et al. (2016). Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. Curr. Biol. *26*, 827–833.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.

Pruvost, M., Grange, T., and Geigl, E.-M. (2005). Minimizing DNA contamination by using UNG-coupled quantitative real-time PCR on degraded DNA samples: application to ancient DNA studies. BioTechniques *38*, 569–575.

Pruvost, M., Schwarz, R., Correia, V.B., Champlot, S., Braguier, S., Morel, N., Fernandez-Jalvo, Y., Grange, T., and Geigl, E.-M. (2007). Freshly excavated fossil bones are best for amplification of ancient DNA. Proc. Natl. Acad. Sci. U. S. A. *104*, 739–744.

Pruvost, M., Schwarz, R., Correia, V.B., Champlot, S., Grange, T., and Geigl, E.-M. (2008). DNA diagenesis and palaeogenetic analysis : Critical assessment and methodological progress. Palaeogeogr. Palaeoclimatol. Palaeoecol. *266*, *issues 3-4*, 211–219.

Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., Sica, L., Mouguiama-Daouda, P., Comas, D., Tzur, S., et al. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter–gatherers and Bantu-speaking farmers. Proc. Natl. Acad. Sci. U. S. A. *105*, 1596–1601.

Racimo, F., Renaud, G., and Slatkin, M. (2016). Joint Estimation of Contamination, Error and Demography for Nuclear DNA from Ancient Humans. PLOS Genet. *12*, e1005972.

Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S., Grønnow, B., Appelt, M., Gulløv, H.C., Friesen, T.M., et al. (2014). The genetic prehistory of the New World Arctic. Science *345*, 1255832.

Ralf, A., Montiel González, D., Zhong, K., Kayser, M., and Su, B. (2018). Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. Mol. Biol. Evol. *35*, 1291–1294.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. *102*, 15942–15947.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053–1060.

Reingruber, A., and Thissen, L. (2011). Depending on 14C Data: Chronological Frameworks in the Neolithic and Chalcolithic of Southeastern Europe. Radiocarbon *51*, 751–770.

Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., et al. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. Am. J. Hum. Genet. *67*, 1251–1276.

Rivollat, M., Mendisco, F., Pemonge, M.-H., Safi, A., Saint-Marc, D., Brémond, A., Couture-Veschambre, C., Rottier, S., and Deguilloux, M.-F. (2015). When the Waves of European Neolithization Met: First Paleogenetic Evidence from Early Farmers in the Southern Paris Basin. PLOS ONE *10*, e0125521.

Rivollat, M., Réveillas, H., Mendisco, F., Pemonge, M.-H., Justeau, P., Couture, C., Lefranc, P., Féliu, C., and Deguilloux, M.-F. (2016). Ancient mitochondrial DNA from the middle neolithic necropolis of Obernai extends the genetic influence of the LBK to west of the Rhine. Am. J. Phys. Anthropol. *161*, 522–529.

Rohde, D.L.T., Olson, S., and Chang, J.T. (2004). Modelling the recent common ancestry of all living humans. Nature *431*, 562–566.

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

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *370*, 20130624.

Rollo, F., Amici, A., Salvi, R., and Garbuglia, A. (1988). Short but faithful pieces of ancient DNA. Nature *335*, 774.

Rueckert, A., and Morgan, H.W. (2007). Removal of contaminating DNA from polymerase chain reaction using ethidium monoazide. J. Microbiol. Methods *68*, 596–600.

Sampietro, M.L., Lao, O., Caramelli, D., Lari, M., Pou, R., Martí, M., Bertranpetit, J., and Lalueza-Fox, C. (2007). Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. Proc. R. Soc. Lond. B Biol. Sci. *274*, 2161–2167.

Sánchez-Quinto, F., Schroeder, H., Ramirez, O., Avila-Arcos, M.C., Pybus, M., Olalde, I., Velazquez, A.M.V., Marcos, M.E.P., Encinas, J.M.V., Bertranpetit, J., et al. (2012). Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. Curr. Biol. CB *22*, 1494–1499.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. PLOS ONE *7*, e34131.

Seguin-Orlando, A., Korneliussen, T.S., Sikora, M., Malaspinas, A.-S., Manica, A., Moltke, I., Albrechtsen, A., Ko, A., Margaryan, A., Moiseyev, V., et al. (2014). Genomic structure in Europeans dating back at least 36,200 years. Science *346*, 1113–1118.

Semino, O. (2000). The Genetic Legacy of Paleolithic Homo sapiens sapiens in Extant Europeans: A Y Chromosome Perspective. Science *290*, 1155–1159.

Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T.P., Götherström, A., and Jakobsson, M. (2012). Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. Science *336*, 466–469.

Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. J. Archaeol. Sci. *40*, 4477–4482.

Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., Hall, P., Tambets, K., Parik, J., Sjögren, K.-G., et al. (2014). Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. Science *344*, 747–750.

Smit, A., Hubley, R., and Green, P. (2015). RepeatMasker Open-4.0. (2013-2015).

Smith, C.I., Chamberlain, A.T., Riley, M.S., Cooper, A., Stringer, C.B., and Collins, M.J. (2001). Neanderthal DNA: Not just old but old and cold? Nature *410*, 771–772.

Smith, T.F., Waterman, M.S., and Burks, C. (1985). The statistical distribution of nucleic acid similarities. Nucleic Acids Res. *13*, 645–656.

Sokal, R.R., and Menozzi, P. (1982). Spatial Autocorrelations of HLA Frequencies in Europe Support Demic Diffusion of Early Farmers. Am. Nat. *119*, 1–17.

Stoneking, M., Jorde, L.B., Bhatia, K., and Wilson, A.C. (1990). Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics *124*, 717–733.

Stringer, C.B., and Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. Science *239*, 1263–1268.

Szécsényi-Nagy, A., Roth, C., Brandt, G., Rihuete-Herrada, C., Tejedor-Rodríguez, C., Held, P., García-Martínez-de-Lagrán, Í., Arcusa Magallón, H., Zesch, S., Knipper, C., et al. (2017). The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age. Sci. Rep. *7*.

Thomas, R.H., Schaffner, W., Wilson, A.C., and Pääbo, S. (1989). DNA phylogeny of the extinct marsupial wolf. Nature *340*, 465–467.

Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., et al. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. *39*, 31–40.

Trinkaus, E., Moldovan, O., Milota, Ş., Bîlgăr, A., Sarcina, L., Athreya, S., Bailey, S.E., Rodrigo, R., Mircea, G., Higham, T., et al. (2003). An early modern human from the Peştera cu Oase, Romania. Proc. Natl. Acad. Sci. *100*, 11231–11236.

Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonné-Tamir, B., Bertranpetit, J., Francalacci, P., et al. (2000). Y chromosome sequence variation and the history of human populations. Nat. Genet. *26*, 358–361.

Valdiosera, C., Günther, T., Vera-Rodríguez, J.C., Ureña, I., Iriarte, E., Rodríguez-Varela, R., Simões, L.G., Martínez-Sánchez, R.M., Svensson, E.M., Malmström, H., et al. (2018). Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. Proc. Natl. Acad. Sci. 201717762.

Van Lieshout, R.J., and MacQueen, G. (2008). Psychological Factors in Asthma. Allergy Asthma Clin. Immunol. *4*, 12.

van Oven Mannis, and Kayser Manfred (2008). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. *30*, E386–E394.

Vianello, D., Sevini, F., Castellani, G., Lomartire, L., Capri, M., and Franceschi, C. (2013). HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. Hum. Mutat. *34*, 1189–1194.

Vigne, J.-D., Briois, F., Zazzo, A., Willcox, G., Cucchi, T., Thiébault, S., Carrère, I., Franel, Y., Touquet, R., Martin, C., et al. (2012). First wave of cultivators spread to Cyprus at least 10,600 y ago. Proc. Natl. Acad. Sci. *109*, 8445–8449.

Wall, J.D., and Kim, S.K. (2007). Inconsistencies in Neanderthal Genomic DNA Sequences. PLOS Genet. *3*, e175.

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., and Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci. Int. Genet. *7*, 98–115.

Watson, J.D., and Crick, F.H.C. (1953). MOLECULAR STRUCTURE OF NUCLEIC ACIDS: A Structure for Deoxyribose Nucleic Acid. Nature 1.

Wess, T., Alberts, I., Hiller, J., Drakopoulos, M., Chamberlain, A.T., and Collins, M. (2001). Microfocus Small Angle X-ray Scattering Reveals Structural Features in Archaeological Bone Samples; Detection of Changes in BoneMineral Habit and Size. Calcif. Tissue Int. *70*, 103–110.

Willcox, G. (2005). The distribution, natural habitats and availability of wild cereals in relation to their domestication in the Near East: multiple events, multiple centres. Veg. Hist. Archaeobotany *14*, 534–541.

Willerslev, E., and Cooper, A. (2005). Ancient DNA. Proc. Biol. Sci. *272*, 3–16.

Wolfe, N.D., Dunavan, C.P., and Diamond, J. (2007). Origins of major human infectious diseases. Nature *447*, 279–283.

Xue, Y., Wang, Q., Long, Q., Ng, B.L., Swerdlow, H., Burton, J., Skuce, C., Taylor, R., Abdellah, Z., Zhao, Y., et al. (2009). Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. Curr. Biol. CB *19*, 1453–1457.

Yang, D.Y., Eng, B., Waye, J.S., Dudar, J.C., and Saunders, S.R. (1998). Technical note: improved DNA extraction from ancient bones using silica-based spin columns. Am. J. Phys. Anthropol. *105*, 539–543.

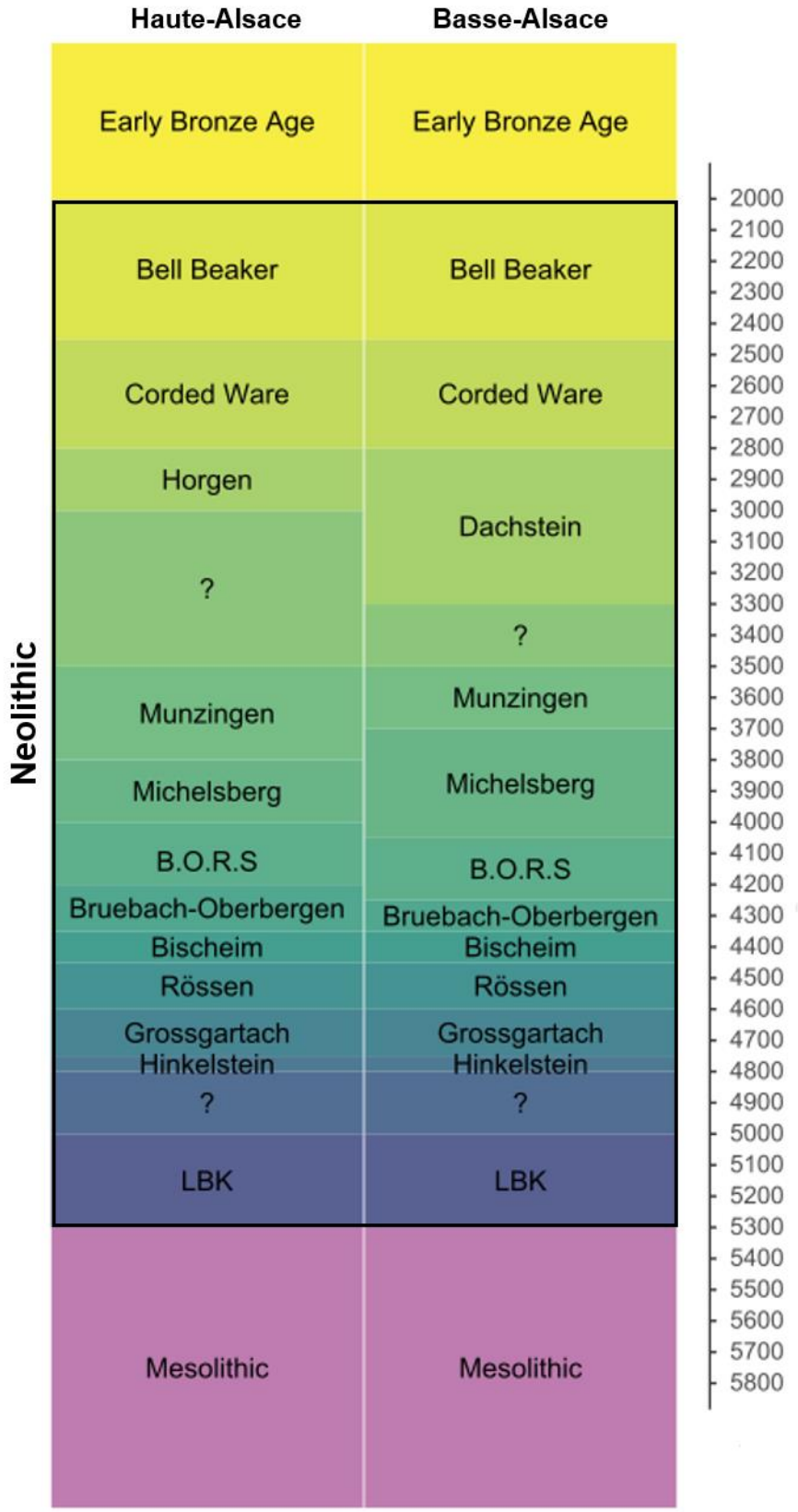
Yasuda, H., Kamide, K., Takiuchi, S., Matayoshi, T., Hanada, H., Kada, A., Yang, J., Miwa, Y., Yoshii, M., Horio, T., et al. (2007). Association of single nucleotide polymorphisms in endothelin family genes with the progression of atherosclerosis in patients with essential hypertension. J. Hum. Hypertens. *21*, 883–892.

Zeder, M.A. (2008). Animal Domestication in the Zagros: an Update and Directions for Future Research. MOM Éditions *49*, 243–277.

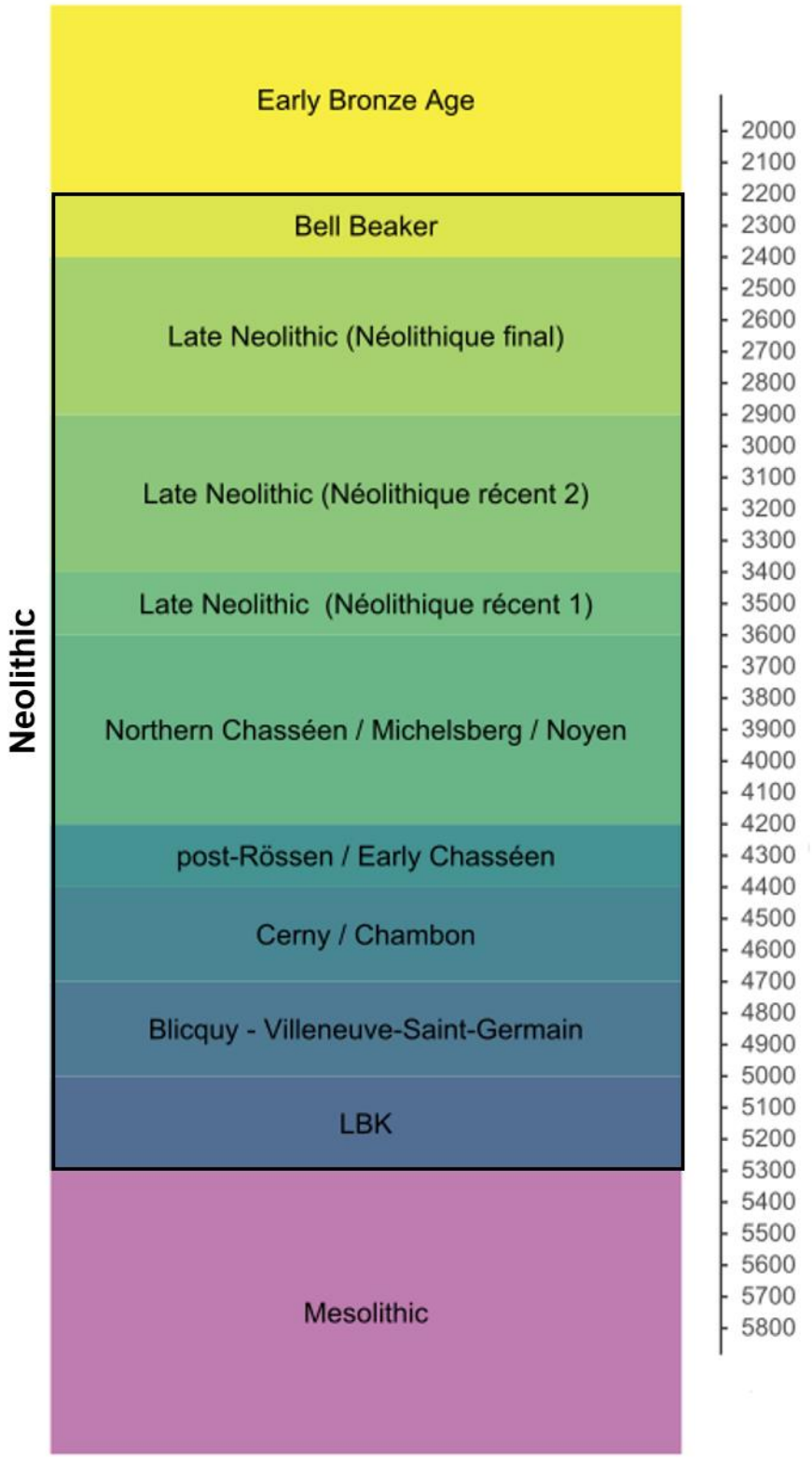
Zhou, L.-T., Qin, L., Zheng, D.-C., Song, Z.-K., and Ye, L. (2012). Meta-analysis of genetic association of chromosome 9p21 with early-onset coronary artery disease. Gene *510*, 185–188.

Supplementary Material

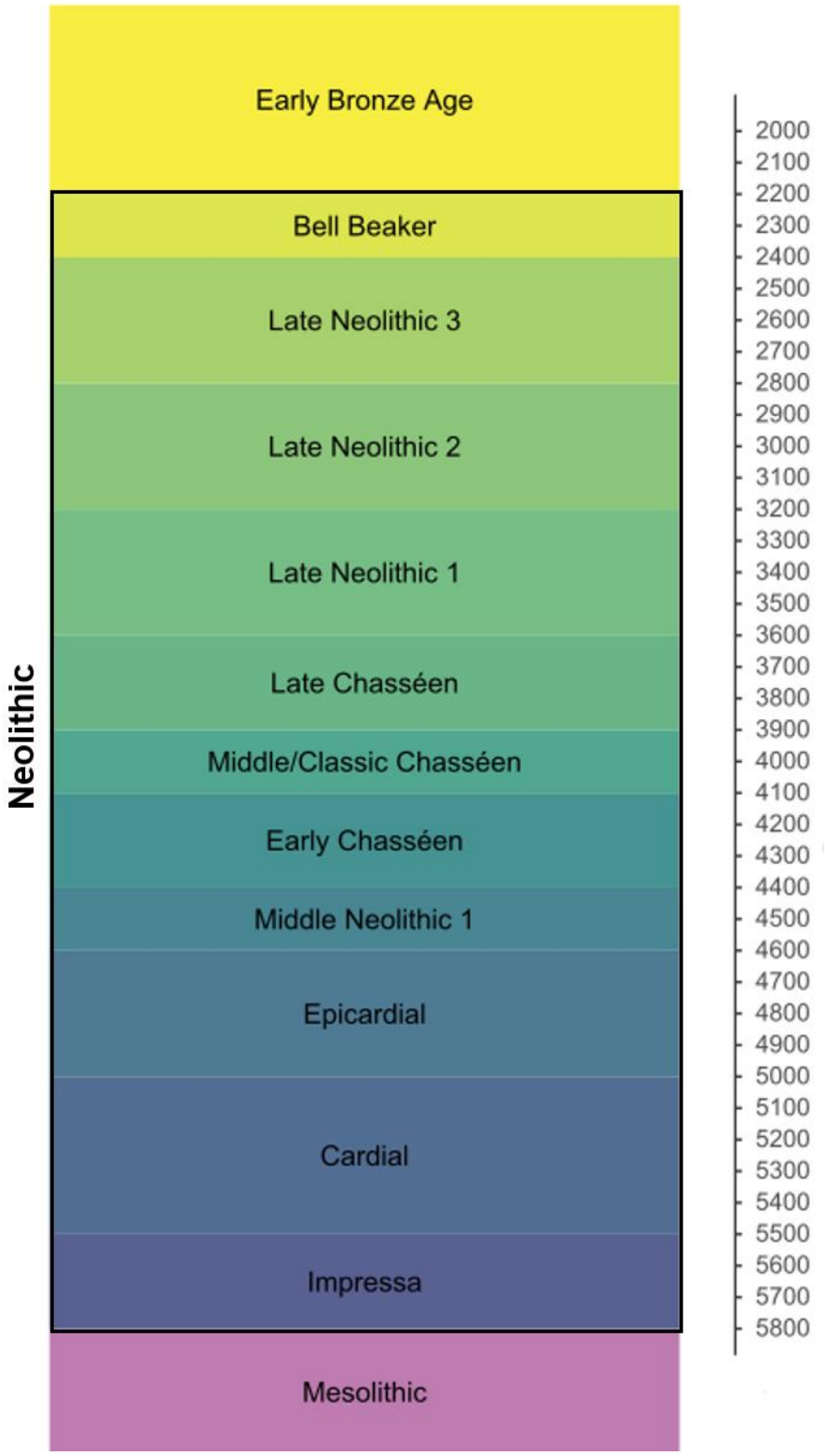
Supplementary Figures and Tables



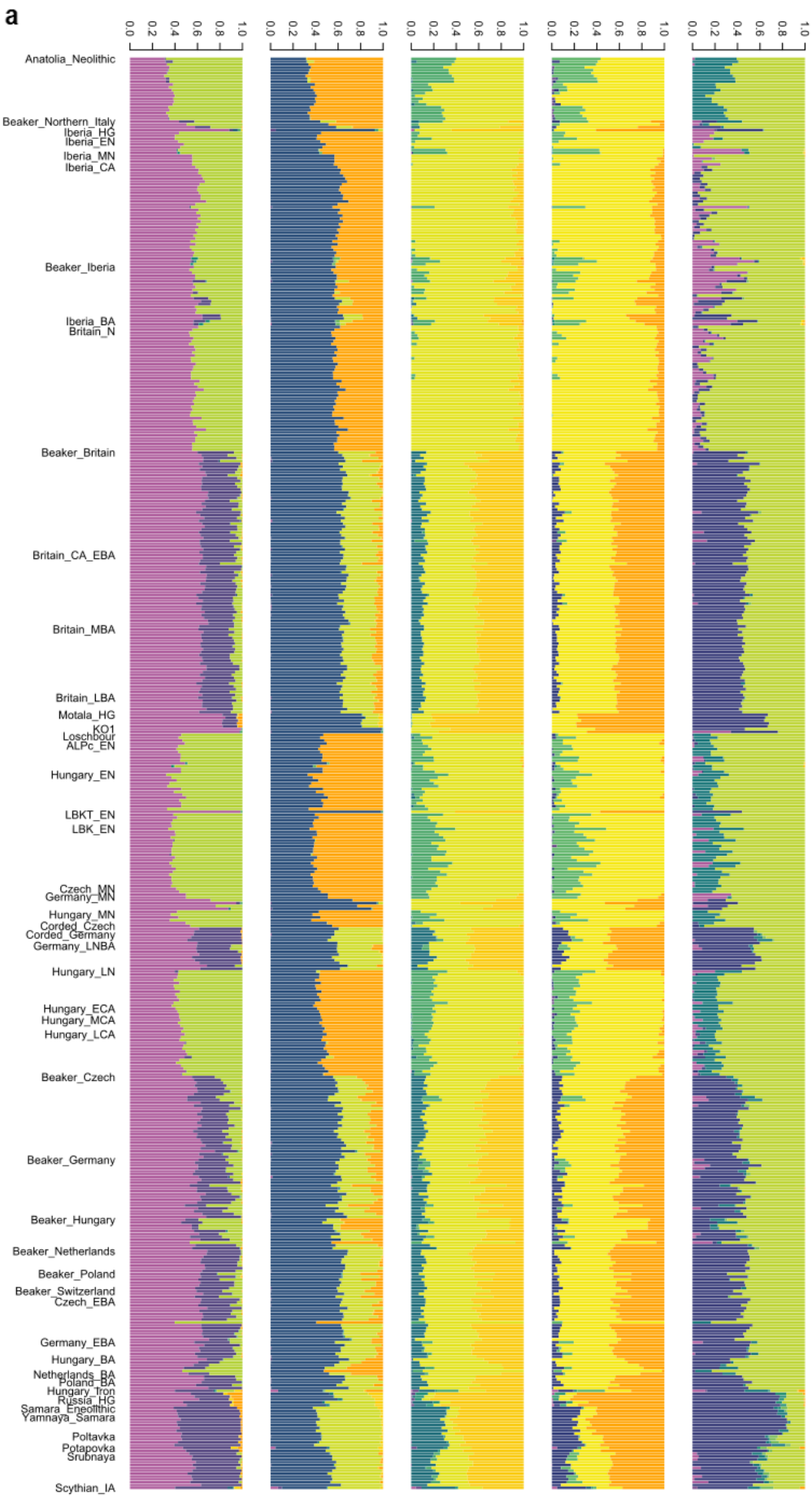
Supplementary Figure 1: Chronology of Neolithic cultures in northeastern France (Grand-Est). Dates are given in years BCE.

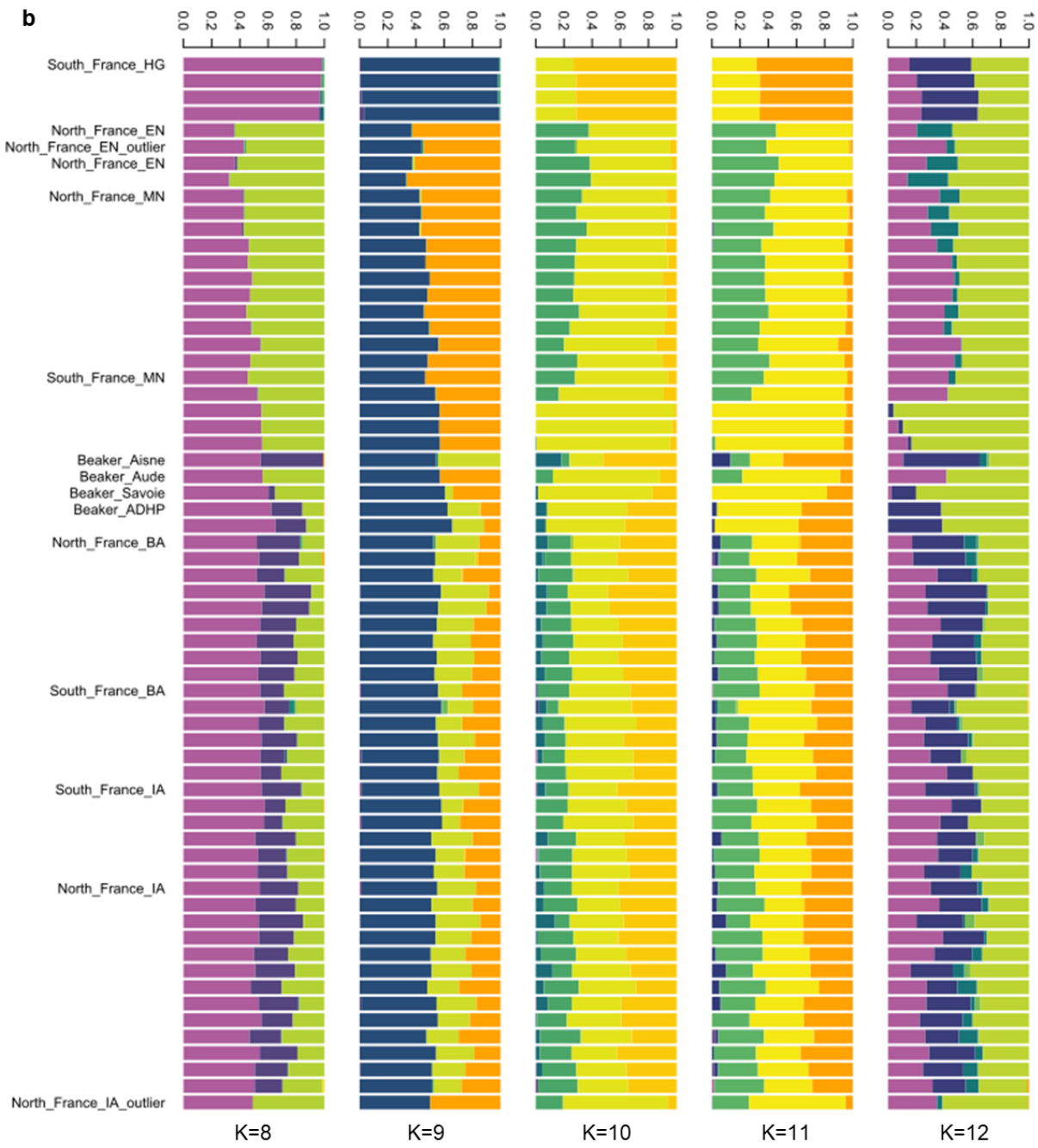


Supplementary Figure 2: Chronology of Neolithic cultures in the Paris Basin (Hauts-de-France). Dates are given in years BCE.



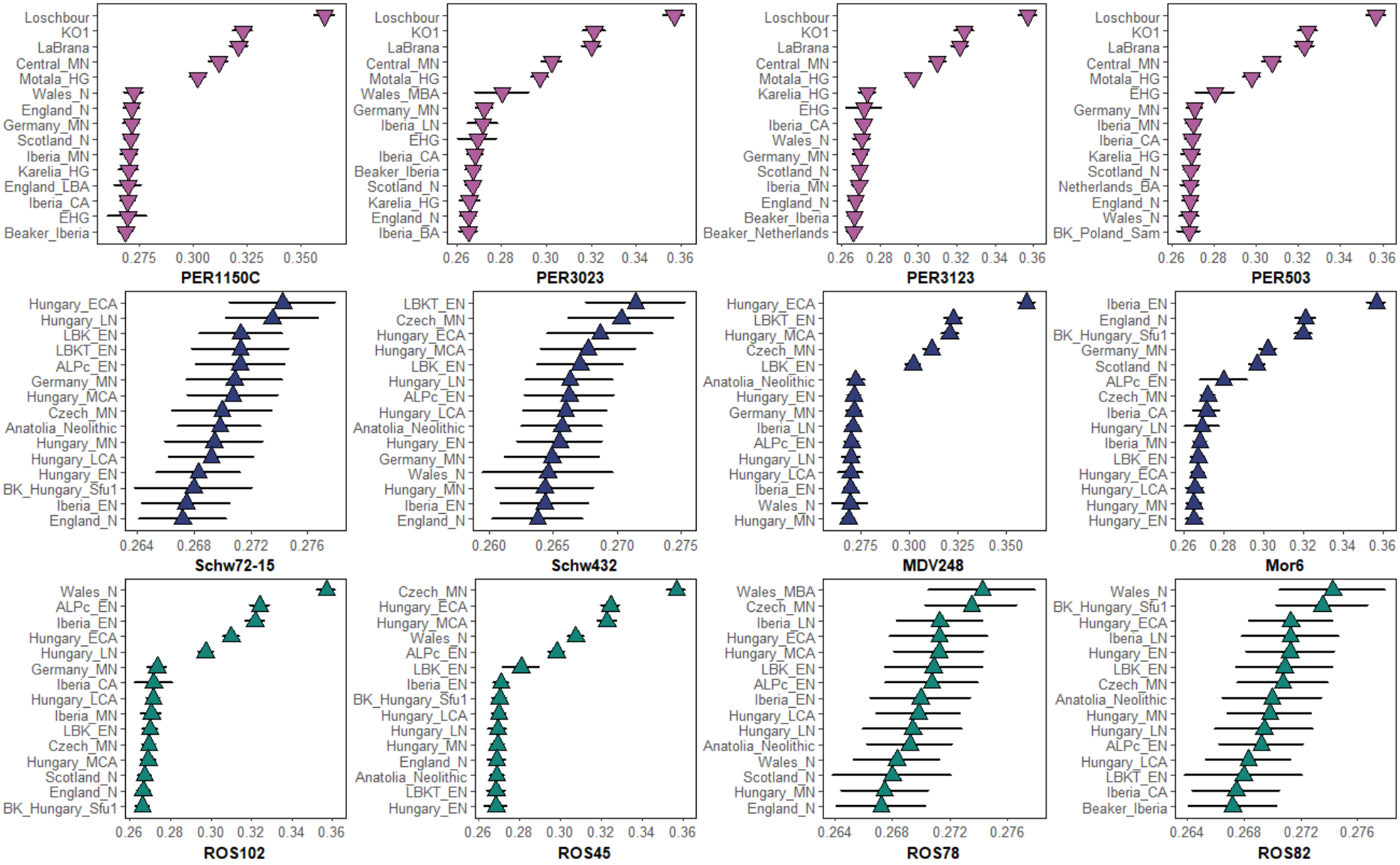
Supplementary Figure 3: Chronology of the Neolithic cultures in Southern France (Occitanie). Dates are given in years BCE.

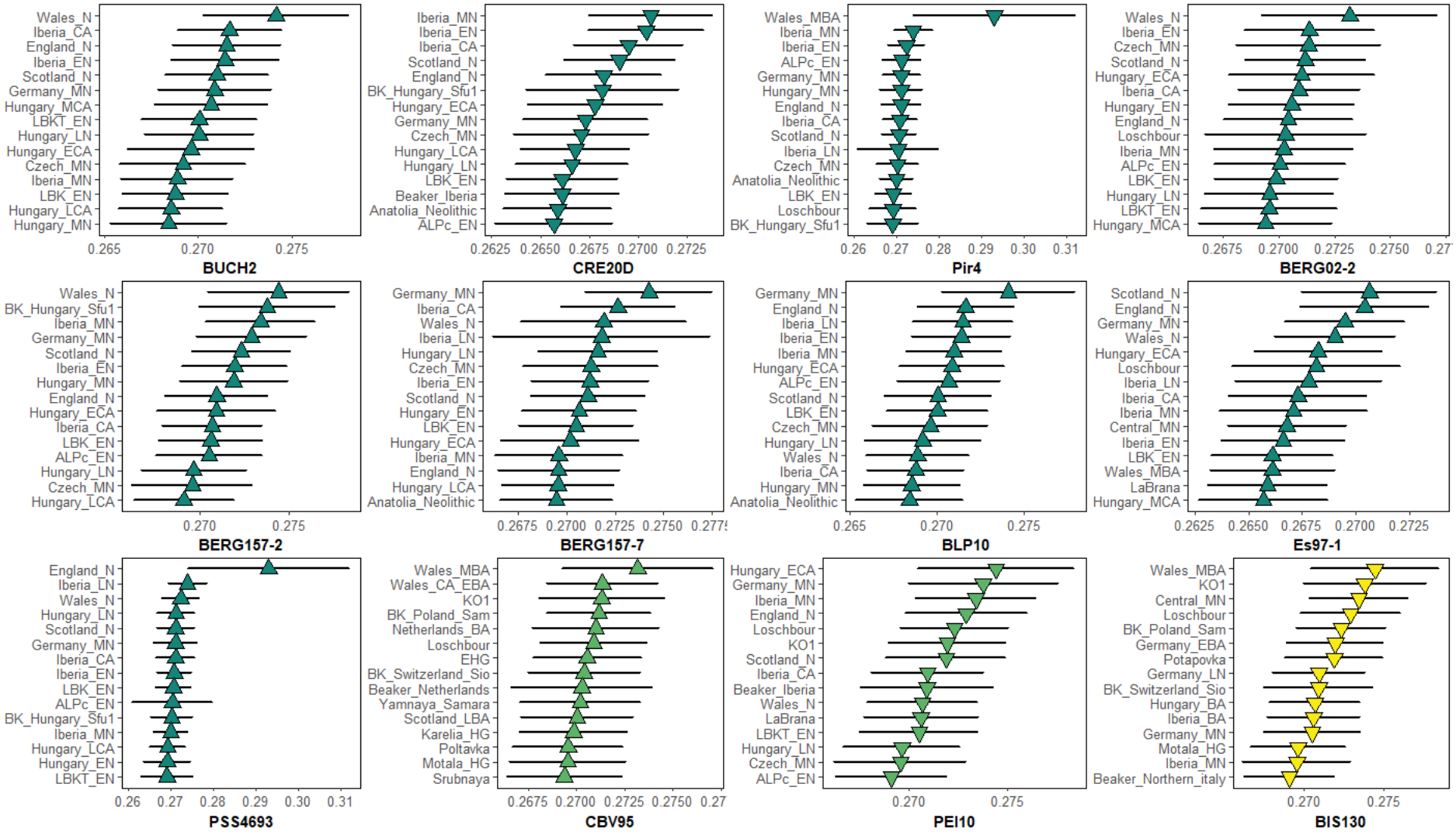

a
0.0
0.2
0.4
0.6
0.8
1.0
Anatolia_Neolithic
Beaker_Northern_Italy
Iberia_HG
Iberia_EN
Iberia_MN
Iberia_CA
Beaker_Iberia
Iberia_BA
Britain_N
Beaker_Britain
Britain_CA_EBA
Britain_MBA
Britain_LBA
Motala_HG
KO1
Loschbour
ALPc_EN
Hungary_EN
LBKT_EN
LBK_EN
Czech_MN
Germany_MN
Hungary_MN
Corded_Czech
Corded_Germany
Germany_LNBA
Hungary_LN
Hungary_ECA
Hungary_MCA
Hungary_LCA
Beaker_Czech
Beaker_Germany
Beaker_Hungary
Beaker_Netherlands
Beaker_Poland
Beaker_Switzerland
Czech_EBA
Germany_EBA
Hungary_BA
Netherlands_BA
Poland_BA
Hungary_Tron
Russia_HG
Samara_Eneolithic
Yamnaya_Samara
Poltavka
Potapovka
Srubnaya
Scythian_IA

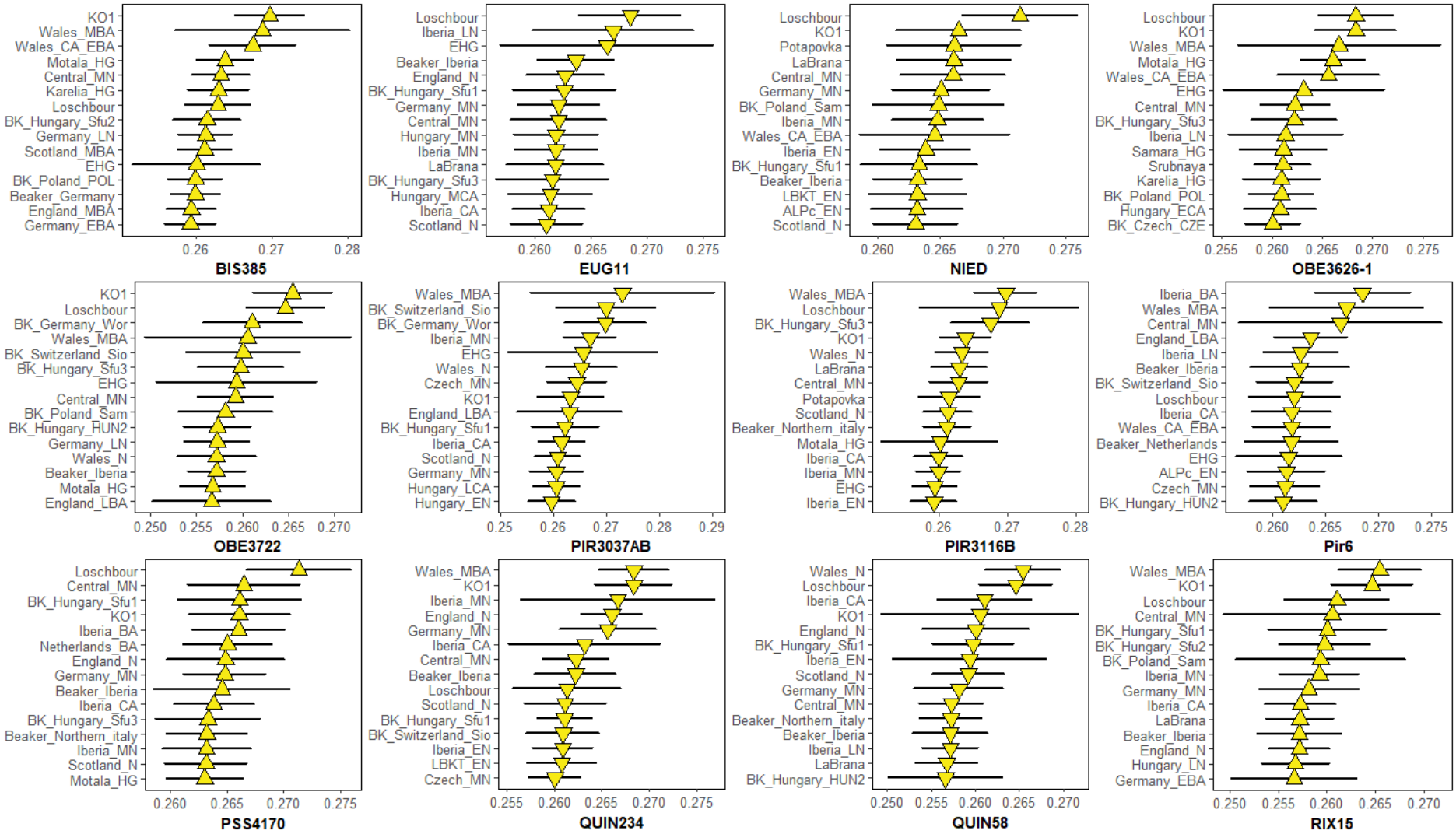


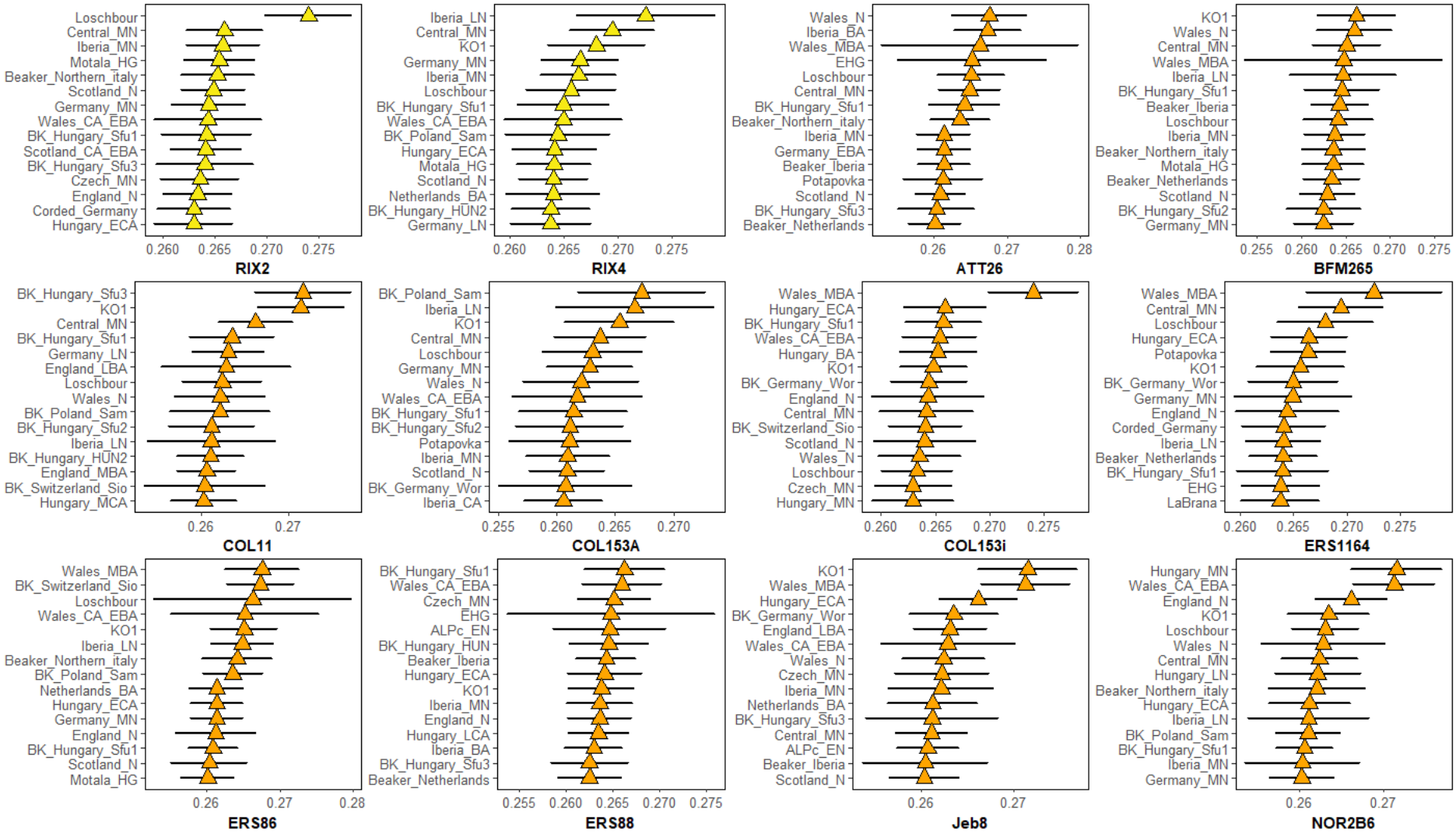
Supplementary Figure 4: Result of the unsupervised ADMIXTURE analysis performed on the whole dataset (ancient individuals and modern worldwide populations) for K=8 to K=12.

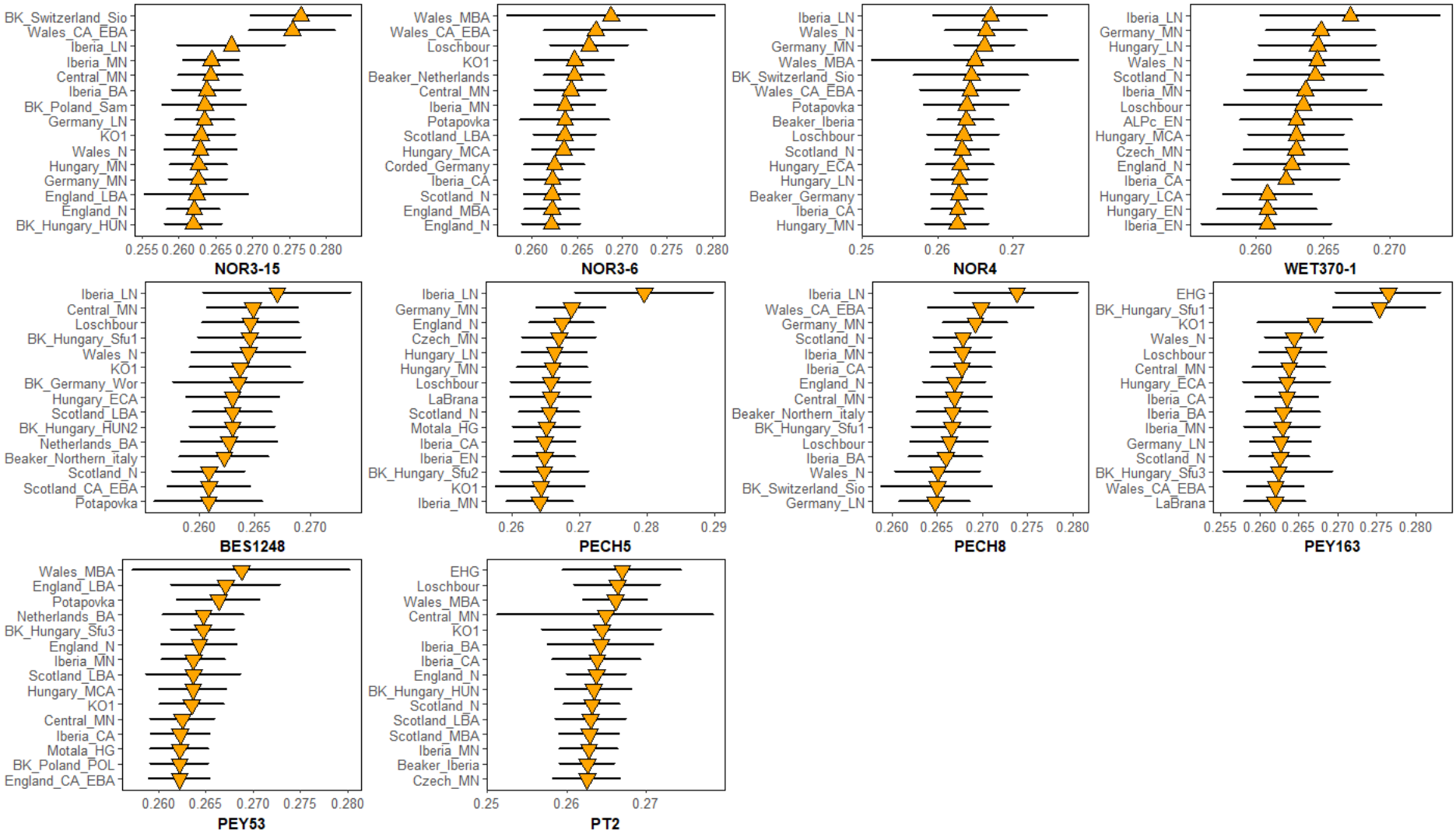
The lowest value of the cross-validation error was obtained for K=10 (CV error= 0.71041) **a.** Results for the published ancient dataset. **b**. Results for the French prehistoric population.


Loschbour
KO1
LaBrana
Central_MN
Motala_HG
Wales_N
England_N
Germany_MN
Scotland_N
Iberia_MN
Karelia_HG
England_LBA
Iberia_CA
EHG
Beaker_Iberia
0.275 0.300 0.325 0.350
PER1150C
Loschbour
KO1
LaBrana
Central_MN
Motala_HG
Wales_MBA
Germany_MN
Iberia_LN
EHG
Iberia_CA
Beaker_Iberia
Scotland_N
Karelia_HG
England_N
Iberia_BA
0.26 0.28 0.30 0.32 0.34 0.36
PER3023
Loschbour
KO1
LaBrana
Central_MN
Motala_HG
Karelia_HG
EHG
Iberia_CA
Wales_N
Germany_MN
Scotland_N
Iberia_MN
England_N
Beaker_Iberia
Beaker_Netherlands
0.26 0.28 0.30 0.32 0.34 0.36
PER3123
Loschbour
KO1
LaBrana
Central_MN
Motala_HG
EHG
Germany_MN
Iberia_MN
Iberia_CA
Karelia_HG
Scotland_N
Netherlands_BA
England_N
Wales_N
BK_Poland_Sam
0.26 0.28 0.30 0.32 0.34 0.36
PER503
Hungary_ECA
Hungary_LN
LBK_EN
LBKT_EN
ALPc_EN
Germany_MN
Hungary_MCA
Czech_MN
Anatolia_Neolithic
Hungary_MN
Hungary_LCA
Hungary_EN
BK_Hungary_Sfu1
Iberia_EN
England_N
0.264 0.268 0.272 0.276
Schw72-15
LBKT_EN
Czech_MN
Hungary_ECA
Hungary_MCA
LBK_EN
Hungary_LN
ALPc_EN
Hungary_LCA
Anatolia_Neolithic
Hungary_EN
Germany_MN
Wales_N
Hungary_MN
Iberia_EN
England_N
0.260 0.265 0.270 0.275
Schw432
Hungary_ECA
LBKT_EN
Hungary_MCA
Czech_MN
LBK_EN
Anatolia_Neolithic
Hungary_EN
Germany_MN
Iberia_LN
ALPc_EN
Hungary_LN
Hungary_LCA
Iberia_EN
Wales_N
Hungary_MN
0.275 0.300 0.325 0.350
MDV248
Iberia_EN
England_N
BK_Hungary_Sfu1
Germany_MN
Scotland_N
ALPc_EN
Czech_MN
Iberia_CA
Hungary_LN
Iberia_MN
LBK_EN
Hungary_ECA
Hungary_LCA
Hungary_MN
Hungary_EN
0.26 0.28 0.30 0.32 0.34 0.36
Mor6
Wales_N
ALPc_EN
Iberia_EN
Hungary_ECA
Hungary_LN
Germany_MN
Iberia_CA
Hungary_LCA
Iberia_MN
LBK_EN
Czech_MN
Hungary_MCA
Scotland_N
England_N
BK_Hungary_Sfu1
0.26 0.28 0.30 0.32 0.34 0.36
ROS102
Czech_MN
Hungary_ECA
Hungary_MCA
Wales_N
ALPc_EN
LBK_EN
Iberia_EN
BK_Hungary_Sfu1
Hungary_LCA
Hungary_LN
Hungary_MN
England_N
Anatolia_Neolithic
LBKT_EN
Hungary_EN
0.26 0.28 0.30 0.32 0.34 0.36
ROS45
Wales_MBA
Czech_MN
Iberia_LN
Hungary_ECA
Hungary_MCA
LBK_EN
ALPc_EN
Iberia_EN
Hungary_LCA
Hungary_LN
Anatolia_Neolithic
Wales_N
Scotland_N
Hungary_MN
England_N
0.264 0.268 0.272 0.276
ROS78
Wales_N
BK_Hungary_Sfu1
Hungary_ECA
Iberia_LN
Hungary_EN
LBK_EN
Czech_MN
Anatolia_Neolithic
Hungary_MN
Hungary_LN
ALPc_EN
Hungary_LCA
LBKT_EN
Iberia_CA
Beaker_Iberia
0.264 0.268 0.272 0.276
ROS82


Wales_N
Iberia_CA
England_N
Iberia_EN
Scotland_N
Germany_MN
Hungary_MCA
LBKT_EN
Hungary_LN
Hungary_ECA
Czech_MN
Iberia_MN
LBK_EN
Hungary_LCA
Hungary_MN
0.265 0.270 0.275
BUCH2
Iberia_MN
Iberia_EN
Iberia_CA
Scotland_N
England_N
BK_Hungary_Sfu1
Hungary_ECA
Germany_MN
Czech_MN
Hungary_LCA
Hungary_LN
LBK_EN
Beaker_Iberia
Anatolia_Neolithic
ALPc_EN
0.2625 0.2650 0.2675 0.2700 0.2725
CRE20D
Wales_MBA
Iberia_MN
Iberia_EN
ALPc_EN
Germany_MN
Hungary_MN
England_N
Iberia_CA
Scotland_N
Iberia_LN
Czech_MN
Anatolia_Neolithic
LBK_EN
Loschbour
BK_Hungary_Sfu1
0.26 0.27 0.28 0.29 0.30 0.31
Pir4
Wales_N
Iberia_EN
Czech_MN
Scotland_N
Hungary_ECA
Iberia_CA
Hungary_EN
England_N
Loschbour
Iberia_MN
ALPc_EN
LBK_EN
Hungary_LN
LBKT_EN
Hungary_MCA
0.2675 0.2700 0.2725 0.2750
BERG02-2
Wales_N
BK_Hungary_Sfu1
Iberia_MN
Germany_MN
Scotland_N
Iberia_EN
Hungary_MN
England_N
Hungary_ECA
Iberia_CA
LBK_EN
ALPc_EN
Hungary_LN
Czech_MN
Hungary_LCA
0.270 0.275
BERG157-2
Germany_MN
Iberia_CA
Wales_N
Iberia_LN
Hungary_LN
Czech_MN
Iberia_EN
Scotland_N
Hungary_EN
LBK_EN
Hungary_ECA
Iberia_MN
England_N
Hungary_LCA
Anatolia_Neolithic
0.2675 0.2700 0.2725 0.2750
BERG157-7
Germany_MN
England_N
Iberia_LN
Iberia_EN
Iberia_MN
Hungary_ECA
ALPc_EN
Scotland_N
LBK_EN
Czech_MN
Hungary_LN
Wales_N
Iberia_CA
Hungary_MN
Anatolia_Neolithic
0.265 0.270 0.275
BLP10
Scotland_N
England_N
Germany_MN
Wales_N
Hungary_ECA
Loschbour
Iberia_LN
Iberia_CA
Iberia_MN
Central_MN
Iberia_EN
LBK_EN
Wales_MBA
LaBrana
Hungary_MCA
0.2625 0.2650 0.2675 0.2700 0.2725
Es97-1
England_N
Iberia_LN
Wales_N
Hungary_LN
Scotland_N
Germany_MN
Iberia_CA
Iberia_EN
LBK_EN
ALPc_EN
BK_Hungary_Sfu1
Iberia_MN
Hungary_LCA
Hungary_EN
LBKT_EN
0.26 0.27 0.28 0.29 0.30 0.31
PSS4693
Wales_MBA
Wales_CA_EBA
KO1
BK_Poland_Sam
Netherlands_BA
Loschbour
EHG
BK_Switzerland_Sio
Beaker_Netherlands
Yamnaya_Samara
Scotland_LBA
Karelia_HG
Poltavka
Motala_HG
Srubnaya
0.2675 0.2700 0.2725 0.2750 0.27
CBV95
Hungary_ECA
Germany_MN
Iberia_MN
England_N
Loschbour
KO1
Scotland_N
Iberia_CA
Beaker_Iberia
Wales_N
LaBrana
LBKT_EN
Hungary_LN
Czech_MN
ALPc_EN
0.270 0.275
PEI10
Wales_MBA
KO1
Central_MN
Loschbour
BK_Poland_Sam
Germany_EBA
Potapovka
Germany_LN
BK_Switzerland_Sio
Hungary_BA
Iberia_BA
Germany_MN
Motala_HG
Iberia_MN
Beaker_Northern_italy
0.270 0.275
BIS130

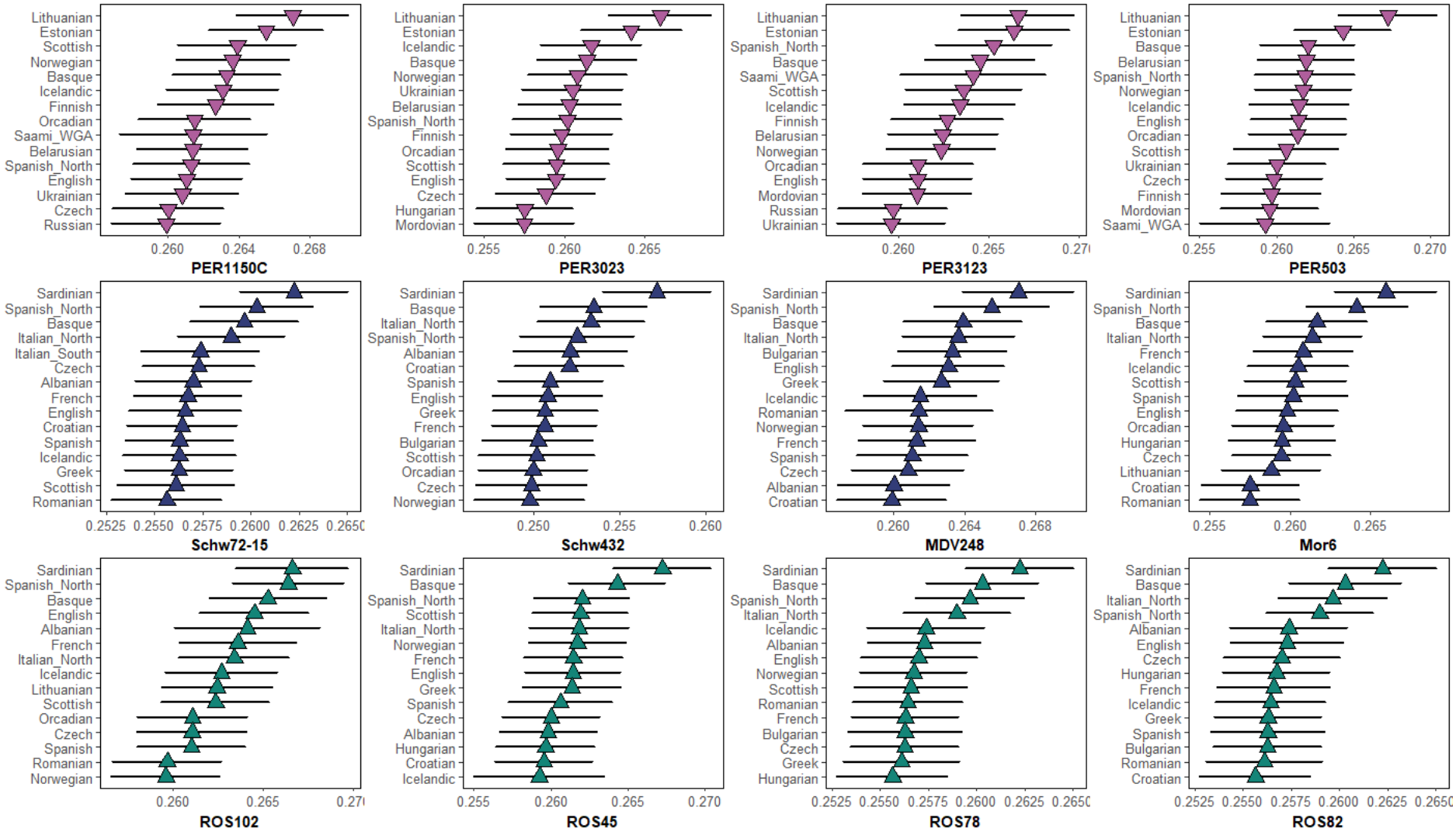

KO1
Wales_MBA
Wales_CA_EBA
Motala_HG
Central_MN
Karelia_HG
Loschbour
BK_Hungary_Sfu2
Germany_LN
Scotland_MBA
EHG
BK_Poland_POL
Beaker_Germany
England_MBA
Germany_EBA
0.26
0.27
0.28
BIS385
Loschbour
Iberia_LN
EHG
Beaker_Iberia
England_N
BK_Hungary_Sfu1
Germany_MN
Central_MN
Hungary_MN
Iberia_MN
LaBrana
BK_Hungary_Sfu3
Hungary_MCA
Iberia_CA
Scotland_N
0.260
0.265
0.270
0.275
EUG11
Loschbour
KO1
Potapovka
LaBrana
Central_MN
Germany_MN
BK_Poland_Sam
Iberia_MN
Wales_CA_EBA
Iberia_EN
BK_Hungary_Sfu1
Beaker_Iberia
LBKT_EN
ALPc_EN
Scotland_N
0.260
0.265
0.270
0.275
NIED
Loschbour
KO1
Wales_MBA
Motala_HG
Wales_CA_EBA
EHG
Central_MN
BK_Hungary_Sfu3
Iberia_LN
Samara_HG
Srubnaya
Karelia_HG
BK_Poland_POL
Hungary_ECA
BK_Czech_CZE
0.255
0.260
0.265
0.270
0.275
OBE3626-1
KO1
Loschbour
BK_Germany_Wor
Wales_MBA
BK_Switzerland_Sio
BK_Hungary_Sfu3
EHG
Central_MN
BK_Poland_Sam
BK_Hungary_HUN2
Germany_LN
Wales_N
Beaker_Iberia
Motala_HG
England_LBA
0.250
0.255
0.260
0.265
0.270
OBE3722
Wales_MBA
BK_Switzerland_Sio
BK_Germany_Wor
Iberia_MN
EHG
Wales_N
Czech_MN
KO1
England_LBA
BK_Hungary_Sfu1
Iberia_CA
Scotland_N
Germany_MN
Hungary_LCA
Hungary_EN
0.25
0.26
0.27
0.28
0.29
PIR3037AB
Wales_MBA
Loschbour
BK_Hungary_Sfu3
KO1
Wales_N
LaBrana
Central_MN
Potapovka
Scotland_N
Beaker_Northern_italy
Motala_HG
Iberia_CA
Iberia_MN
EHG
Iberia_EN
0.26
0.27
0.28
PIR3116B
Iberia_BA
Wales_MBA
Central_MN
England_LBA
Iberia_LN
Beaker_Iberia
BK_Switzerland_Sio
Loschbour
Iberia_CA
Wales_CA_EBA
Beaker_Netherlands
EHG
ALPc_EN
Czech_MN
BK_Hungary_HUN2
0.260
0.265
0.270
0.275
Pir6
Loschbour
Central_MN
BK_Hungary_Sfu1
KO1
Iberia_BA
Netherlands_BA
England_N
Germany_MN
Beaker_Iberia
Iberia_CA
BK_Hungary_Sfu3
Beaker_Northern_italy
Iberia_MN
Scotland_N
Motala_HG
0.260
0.265
0.270
0.275
PSS4170
Wales_MBA
KO1
Iberia_MN
England_N
Germany_MN
Iberia_CA
Central_MN
Beaker_Iberia
Loschbour
Scotland_N
BK_Hungary_Sfu1
BK_Switzerland_Sio
Iberia_EN
LBKT_EN
Czech_MN
0.255
0.260
0.265
0.270
0.275
QUIN234
Wales_N
Loschbour
Iberia_CA
KO1
England_N
BK_Hungary_Sfu1
Iberia_EN
Scotland_N
Germany_MN
Central_MN
Beaker_Northern_italy
Beaker_Iberia
Iberia_LN
LaBrana
BK_Hungary_HUN2
0.250
0.255
0.260
0.265
0.270
QUIN58
Wales_MBA
KO1
Loschbour
Central_MN
BK_Hungary_Sfu1
BK_Hungary_Sfu2
BK_Poland_Sam
Iberia_MN
Germany_MN
Iberia_CA
LaBrana
Beaker_Iberia
England_N
Hungary_LN
Germany_EBA
0.250
0.255
0.260
0.265
0.270
RIX15

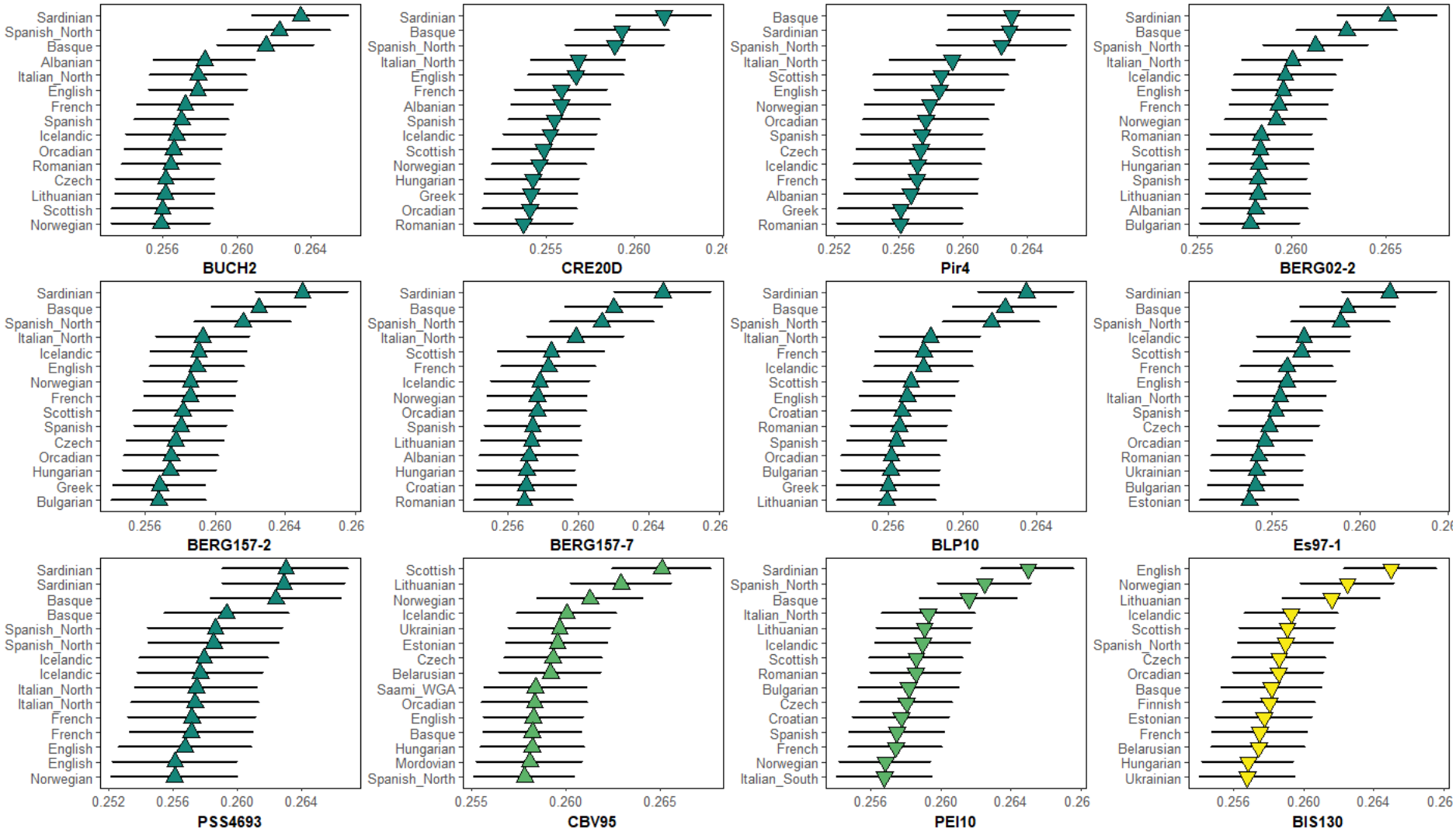

Loschbour
Central_MN
Iberia_MN
Motala_HG
Beaker_Northern_italy
Scotland_N
Germany_MN
Wales_CA_EBA
BK_Hungary_Sfu1
Scotland_CA_EBA
BK_Hungary_Sfu3
Czech_MN
England_N
Corded_Germany
Hungary_ECA
0.260 0.265 0.270 0.275
RIX2
Iberia_LN
Central_MN
KO1
Germany_MN
Iberia_MN
Loschbour
BK_Hungary_Sfu1
Wales_CA_EBA
BK_Poland_Sam
Hungary_ECA
Motala_HG
Scotland_N
Netherlands_BA
BK_Hungary_HUN2
Germany_LN
0.260 0.265 0.270 0.275
RIX4
Wales_N
Iberia_BA
Wales_MBA
EHG
Loschbour
Central_MN
BK_Hungary_Sfu1
Beaker_Northern_italy
Iberia_MN
Germany_EBA
Beaker_Iberia
Potapovka
Scotland_N
BK_Hungary_Sfu3
Beaker_Netherlands
0.26 0.27 0.28
ATT26
KO1
Wales_N
Central_MN
Wales_MBA
Iberia_LN
BK_Hungary_Sfu1
Beaker_Iberia
Loschbour
Iberia_MN
Beaker_Northern_italy
Motala_HG
Beaker_Netherlands
Scotland_N
BK_Hungary_Sfu2
Germany_MN
0.255 0.260 0.265 0.270 0.275
BFM265
BK_Hungary_Sfu3
KO1
Central_MN
BK_Hungary_Sfu1
Germany_LN
England_LBA
Loschbour
Wales_N
BK_Poland_Sam
BK_Hungary_Sfu2
Iberia_LN
BK_Hungary_HUN2
England_MBA
BK_Switzerland_Sio
Hungary_MCA
0.26 0.27
COL11
BK_Poland_Sam
Iberia_LN
KO1
Central_MN
Loschbour
Germany_MN
Wales_N
Wales_CA_EBA
BK_Hungary_Sfu1
BK_Hungary_Sfu2
Potapovka
Iberia_MN
Scotland_N
BK_Germany_Wor
Iberia_CA
0.255 0.260 0.265 0.270
COL153A
Wales_MBA
Hungary_ECA
BK_Hungary_Sfu1
Wales_CA_EBA
Hungary_BA
KO1
BK_Germany_Wor
England_N
Central_MN
BK_Switzerland_Sio
Scotland_N
Wales_N
Loschbour
Czech_MN
Hungary_MN
0.260 0.265 0.270 0.275
COL153i
Wales_MBA
Central_MN
Loschbour
Hungary_ECA
Potapovka
KO1
BK_Germany_Wor
Germany_MN
England_N
Corded_Germany
Iberia_LN
Beaker_Netherlands
BK_Hungary_Sfu1
EHG
LaBrana
0.260 0.265 0.270 0.275
ERS1164
Wales_MBA
BK_Switzerland_Sio
Loschbour
Wales_CA_EBA
KO1
Iberia_LN
Beaker_Northern_italy
BK_Poland_Sam
Netherlands_BA
Hungary_ECA
Germany_MN
England_N
BK_Hungary_Sfu1
Scotland_N
Motala_HG
0.26 0.27 0.28
ERS86
BK_Hungary_Sfu1
Wales_CA_EBA
Czech_MN
EHG
ALPc_EN
BK_Hungary_HUN
Beaker_Iberia
Hungary_ECA
KO1
Iberia_MN
England_N
Hungary_LCA
Iberia_BA
BK_Hungary_Sfu3
Beaker_Netherlands
0.255 0.260 0.265 0.270 0.275
ERS88
KO1
Wales_MBA
Hungary_ECA
BK_Germany_Wor
England_LBA
Wales_CA_EBA
Wales_N
Czech_MN
Iberia_MN
Netherlands_BA
BK_Hungary_Sfu3
Central_MN
ALPc_EN
Beaker_Iberia
Scotland_N
0.26 0.27
Jeb8
Hungary_MN
Wales_CA_EBA
England_N
KO1
Loschbour
Wales_N
Central_MN
Hungary_LN
Beaker_Northern_italy
Hungary_ECA
Iberia_LN
BK_Poland_Sam
BK_Hungary_Sfu1
Iberia_MN
Germany_MN
0.26 0.27
NOR2B6

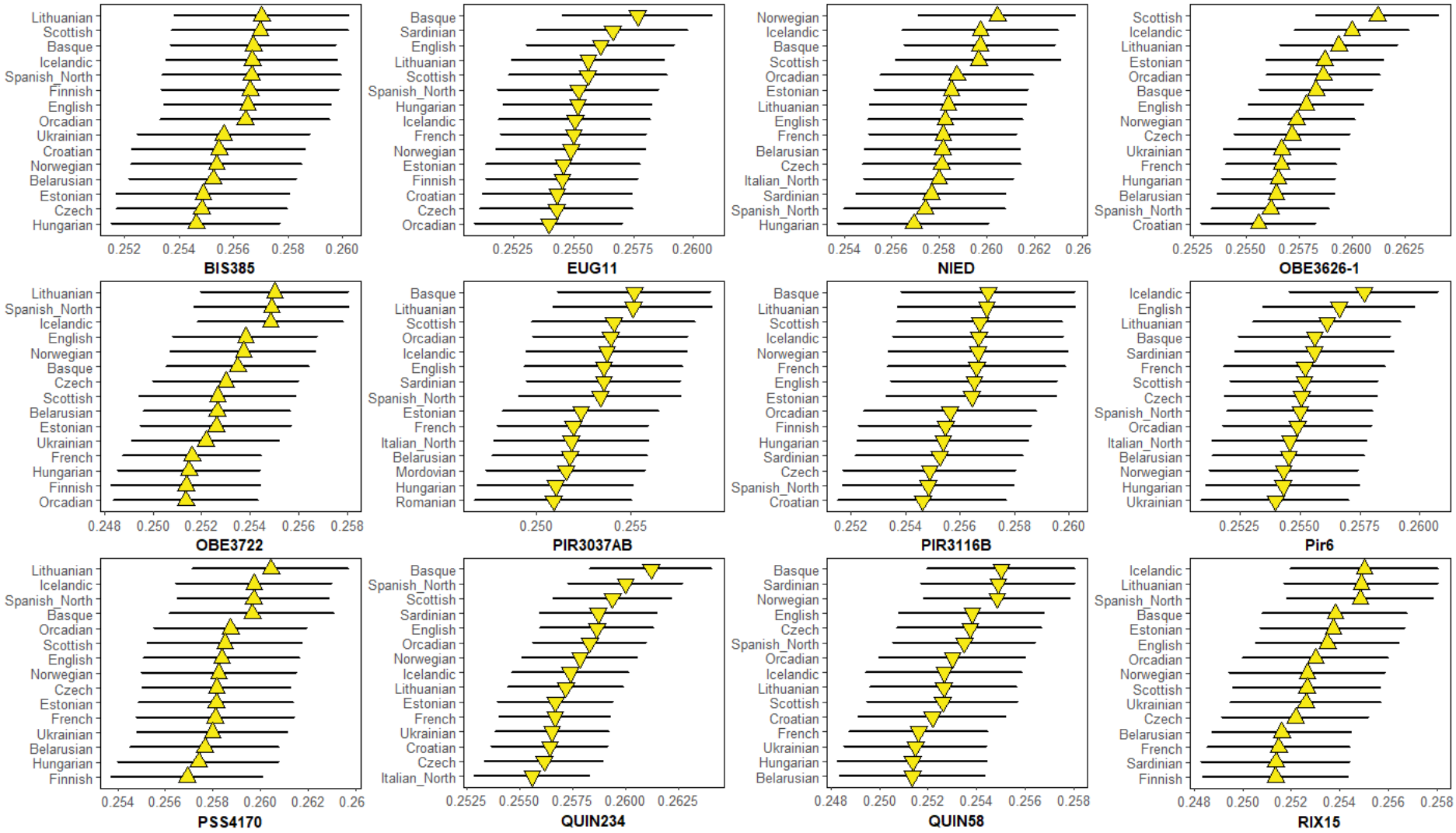

BK_Switzerland_Sio
Wales_CA_EBA
Iberia_LN
Iberia_MN
Central_MN
Iberia_BA
BK_Poland_Sam
Germany_LN
KO1
Wales_N
Hungary_MN
Germany_MN
England_LBA
England_N
BK_Hungary_HUN
0.255 0.260 0.265 0.270 0.275 0.280
NOR3-15
Wales_MBA
Wales_CA_EBA
Loschbour
KO1
Beaker_Netherlands
Central_MN
Iberia_MN
Potapovka
Scotland_LBA
Hungary_MCA
Corded_Germany
Iberia_CA
Scotland_N
England_MBA
England_N
0.260 0.265 0.270 0.275 0.280
NOR3-6
Iberia_LN
Wales_N
Germany_MN
Wales_MBA
BK_Switzerland_Sio
Wales_CA_EBA
Potapovka
Beaker_Iberia
Loschbour
Scotland_N
Hungary_ECA
Hungary_LN
Beaker_Germany
Iberia_CA
Hungary_MN
0.25 0.26 0.27
NOR4
Iberia_LN
Germany_MN
Hungary_LN
Wales_N
Scotland_N
Iberia_MN
Loschbour
ALPc_EN
Hungary_MCA
Czech_MN
England_N
Iberia_CA
Hungary_LCA
Hungary_EN
Iberia_EN
0.260 0.265 0.270
WET370-1
Iberia_LN
Central_MN
Loschbour
BK_Hungary_Sfu1
Wales_N
KO1
BK_Germany_Wor
Hungary_ECA
Scotland_LBA
BK_Hungary_HUN2
Netherlands_BA
Beaker_Northern_Italy
Scotland_N
Scotland_CA_EBA
Potapovka
0.260 0.265 0.270
BES1248
Iberia_LN
Germany_MN
England_N
Czech_MN
Hungary_LN
Hungary_MN
Loschbour
LaBrana
Scotland_N
Motala_HG
Iberia_CA
Iberia_EN
BK_Hungary_Sfu2
KO1
Iberia_MN
0.26 0.27 0.28 0.29
PECH5
Iberia_LN
Wales_CA_EBA
Germany_MN
Scotland_N
Iberia_MN
Iberia_CA
England_N
Central_MN
Beaker_Northern_Italy
BK_Hungary_Sfu1
Loschbour
Iberia_BA
Wales_N
BK_Switzerland_Sio
Germany_LN
0.260 0.265 0.270 0.275 0.280
PECH8
EHG
BK_Hungary_Sfu1
KO1
Wales_N
Loschbour
Central_MN
Hungary_ECA
Iberia_CA
Iberia_BA
Iberia_MN
Germany_LN
Scotland_N
BK_Hungary_Sfu3
Wales_CA_EBA
LaBrana
0.255 0.260 0.265 0.270 0.275 0.280
PEY163
Wales_MBA
England_LBA
Potapovka
Netherlands_BA
BK_Hungary_Sfu3
England_N
Iberia_MN
Scotland_LBA
Hungary_MCA
KO1
Central_MN
Iberia_CA
Motala_HG
BK_Poland_POL
England_CA_EBA
0.260 0.265 0.270 0.275 0.280
PEY53
EHG
Loschbour
Wales_MBA
Central_MN
KO1
Iberia_BA
Iberia_CA
England_N
BK_Hungary_HUN
Scotland_N
Scotland_LBA
Scotland_MBA
Iberia_MN
Beaker_Iberia
Czech_MN
0.25 0.26 0.27
PT2

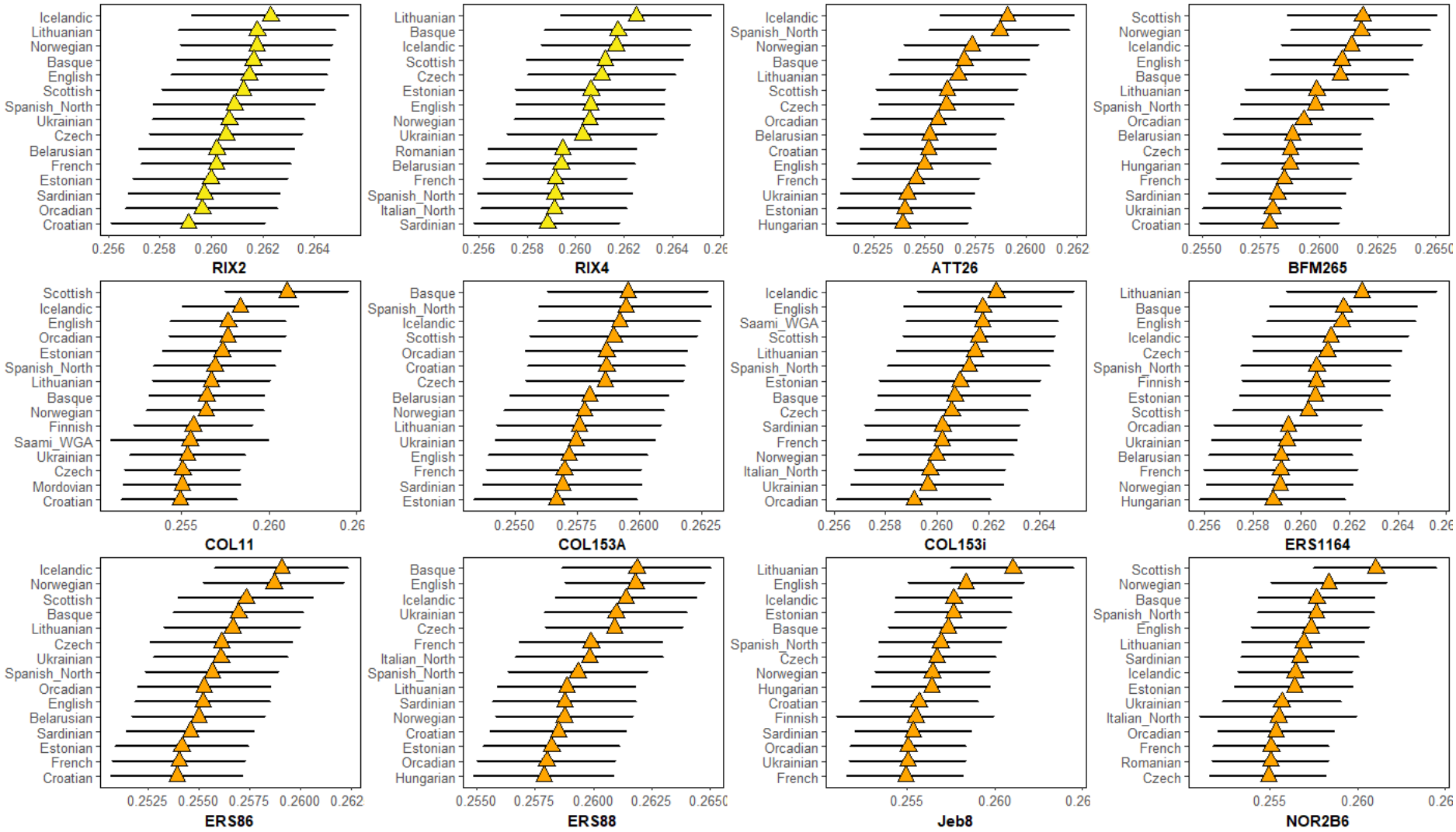
Supplementary Figure 5: Values of f3(Mbuti; ancient French individual, ancient population) for all 58 newly reported individuals.

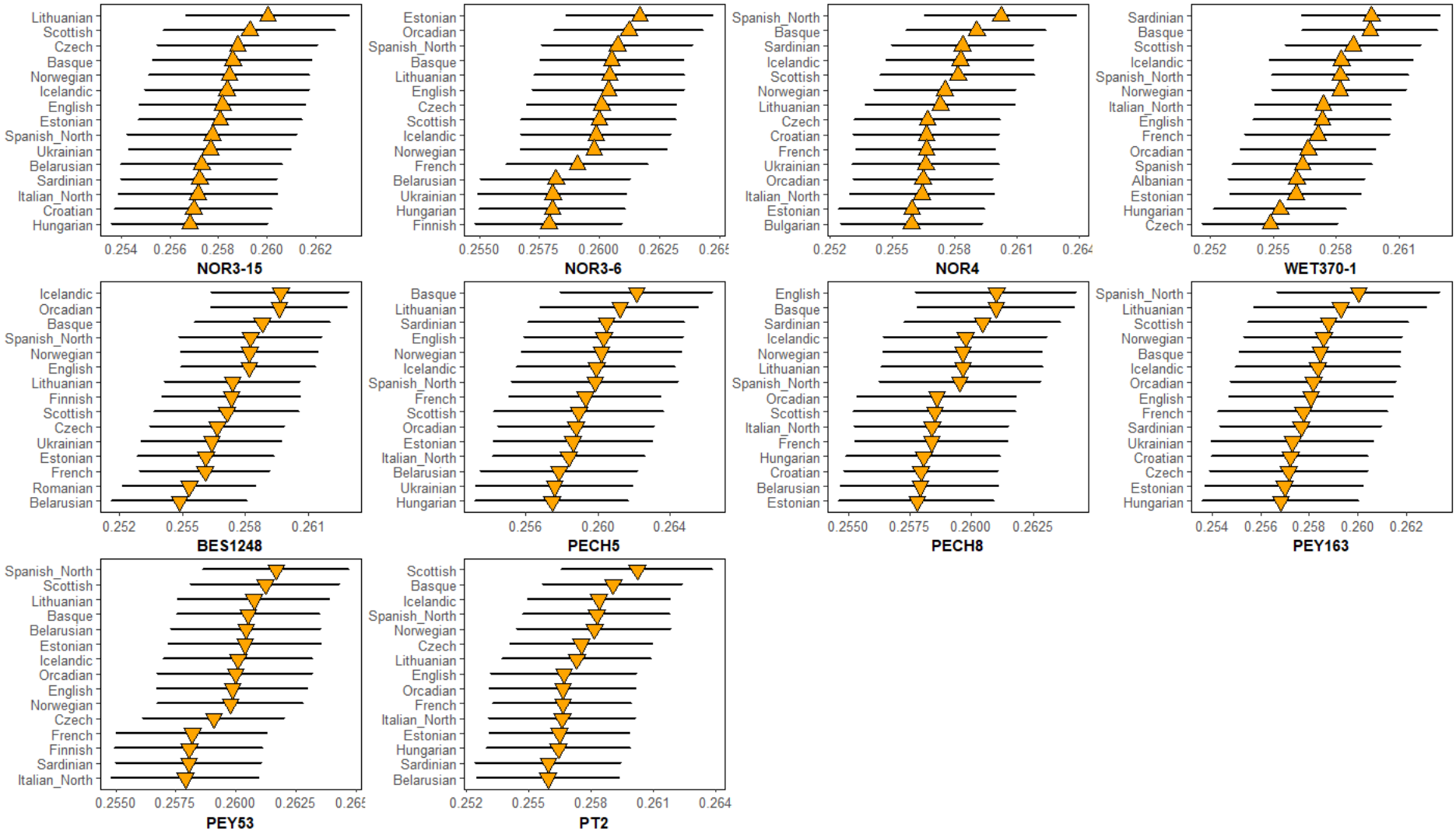
The 15 highest values are shown.


Lithuanian
Estonian
Scottish
Norwegian
Basque
Icelandic
Finnish
Orcadian
Saami_WGA
Belarusian
Spanish_North
English
Ukrainian
Czech
Russian
0.260
0.264
0.268
PER1150C
Lithuanian
Estonian
Icelandic
Basque
Norwegian
Ukrainian
Belarusian
Spanish_North
Finnish
Orcadian
Scottish
English
Czech
Hungarian
Mordovian
0.255
0.260
0.265
PER3023
Lithuanian
Estonian
Spanish_North
Basque
Saami_WGA
Scottish
Icelandic
Finnish
Belarusian
Norwegian
Orcadian
English
Mordovian
Russian
Ukrainian
0.260
0.265
PER3123
Lithuanian
Estonian
Basque
Belarusian
Spanish_North
Norwegian
Icelandic
English
Orcadian
Scottish
Ukrainian
Czech
Finnish
Mordovian
Saami_WGA
0.255
0.260
0.265
0.270
PER503
Sardinian
Spanish_North
Basque
Italian_North
Italian_South
Czech
Albanian
French
English
Croatian
Spanish
Icelandic
Greek
Scottish
Romanian
0.2525
0.2550
0.2575
0.2600
0.2625
Schw72-15
Sardinian
Basque
Italian_North
Spanish_North
Albanian
Croatian
Spanish
English
Greek
French
Bulgarian
Scottish
Orcadian
Czech
Norwegian
0.250
0.255
0.260
Schw432
Sardinian
Spanish_North
Basque
Italian_North
Bulgarian
English
Greek
Icelandic
Romanian
Norwegian
French
Spanish
Czech
Albanian
Croatian
0.260
0.264
0.268
MDV248
Sardinian
Spanish_North
Basque
Italian_North
French
Icelandic
Scottish
Spanish
English
Orcadian
Hungarian
Czech
Lithuanian
Croatian
Romanian
0.255
0.260
0.265
Mor6
Sardinian
Spanish_North
Basque
English
Albanian
French
Italian_North
Icelandic
Lithuanian
Scottish
Orcadian
Czech
Spanish
Romanian
Norwegian
0.260
0.265
ROS102
Sardinian
Basque
Spanish_North
Scottish
Italian_North
Norwegian
French
English
Greek
Spanish
Czech
Albanian
Hungarian
Croatian
Icelandic
0.255
0.260
0.265
0.270
ROS45
Sardinian
Basque
Spanish_North
Italian_North
Icelandic
Albanian
English
Norwegian
Scottish
Romanian
French
Bulgarian
Czech
Greek
Hungarian
0.2525
0.2550
0.2575
0.2600
0.2625
ROS78
Sardinian
Basque
Italian_North
Spanish_North
Albanian
English
Czech
Hungarian
French
Icelandic
Greek
Spanish
Bulgarian
Romanian
Croatian
0.2525
0.2550
0.2575
0.2600
0.2625
ROS82


Sardinian
Spanish_North
Basque
Albanian
Italian_North
English
French
Spanish
Icelandic
Orcadian
Romanian
Czech
Lithuanian
Scottish
Norwegian
0.256
0.260
0.264
BUCH2
Sardinian
Basque
Spanish_North
Italian_North
English
French
Albanian
Spanish
Icelandic
Scottish
Norwegian
Hungarian
Greek
Orcadian
Romanian
0.255
0.260
CRE20D
Basque
Sardinian
Spanish_North
Italian_North
Scottish
English
Norwegian
Orcadian
Spanish
Czech
Icelandic
French
Albanian
Greek
Romanian
0.252
0.256
0.260
0.264
Pir4
Sardinian
Basque
Spanish_North
Italian_North
Icelandic
English
French
Norwegian
Romanian
Scottish
Hungarian
Spanish
Lithuanian
Albanian
Bulgarian
0.255
0.260
0.265
BERG02-2
Sardinian
Basque
Spanish_North
Italian_North
Icelandic
English
French
Norwegian
Scottish
Spanish
Czech
Orcadian
Hungarian
Greek
Bulgarian
0.256
0.260
0.264
BERG157-2
Sardinian
Basque
Spanish_North
Italian_North
Scottish
French
Icelandic
Norwegian
Orcadian
Spanish
Lithuanian
Albanian
Hungarian
Croatian
Romanian
0.256
0.260
0.264
BERG157-7
Sardinian
Basque
Spanish_North
Italian_North
French
Icelandic
Scottish
English
Croatian
Romanian
Spanish
Orcadian
Bulgarian
Greek
Lithuanian
0.256
0.260
0.264
BLP10
Sardinian
Basque
Spanish_North
Icelandic
Scottish
French
English
Italian_North
Spanish
Czech
Orcadian
Romanian
Ukrainian
Bulgarian
Estonian
0.255
0.260
Es97-1
Sardinian
Sardinian
Basque
Basque
Spanish_North
Spanish_North
Icelandic
Icelandic
Italian_North
Italian_North
French
French
English
English
Norwegian
0.252
0.256
0.260
0.264
PSS4693
Scottish
Lithuanian
Norwegian
Icelandic
Ukrainian
Estonian
Czech
Belarusian
Saami_WGA
Orcadian
English
Basque
Hungarian
Mordovian
Spanish_North
0.255
0.260
0.265
CBV95
Sardinian
Spanish_North
Basque
Italian_North
Lithuanian
Icelandic
Scottish
Romanian
Bulgarian
Czech
Croatian
Spanish
French
Norwegian
Italian_South
0.256
0.260
0.264
PEI10
English
Norwegian
Lithuanian
Icelandic
Scottish
Spanish_North
Czech
Orcadian
Basque
Finnish
Estonian
French
Belarusian
Hungarian
Ukrainian
0.256
0.260
0.264
BIS130


Lithuanian
Scottish
Basque
Icelandic
Spanish_North
Finnish
English
Orcadian
Ukrainian
Croatian
Norwegian
Belarusian
Estonian
Czech
Hungarian
0.252 0.254 0.256 0.258 0.260
BIS385
Basque
Sardinian
English
Lithuanian
Scottish
Spanish_North
Hungarian
Icelandic
French
Norwegian
Estonian
Finnish
Croatian
Czech
Orcadian
0.2525 0.2550 0.2575 0.2600
EUG11
Norwegian
Icelandic
Basque
Scottish
Orcadian
Estonian
Lithuanian
English
French
Belarusian
Czech
Italian_North
Sardinian
Spanish_North
Hungarian
0.254 0.256 0.258 0.260 0.262 0.26
NIED
Scottish
Icelandic
Lithuanian
Estonian
Orcadian
Basque
English
Norwegian
Czech
Ukrainian
French
Hungarian
Belarusian
Spanish_North
Croatian
0.2525 0.2550 0.2575 0.2600 0.2625
OBE3626-1
Lithuanian
Spanish_North
Icelandic
English
Norwegian
Basque
Czech
Scottish
Belarusian
Estonian
Ukrainian
French
Hungarian
Finnish
Orcadian
0.248 0.250 0.252 0.254 0.256 0.258
OBE3722
Basque
Lithuanian
Scottish
Orcadian
Icelandic
English
Sardinian
Spanish_North
Estonian
French
Italian_North
Belarusian
Mordovian
Hungarian
Romanian
0.250 0.255
PIR3037AB
Basque
Lithuanian
Scottish
Icelandic
Norwegian
French
English
Estonian
Orcadian
Finnish
Hungarian
Sardinian
Czech
Spanish_North
Croatian
0.252 0.254 0.256 0.258 0.260
PIR3116B
Icelandic
English
Lithuanian
Basque
Sardinian
French
Scottish
Czech
Spanish_North
Orcadian
Italian_North
Belarusian
Norwegian
Hungarian
Ukrainian
0.2525 0.2550 0.2575 0.2600
Pir6
Lithuanian
Icelandic
Spanish_North
Basque
Orcadian
Scottish
English
Norwegian
Czech
Estonian
French
Ukrainian
Belarusian
Hungarian
Finnish
0.254 0.256 0.258 0.260 0.262 0.26
PSS4170
Basque
Spanish_North
Scottish
Sardinian
English
Orcadian
Norwegian
Icelandic
Lithuanian
Estonian
French
Ukrainian
Croatian
Czech
Italian_North
0.2525 0.2550 0.2575 0.2600 0.2625
QUIN234
Basque
Sardinian
Norwegian
English
Czech
Spanish_North
Orcadian
Icelandic
Lithuanian
Scottish
Croatian
French
Ukrainian
Hungarian
Belarusian
0.248 0.250 0.252 0.254 0.256 0.258
QUIN58
Icelandic
Lithuanian
Spanish_North
Basque
Estonian
English
Orcadian
Norwegian
Scottish
Ukrainian
Czech
Belarusian
French
Sardinian
Finnish
0.248 0.250 0.252 0.254 0.256 0.258
RIX15


Icelandic
Lithuanian
Norwegian
Basque
English
Scottish
Spanish_North
Ukrainian
Czech
Belarusian
French
Estonian
Sardinian
Orcadian
Croatian
0.256 0.258 0.260 0.262 0.264
RIX2
Lithuanian
Basque
Icelandic
Scottish
Czech
Estonian
English
Norwegian
Ukrainian
Romanian
Belarusian
French
Spanish_North
Italian_North
Sardinian
0.256 0.258 0.260 0.262 0.264 0.26
RIX4
Icelandic
Spanish_North
Norwegian
Basque
Lithuanian
Scottish
Czech
Orcadian
Belarusian
Croatian
English
French
Ukrainian
Estonian
Hungarian
0.2525 0.2550 0.2575 0.2600 0.262
ATT26
Scottish
Norwegian
Icelandic
English
Basque
Lithuanian
Spanish_North
Orcadian
Belarusian
Czech
Hungarian
French
Sardinian
Ukrainian
Croatian
0.2550 0.2575 0.2600 0.2625 0.265
BFM265
Scottish
Icelandic
English
Orcadian
Estonian
Spanish_North
Lithuanian
Basque
Norwegian
Finnish
Saami_WGA
Ukrainian
Czech
Mordovian
Croatian
0.255 0.260 0.26
COL11
Basque
Spanish_North
Icelandic
Scottish
Orcadian
Croatian
Czech
Belarusian
Norwegian
Lithuanian
Ukrainian
English
French
Sardinian
Estonian
0.2550 0.2575 0.2600 0.2625
COL153A
Icelandic
English
Saami_WGA
Scottish
Lithuanian
Spanish_North
Estonian
Basque
Czech
Sardinian
French
Norwegian
Italian_North
Ukrainian
Orcadian
0.256 0.258 0.260 0.262 0.264
COL153i
Lithuanian
Basque
English
Icelandic
Czech
Spanish_North
Finnish
Estonian
Scottish
Orcadian
Ukrainian
Belarusian
French
Norwegian
Hungarian
0.256 0.258 0.260 0.262 0.264 0.26
ERS1164
Icelandic
Norwegian
Scottish
Basque
Lithuanian
Czech
Ukrainian
Spanish_North
Orcadian
English
Belarusian
Sardinian
Estonian
French
Croatian
0.2525 0.2550 0.2575 0.2600 0.262
ERS86
Basque
English
Icelandic
Ukrainian
Czech
French
Italian_North
Spanish_North
Lithuanian
Sardinian
Norwegian
Croatian
Estonian
Orcadian
Hungarian
0.2550 0.2575 0.2600 0.2625 0.265
ERS88
Lithuanian
English
Icelandic
Estonian
Basque
Spanish_North
Czech
Norwegian
Hungarian
Croatian
Finnish
Sardinian
Orcadian
Ukrainian
French
0.255 0.260 0.26
Jeb8
Scottish
Norwegian
Basque
Spanish_North
English
Lithuanian
Sardinian
Icelandic
Estonian
Ukrainian
Italian_North
Orcadian
French
Romanian
Czech
0.255 0.260 0.26
NOR2B6


Lithuanian
Scottish
Czech
Basque
Norwegian
Icelandic
English
Estonian
Spanish_North
Ukrainian
Belarusian
Sardinian
Italian_North
Croatian
Hungarian
0.254 0.256 0.258 0.260 0.262
NOR3-15
Estonian
Orcadian
Spanish_North
Basque
Lithuanian
English
Czech
Scottish
Icelandic
Norwegian
French
Belarusian
Ukrainian
Hungarian
Finnish
0.2550 0.2575 0.2600 0.2625
NOR3-6
Spanish_North
Basque
Sardinian
Icelandic
Scottish
Norwegian
Lithuanian
Czech
Croatian
French
Ukrainian
Orcadian
Italian_North
Estonian
Bulgarian
0.252 0.255 0.258 0.261
NOR4
Sardinian
Basque
Scottish
Icelandic
Spanish_North
Norwegian
Italian_North
English
French
Orcadian
Spanish
Albanian
Estonian
Hungarian
Czech
0.252 0.255 0.258 0.261
WET370-1
Icelandic
Orcadian
Basque
Spanish_North
Norwegian
English
Lithuanian
Finnish
Scottish
Czech
Ukrainian
Estonian
French
Romanian
Belarusian
0.252 0.255 0.258 0.261
BES1248
Basque
Lithuanian
Sardinian
English
Norwegian
Icelandic
Spanish_North
French
Scottish
Orcadian
Estonian
Italian_North
Belarusian
Ukrainian
Hungarian
0.256 0.260 0.264
PECH5
English
Basque
Sardinian
Icelandic
Norwegian
Lithuanian
Spanish_North
Orcadian
Scottish
Italian_North
French
Hungarian
Croatian
Belarusian
Estonian
0.2550 0.2575 0.2600 0.2625
PECH8
Spanish_North
Lithuanian
Scottish
Norwegian
Basque
Icelandic
Orcadian
English
French
Sardinian
Ukrainian
Croatian
Czech
Estonian
Hungarian
0.254 0.256 0.258 0.260 0.262
PEY163
Spanish_North
Scottish
Lithuanian
Basque
Belarusian
Estonian
Icelandic
Orcadian
English
Norwegian
Czech
French
Finnish
Sardinian
Italian_North
0.2550 0.2575 0.2600 0.2625
PEY53
Scottish
Basque
Icelandic
Spanish_North
Norwegian
Czech
Lithuanian
English
Orcadian
French
Italian_North
Estonian
Hungarian
Sardinian
Belarusian
0.252 0.255 0.258 0.261 0.264
PT2

Supplementary Figure 6: Values of f3(Mbuti; ancient French individual, modern Western Eurasian population) for all 58 newly reported individuals.
The 15 highest values are shown.

Supplementary Table 1: Summary of all targeted loci and the oligonucleotides used for their enrichment

Locus	ID	Position on GRCh37	Forward primer	Reverse primer	Amplicon Size (bp)
mtDNA	Hs-800-2500	-	CACCCCCACGGGAAACAG	TACCGCGGCCGTTAAACA	1777
mtDNA	Hs-2400-4200	-	ACACAGGCATGCTCATAAGGAA	AGATTGTAATGGGTATGGAGA	1811
mtDNA	Hs-4100-5700	-	CCGATTCCGCTACGACCAAC	CGGGAGAAGTAGATTGAAG	1599
mtDNA	Hs-5400-7100	-	TTACCACGCTACTCCTACCTATCT	AAGCCTCCTATGATGGCAAATAC	1618
mtDNA	Hs-6800-8500	-	TCACCTCCGCTACCATAATCATC	CAATGAATGAAGCGAAC	1741
mtDNA	Hs-8200-9900F	-	CCCCTCTAGAGCCCACTGTAAA	ATGAAGCAGATAGTGAGGAAA	1584
mtDNA	Hs-9800-11400F	-	ACGGCATCTACGGCTCAAC	AAGTGGAGTCCGTAAAGAGGTATC	1626
mtDNA	Hs-11300-12800	-	CTCACTGCCCAAGAACTATCAAAC	TCTGCTCGGGCGTATCATC	1525
mtDNA	Hs-12700-14200	-	CCGCTAACAACCTATTCCAACT	TCAGGGTTGATTCGGGAGGAT	1547
mtDNA	Hs-14200-15700	-	CCAACAAACAATGTTCAACCAG	TTGTCCTCCGATTCAGGTTAGA	1584
mtDNA	Hs-15700-800	-	CCCCATCCTCCATATATCCAA	TTTCCCGTGGGGGTGT	1728
CSMD1	rs10108270	8:4190793	AAAACAGAAGTCACACCCTGCC	GCGCTCGCCCTTTCTGTATC	144
RAAN	rs4891825	18:67867663	AGCCAGACCCTCAATCAAGAC	TGAAGGGAGGGGGAATCTCT	130
LIMGH1	rs10007810	4:41554364	TGAGCTGTGTGAAGAGTCTGT	TTCCTCTCAGATGGATCCTGATG	149
HTBP2	rs4918842	10:115316812	TGTCTGCCTGGTTCTGCCTTACTG	GGTGTGGATGCTTGAATTAATCGG	133
LOC442008	rs1876482	2:17362568	GGCTGTACCCTCACTATTGGT	TGGAGCATAGTGAGCTGTTGA	92
BRF1	rs3784230	14:105679055	GGAGAGGACGCAGGCATTAC	CTGCCTCAGGCCCCCTAC	160
DARC	rs2814778	1:159174683	CCCTCATTAGTCCTTGGCTCTT	TTCCCCATGGCACCGTTTG	98
_	rs2065982	13:34864240	GGGCTCTGTGGGACTCCTTG	AGTGGTCAGTGCCAGTAGGAA	122
_	rs6451722	5:43711378	TGCATTTCCACCCATTCTCTCT	GGTGGCGAATCTGCTCCAAA	127
_	rs730570	14:101142890	GCACTCACCTGCATCTCACAC	GACTGTGAGCACCTTCCTTCC	166
_	rs2714758	15:25479337	AGCCCATGGTGTGCTGAAG	CTCCCCACGCATGCATCTAG	102
OCA2-1	rs1800407	15:28230318	GACAGAGCATGATGATCATGG	CTTGTACTCTCTCTGTGTGTG	82
ASIP-2	rs2378249	20:33218090	AAGCTAACTAAGGGCACAAGTC	ATCTCATTGCTTTTCAGCCCAC	82
DHCR7/NA DSYN1	rs7944926	11:71165625	CTGCTTTCACCTCATTTGCTC	AGTTTTGGCCAAGCCTTCTTC	62
SLC22A4-1	rs1050152	5:131676320	ATGGGTAGTCTGACTGTCCTG	TTTCACTTTCTGCATCTGCTCT	102
SLC22A4-2	rs272872	5:131675864	CTAGTTTGGGCAAAAGCAGAGT	AGGAGAAGGAGGTTAACCAGGA	84
FADS1-FADS2	rs174546	11:61569830	TATGTCCCCAAACCCAACCC	CAGTCTGGAAGAAGAGGAGGAA	106
MCM6	rs41525747	2:136608643	CGCTGGCAATACAGATAAGAT	CAAATGCAACCTAAGGAGGAG	69
MCM6	rs4988235	2:136608646	CGCTGGCAATACAGATAAGAT	CAAATGCAACCTAAGGAGGAG	69
MCM6	rs41380347	2:136608651	CGCTGGCAATACAGATAAGAT	CAAATGCAACCTAAGGAGGAG	69
MCM6	rs145946881	2:136608746	AGGCCTACTAGTACATTGTAGGGTC	GCATTTGAGTGTAGTTGTTAGACGGA	131
PLRP2	rs4751995	10:118397884	GGGTTGTATTTCTTTGGACAGGTT	TCCTGATGTACCCATTCACTTTCT	84
PLRP2	rs4751996	10:118397894	GGGTTGTATTTCTTTGGACAGGTT	TCCTGATGTACCCATTCACTTTCT	84
SLC24A4	rs12896399	14:92773663	CTGGCGATCCAATTCTTTGTTCT	GCCCTGGGTCTTGATGTTGTATT	100
TYR	rs1393350	11:89011046	TGCATATCCACCAACTCCTACTC	TGGGAAGGTGAATGATAACACGAA	104
OCA2	rs1800407	15:28230318	AGGAGCATGGTGGTGACGTT	GAGCCTGCTCACTCTGGCTT	151
ASIP	rs6119471	20:32785212	CCTGGGCAGCAGAGCAAGA	GGCACTTGAGAGGAGGCTAAC	120
BNC2	rs2153271	9:16864521	GCACAGGCAAATCTTGGCAAC	AATGCACAACATGGGGGATAGG	143
KITLG	rs12821256	12:89328335	GTGTGAAGTTGTGTGGCAGAAGT	CATGAGTCATGAGTGCTTTGTTCC	139
TYR	rs1042602	11:88911696	TGTGTCAATGGATGCACTGCT	TTCCCACCGCAACAAGAAGAG	112
EXOC2	rs4959270	6:457748	TGGGGTTTACGATTCAACATGAG	ATGGAAAAGAACCACCATGTCAG	107
HERC2	rs12913832	15:28365618	GGCTCTCTGTGTCTGATCCAA	CGGCCCCTGATGATGATAGC	87
IRF4	rs12203592	6:396321	AACCACAGGGCAGCTGATCT	TATGGCTAAACCTGGCACCAA	123
EDAR	rs3827760	2:109513601	GCCACGTTTTCACAACAGCC	TTGCCTCGAGAAGACTAGCC	71
AR	rs2497938	X:66563018	TACCAAGCCCTCAGAAAATGGTAA	TCCAAATCTGGACGCTCCTTTCCAC	129
AMELO	rs2106416		CACCACCAAATCATCCCC	GGCTGCTGAGCTGGCAC	95
AMEAB	rs9462529		CCCTGGGCTCTGTAAAGAATAG	CTGTAAAATTGGGACCACTTGAGAAAC	78
SLC24A5	rs1426654	15:48426484	TGTTCAGCCCTTGGATTGTCT	AAATCACACTGAGTAAGCAAGAAGT	134
GRM5	rs7119749	11:88515022	ACCTGTTGAAACTGACAAGCC	TCACCCACTGTCATGCAGCAA	102
SLC45A2	rs16891982	5:33951693	TCCAAGTTGTGCTAGACCAGAA	AGGAGTCGAGGTTGGATGTTG	122
MC1R	rs885479	16:89986154	GTGGACCGCTACATCTCCAT	CGATGAAGAGCGTGCTGAAG	124
MC1R	rs1110400	16:89986130	GGACCGCTACATCTCCATCTT	GCGTGCTGAAGACGACACT	113
MC1R	rs11547464	16:89986091	CCTGCAGCTCCATGCTGTC	GGTAGCGCAGTGCGTAGAA	87
MC1R	rs1805006	16:89985918	TGTACTGCTTCATCTGCTGCC	ATCACGTCAATGACATTGTCCAG	151
MC1R	rs1805007	16:89986117	GGACCGCTACATCTCCATCTT	GCGTGCTGAAGACGACACT	113
MC1R	rs1805008	16:89986144	GGACCGCTACATCTCCATCTT	GCGTGCTGAAGACGACACT	113
MC1R	rs1805009	16:89986546	ACCTCTTTCTCGCCCTCATCA	GAGCATGTCAGCACCTCCTT	106
MC1R	rs2228479	16:89985940	TGTACTGCTTCATCTGCTGCC	ATCACGTCAATGACATTGTCCAG	151
M42	rs2032630	Y:21866840	AGATAACTTGTGTCAGTGCAACTT	AAGTCACCAGCTCTCTTTTTCAT	79
M216	rs2032666	Y:15437564	CCTCAACCAGTTTTTATGAAGCTAGA	AGCAAAAGATAATTGTTCCAGGGTAAG	145
V86	rs192993367	Y:6909957	AAGCAGGTATCATCCACAAGAG	ACTGTAGGAAGCTGCTGATCAA	154
V20	rs182352067	Y:6845955	ACCACCACCCAAATAGGCTACA	TTCCACTTTCACTCCAGAGAACA	161
M38	rs369611932	Y:21742158	CAATGGTATGTAGGCAATGAAAA	TTACAACACAGGGGAATGCTTAC	87
M217	rs2032668	Y:15437333	GTGAAGGAGAATGAAAAAGTTGGG	AAAGCTGCTGTGGCTTTCATCA	88
M168	rs2032595	Y:14813991	GGTGGCTGTGTGGAGTATGTGT	CTCTTACCCAAACTGCTAAAACTATT	80
M96	rs9306841	Y:21778998	TAAAGGACCATATATTTTGCCAT	ACTTGGAAAACAGGTCTCTCATA	82
M33	rs368762706	Y:21740450	GCTGAGACAAGATCTGTTCAGTTTA	GGGAGTCACTCAGTTACAAAAGTA	84

Locus	ID	Position on GRCh37	Forward primer	Reverse primer	Amplicon Size (bp)
P2	rs9785756	Y:21610831	AGAGGTGCCCCTAGGAGGAGA	TCATGCTGGTTAAAGGAGCATT	59
M215	rs2032654	Y:15467824	TACATACTTGCTGCATTAAGACAAAC	AAATAGAGATAATTTAAATCCAGCAC	83
L336	rs112779735	Y:21903853	TCTATGCTATCTACACTTCTCCACC	GCTTTTTGTGGACCCCGCTG	82
M78	rs368977028	Y:21893303	TTTCTGCATTACTCCGTATGTTC	AGTACTATGACCAGCTTATTTTGA	106
M75	rs2032639	Y:21890177	AAAAGACAATTATCAAACCACATC	ATTGAACAGAGGCATTTGTGAA	84
P14	rs9786420	Y:17398598	CTGTCACTAGACTGTGATCAGCATT	GCATGTATTTCCATCTATCTGCATGT	163
M201	rs2032636	Y:15027529	TCAGATCTAATAATCCAGTATCAACTGAG	GTTCAAATCCCATATCCAGCA	145
U5	rs2178500	Y:23973594	CCTGTGAGGCCCAGGATGAA	AGCAGGAATCGGGTCCCATAA	158
L31	rs35617575	Y:14028148	ACACATACACCCCATGATCCA	GGTCTGCACTCAGAGGCGATT	101
M52	rs376769460	Y:19591313	AGATTACAGTTATGAGCCACCATCC	TGACGAAGCAAACATTTCAAGAGAG	128
m258	rs9341301	Y:15023364	GAAACTAACAATGGGCTTGCAC	TAAGAGGGATTCCAAGTTCCCA	68
L68	rs35547782	Y:18700150	ATGTGAGTGGATAGAAAAGTGTCAG	ACTCCTATAGCAGTGAACAAGCAA	155
M438	rs17307294	Y:16638804	CAGTTTTGGGCCTGGAATGTAGA	ACACACAAAAATGAAAAAGCCAA	85
M436	rs17315680	Y:17315680	ATTGTGCTATAAGTGTTGATTATCC	CCGCAGTATCTATTCATTTCCATCC	88
P219	rs17221964	Y:15517851	AATAATGTCAATCACAATCGTGTTC	AGTCCCTATCTGCAAGCTAATATAA	91
M522	rs9786714	Y:7173143	ATGTCCTCTCGCCACCATCT	AGGAGGAAGTACAGTGCAGAAA	110
M304	rs13447352	Y:22749853	GTATGTGGGATTTTTTTAGATGTGT	AACGTCTTATACCAAAATATCACCA	78
M267	rs9341313	Y:22741818	AGTCTGGATAGCGGATTCGATGG	GGCATCAGCTAGATTGTGTTCTT	84
L559	rs373707621	Y:19512441	CAGGGTGCAAAGGGGTTGA	CAAGAGATGTTGAGAAGATACTGTGA	119
M102	rs2032608	Y:21926112	AGATAAAATTCACATAGTGCAATGG	TGCTGTTTATTCTTATTGTCTTTTC	62
M221	rs2032667	Y:15436316	CTGTATATGATATCTATATTATCACTT	CTTTCTCACTCAATTTTATACCTCT	84
P131	rs9786043	Y:15472863	CCTTCCACTTAGTAACACTCAGAACA	CATATGCACACACATAGAATTCCTCC	161
P326	rs372687543	Y:8467290	TCAGATATCAGGCCGCTTTCA	TGCCAGGCTATCAGAGCTGT	73
M231	rs9341278	Y:15469724	AAAATGTGGGCTCGTTTTAATTATATTC	TTTGACACCACAGAAATTACAGGTATGA	151
P186	rs16981290	*Y:7568568*	TTTACAGGGTGGAAAGATTAAATGAGG	GAAACACTTTGGGGAATGAAAAAATAGA	168
M119	rs72613040	Y:21762685	GGTTATTCCAATTCAGCATACAG	TGTGTTTATTTGGGGAGACAGA	76
M122	rs78149062	Y:21764674	GATAGCTTTATTCAGATTTTCCCC	AATCAAGGTAGAAAAGCAATTGAG	67
M45	rs2032631	Y:21867787	AGAGAGGATATCAAAAATTGGCAG	CTGGACCTCAGAAGGAGCTTT	66
M242	rs8179021	Y:15018582	GCAAAAAGGTGACCAAGGTG	GGTGGGGGAAAAAACCATGT	90
M207	rs2032658	Y:15581983	ACTATGGGGCAAATGTAAGTCA	ACTTCAACCTCTTGTTGGAAGAT	86
M173	rs2032624	Y:15026424	TTTTCTTACAATTCAAGGGCATTTA	CTGAAAACAAAACACTGGCTTATC	81
L146	rs17250535	Y:23473201	CCTGGCAAACTTTTCATTGCT	AACACTTGCCTTCCAATATGGGAC	100
M449	rs17222279	Y:16839499	ATCATCCTGTTTATCAACTGCAA	CCTGAAAGTGTGGTGGTGAAA	70
M459	rs17316227	Y:6906074	TTTTGCTGATGAATCAAGCCAAC	CACTCAGAAATGGCATCACAAAC	84
L562	-	Y:9790893	TGCATATTTTGGGTATTTTGGCCTC	GTAACAGAAACAAAGGCAACTAGGA	108
M415	rs9786194	Y:9170545	CATGACTACCCCATGTGTTTACT	ATTGAAAACTACTCAGTTTCCCG	75
P202	rs369616152	Y:14001024	TCACAAAACCATCTGCTTTGTATT	CTTCCCAGTTTGTGGTTCTTTGTT	77
M184	rs20320	Y:14898163	ATTTTAGTCTGTGTCTTTTTCCTTT	CTTCTCCAAGTTTTGCATGGTC	89
CDKN2B-AS1	rs2383206	9:22115026	CATGGCCCGATGATTTTCAGT	AGGATTCAGGCCATCTTGCA	110
EDNRB-AS1	rs5351	13:78475313	AGGTCATTAGTGTATAAAAAAATGCAGT	ATCGTAAAAATTCTCTCATCCCTATAGT	118
HLA-DQ1	rs2187668	6:32605884	GGTGGGGAGTGGGTAAAGGA	ACACATATGAGGCAGCTGAGA	95
HFE	rs1800562	6:26093141	TGGGGAAGAGCAGAGATATACG	CCTGGCTCTCATCAGTCACA	94
CCR5	CCR5-Δ32/ rs333	3:46414947-46414978	TTGCGTCTCTCCCAGGAATC	GGACCAGCCCCAAGATGAC	141
G6PD	rs5030868	X:153762634	TGGTCGATGCGGTAGATCTG	GCTGGAACCGCATCATCGT	115
STAT6	rs324013	12:57510661	GAGGGATCAGCCTAGCTGCAA	CCTAACCAGTCTCCATTCCACA	150
STAT6	rs324015	12:57490100	TCCTGCTCAGCCCCATCAC	GCACTGACTGGAAGGGAAGT	120
IL13	rs1800925	5:131992809	GGGGGTTTCTGGAGGACTTC	TCTGGGGGCCTCAAATTTCC	129
IFIH1/MDA5	rs10930046	2:163137983	CCCAGTATCTGAGGAAGGGGAA	TCATCACACCAACAAAGAAGCAG	123
SOCS2	rs10745657	12:93973501	TCCTCTGTGTGGCTTGAGTT	CCAGCTCTGCTCCATGACTT	120
TOLLIP	rs5743899	11:1323564	CCGTGGGCAGTCAGGACAG	TGAGTGTCGGTGAGTCCTT	174
IL29	rs30461	19:39789115	CTGGAGGCATCTGTCACCTT	ACTCAGGGTGGGTTGACGTT	103
TGFB2	rs900	1:218614905	GGCGGTACTAGTTCAGACACTTTGG	TCTTGTCTCTCTCACTTACAAAGTAGG	122
CCL18	rs2015086	17:34391617	CAACCCTGGGCTGAGAACTC	CAACTCCTGGCCTCCTTCTG	137
CCL18	rs14304	17:34398495	GGCCCAGGAGGGAACAGGA	TAATGTGGCTGGGCATAGCAGA	98
NOS2A	rs7215373	17:26075500	ACATTCAGCCAGAGTGATGCTA	AGCAGGTGCACCTGTAGATT	117
SH2B3	rs3184504	12:111884608	TCGTAGAGTTCAAGGCCCAAG	ATTGCCCCAGGGTGTGAAAAG	142
PPT2-EGFL8	rs2269424	6:32132233	GAAATAGGGGCCTCAGAGGGT	GCTCAGTAGGGTCAAGTCCCA	91
TLR1	rs4833095	4:38799710	TGACCCTGTAGCTTCACGTT	TGGCGAAACTTCAAACAAATCCA	148
TLR3	rs3775291	4:187004074	TCACTTGCTCATTCTCCCTTACAC	CATTCCTGGCCTGTGAGTTCTT	149

Locus	ID	Position on GRCh37	Forward primer	Reverse primer	Amplicon Size (bp)
TLR6	rs5743810	4:38830350	GTGGTTGAGGGTAAAATTCAGTAAGGT	CGCTATCCAAGTGAACATATCAGTTAAT	148
TLR1-6-10	rs4833103	4:38815502	ATTGCCACTGATATATCAGTACC	GCAAAGGGCAAATCTCTGCTC	87
CASP12	rs497116	11:104763117	GCCATTACCTGAGCTGTGAGA	CTGGCCTCAACATCCGCAAC	145

Amplicon size does not account for the extra sequences brought by the 5' and 3' tag sequences and the T3 promoter. Only the annealing portion of each primer is reported.

Supplementary Table 2: Results of the amplifications of nuclear capture targets and their respective negative controls

	Amplification 1				Amplification 2			
	4ng template DNA		NTC		4ng template DNA		NTC	
Marker	Ct	Tm	Ct	Tm	Ct	Tm	Ct	Tm
LP1	26.12	79.86		74.58	26.45	80.08	35	80.01
MCM6 4	25.58	78.11		73.53	25.755	78.185		79.875
HERC2	25.88	82.38			25.905	82.565		
IRF4	25.95	80.89			43216	81.055	35	80.235
SLC45A2	25.46	80.1	35	84.78	25.465	80.275		
SLC24A4	43337	79.6		73.53	43369	79.77		79.86
TYR 2	25.89	79.08		74.58	25.935	79.29		
OCA2 1	25.47	86.87	32.605	81.09	25.45	87.035	33.595	81.49
OCA2 2	25.99	73.92	33.995	76.11	25.92	74.45	35	78.04
SLC24A5	25.84	82.7			25.905	82.81		
ASIP	26.22	81.63	35	80.87	26.29	81.74		
MC1R 1 2	26.19	90.22	35	78.965	26.505	90.285	35	81.16
MC1R 3	27.85	86.51			28.595	86.575		
MC1R 4 5 6	26.86	89.81	35	76.74	27.165	89.865		73.65
MC1R 7	27.82	89.8	35	79.645	27.985	89.895	35	79.715
MC1R 8	26.96	85.58			27.185	85.7	35	80.04
KITLG	25.44	80.25	34.255	78.985	25.63	80.445	34.685	79.26
TYR 1	43338	81.84	34.685	79.34	26.125	82.07	35	79.95
EXOC2	26.57	80.2	35	79.83	26.555	80.685	35	78.07
BNC2	25.43	82.26			25.615	82.44		
CASP12	25.51	79.47	35	80.825	25.47	79.595	35	83.27
AMELO	25.76	87.01		78.335	25.76	87.2	34.66	81.245
AMEAB	26.61	79.03			26.85	79.17		
PLRP2	25.99	78.81	34.07	78.155	26.34	78.92	35	78.85
AR	26.91	80.39	35	79.71	27.025	80.51		
CSMD1	26.44	82.08			26.51	82.19	35	80.99
RAAN	25.77	79.51	35	78.645	25.94	79.71	35	78.48
LIMGH1	26.57	80.48	35	79.47	26.705	80.59		
HTBP2	25.64	82.12	35	79.64	25.715	82.285		79.42
LOC442008	25.82	81.73	35	80.31	43157	81.89	35	82.67
BRF1	26.54	89.7	35	85.99	26.525	89.945	35	80.015
DARC	25.61	86.37	35	79.42	25.67	86.425	35	81.21
AIM A	25.82	80.26	35	80.97	25.95	80.355	35	82.16
AIM B	25.73	80.45	33.93	78.9	43277	80.535	33.795	78.86
AIM C	26.49	85.04	35	79.995	26.13	85.185	35	80.28
AIM D	25.28	84.16			25.365	84.245	35	81.64
M42	26.85	77.32	35	80.39	27	77.5		79.87
M168	27.41	79.14	35	78.08	27.775	79.235	35	78.105
M216	27.18	75.91	35	77.01	43247	76.035	35	78.955
M38	27.55	75.55			27.585	75.805	35	77.98
M217	26.55	76.48	35	78.725	26.53	76.58	35	77.545
V20	25.57	82.42			25.625	82.575		
V86	26.53	81.46			26.525	81.715		75.71
M96	26.69	74.95			26.795	75.14		77.83
P2	27.42	78.07		71.6	27.51	78.22		
M33	26.82	78.33			43158	78.515		
M75	27.13	74.47	35	79.77	27.21	74.58		77.66
M201	43369	77.94	35	80.175	26.975	78.23		
U5	26.47	84.32	35	76.795	26.755	84.38		
L31	26.98	79.7		73.53	27.095	79.895	35	80.41
M522	27.2	81.66	35	82.455	27.335	81.805		70.86
m258	26.89	78.37	35	78.305	26.955	78.57	35	78.84
L68	26.75	78.57		73.53	26.79	78.63		
M304	26.73	76.06	35	80.25	43186	76.19	31.775	76.045
L559	26.46	77.65			26.565	77.74	35	79.28
P131	26.51	74.87		74.58	26.59	74.565	43368	77.15
P326	27.6	81.17		74.58	43431	81.33		77.49
P202	35	78.52		74.58				
M45	27.7	76.6		77.735	27.29	76.655		
M242	27.45	77.61		80.75	27.565	77.655	35	78.335
M173	26.84	75.85		74.61	26.815	76.005		
M207	43339	75.38			27.815	75.535		74.685
M231	29.82	72.94 - 76.74			30.27	73.18		
P186	28.81	76.16			28.915	76.3		
M119	26.82	76.68			26.92	76.865	35	78.275
M122	27.13	76.44			27.455	76.59	35	76.02
M52	26.66	77.3	35	80.125	28.085	77.42		
P14	43159	78.04	35	74.925	28.425	78.12		79.695

CDKN2B-AS1	25.66	79.29			25.65	79.45	35	80.1
EDNRB-AS1	26.24	78.61		76.31	26.92	78.845		
HLA-DQ1	25.99	81.58	43343	75.98	26.235	81.835		73.225
HFE	25.95	83.98			26.165	84.12	35	80.04
G6PD	27.21	85.83	30.97	77.35	27.265	79.405	35	76.51
STAT6 1	26.42	80.69	35	80.12	26.14	85.99		73.64
STAT6 2	27.18	86.82			27.365	80.89	35	76.205
IL13	28.43	85.26	35	78.4	28.78	87.035	34.32	78.655
TLR1	26.7	78.18	35	79.33	43126	85.53		77.965
TLR3	26.57	80.99		76.54	26.515	78.285	35	75.59
TLR6	27.4	76.94			27.32	81.15		
MDA5	25.55	77.48		74.42	25.725	77.14		76.96
SH2B3	25.85	85.1			25.79	77.575	35	80.76
SOCS2	25.73	80.95	35	78.29	25.77	85.195	35	80
TOLLIP	28.78	88.9	34.145	75.87	28.915	81.1	35	78.27
IL29	27.17	83.9	35	77.65	27.145	89.07	35	80.81
TGFB2		84.1	35	79.56	26.275	84.115		77.23
CCL18 1				79.61	25.51	78.91	35	81.77
CCL18 2					26.38	82.35		
NOS2A		79.38		73.4	26.515	86.52		75.285
CCR5		79.09		72.785	<10	81.82		73.99
OCA2-2	24.295	87.565	35	81.59	*added 16/06/16*			
ASIP-2	23.72	83.365 - 88.775	35	80.44	*added 16/06/16*			
EDAR	43427	84.685		82.74	*added 16/06/16*			
FADS1-FADS2	21.605	83.855	35	81.4	*added 16/06/16*			
DHCR7/NADSYN1	24.145	82.75	35	83.45	*added 16/06/16*			
SLC22A4-1	24.63	81.43	35	81.97	*added 16/06/16*			
SLC22A4-2	24.855	82.805		67.78	*added 16/06/16*			
PPT2	23.855	85.425	35	82.99	*added 16/06/16*			
GRM5	24.945	85.335		65.46	*added 16/06/16*			
TLR1-6-10	25.625	83.475			*added 16/06/16*			
M436	25.56	78.42			*added 16/06/16*			
P219	25.985	77.705			*added 16/06/16*			
M102	26.105	78.055			*added 16/06/16*			
M221	26.62	79.945			*added 16/06/16*			
L146	24.625	81.375			*added 16/06/16*			
M415	26.23	80.1			*added 16/06/16*			
M215	25.47	79.98			*added 16/06/16*			
L336	25.6	85.585			*added 16/06/16*			
M78	25.55	79.715			*added 16/06/16*			
M438	25.46	80.705			*added 16/06/16*			
M184	43246	80.84			*added 16/06/16*			
M267	26.4	80.64			*added 16/06/16*			
P209	25.705	78.57 - 85.67			*added 16/06/16*			
S21	22.98	78.835			*added 16/06/16*			
M449	25.89	81.575			*added 16/06/16*			

Supplementary Table 3: Results of all shotgun sequencing runs performed during this study.

Sample	Type of bone	Library preparation	USER	Type of run	Total	Unique	% endogenous
ATT13	PB	8-9/07/17	no	NextSeq (screening)	223288	89078	39.9
ATT26	PB	30/11/17	yes	NextSeq (genome)	5545147	3688046	66.5
ATT26	PB	8-9/07/17	no	NextSeq (screening)	638971	441727	69.1
Att52-1	PB	07/06/16	no	NextSeq (screening)	74891	7016	9.4
BERG02-2	PB	24/03/17	yes	NextSeq (genome)	27870147	11918249	42.8
BERG02-2	PB	07/06/16	no	NextSeq (screening)	37581	8072	21.5
BERG02-3	PB	07/06/16	no	NextSeq (screening)	44016	11870	27
BERG02-5	PB		no	NextSeq (screening)	70879	12938	18.3
BERG118-1	PB	07/06/16	no	NextSeq (screening)	50718	13569	26.8
BERG157-1	PB	19/07/16	no	NextSeq (screening)	36455	12404	34
BERG157-2	PB	24/03/17	yes	NextSeq (genome)	14072235	10250332	72.8
BERG157-2	PB	7/7/16	no	NextSeq (screening)	891009	513628	57.6
BERG157-3	PB	07/06/16	no	NextSeq (screening)	178786	23025	12.9
BERG157-5	PB	7/7/16	no	NextSeq (screening)	1165185	347936	29.9
BERG157-6	PB	7/7/16	no	NextSeq (screening)	914797	294953	32.2
BERG157-7	PB	24/03/17	yes	NextSeq (genome)	13849797	8395221	60.6
BERG157-7	PB	7/7/16	no	NextSeq (screening)	840597	393989	46.9
BERG157-9	PB	7/7/16	no	NextSeq (screening)	668832	329296	49.2
Berg34-1	PB	20/09/16	no	NextSeq (screening)	779258	193891	24.9
Berg34-2	PB	20/09/16	no	NextSeq (screening)	952082	279617	29.4
BERG61	PB	07/06/16	no	NextSeq (screening)	27460	7992	29.1
BERG643-1	PB	07/06/16	no	NextSeq (screening)	15444	4806	31.1
BERG79	PB	07/06/16	no	NextSeq (screening)	74105	8574	11.6
BES1096B	PB	8-9/07/17	no	NextSeq (screening)	343781	91233	26.5
BES1154	PB	8-9/07/17	no	NextSeq (screening)	85211	32687	38.4
BES1248	PB	30/11/17	yes	NextSeq (genome)	7907786	4948778	62.6
BES1248	PB	8-9/07/17	no	NextSeq (screening)	228498	152409	66.7
BES1249	PB	8-9/07/17	no	NextSeq (screening)	286654	128706	44.9
BFM262	PB	8-9/07/17	no	NextSeq (screening)	471119	212944	45.2
BFM265	PB	30/11/17	yes	NextSeq (genome)	11563665	6524970	56.4
BFM265	PB	8-9/07/17	no	NextSeq (screening)	310901	166809	53.7
BFT228	PB	8-9/07/17	no	NextSeq (screening)	314788	60913	19.4
BIS130	PB	30/11/17	yes	NextSeq (genome)	8544634	3805484	44.5
BIS130	PB	8-9/07/17	no	NextSeq (screening)	195426	85377	43.7
BIS159	PB	8-9/07/17	no	NextSeq (screening)	243936	99139	40.6
BIS358	PB	8-9/07/17	no	NextSeq (screening)	137902	29619	21.5
BIS382	PB	8-9/07/17	no	NextSeq (screening)	336010	120200	35.8
BIS385	PB	30/11/17	yes	NextSeq (genome)	10567309	6594540	62.4
BIS385	PB	8-9/07/17	no	NextSeq (screening)	111811	60367	54
BIS3881	PB	8-9/07/17	no	NextSeq (screening)	121480	33259	27.4
BIS3882	PB	8-9/07/17	no	NextSeq (screening)	282192	77897	27.6
BLH447	PB	8-9/07/17	no	NextSeq (screening)	285127	55751	19.6
BLP10	PB	24/03/17	yes	NextSeq (genome)	21195766	7100652	33.5
BLP10	PB	16/11/16	no	NextSeq (screening)	130238	44986	34.5
BLP10--Lib1	LB	2015	no	MiSeq (screening)	407837	5	0
BLP10--Lib2	LB	2015	no	MiSeq (screening)	392513	4	0
BLP31A	LB	2015	no	MiSeq (screening)	161760	92	0.1
BLP31-Lib1	LB	2015	no	MiSeq (screening)	437649	28	0
BLP31-Lib2	LB	2015	no	MiSeq (screening)	504711	51	0
BLP32A	LB	2015	no	MiSeq (screening)	205918	47	0
BLP32-Lib1	LB	2015	no	MiSeq (screening)	389806	97	0
BLP32-Lib2	LB	2015	no	MiSeq (screening)	282572	136	0
BLP35--Lib1	LB	2015	no	MiSeq (screening)	309808	42	0
BLP35--Lib2	LB	2015	no	MiSeq (screening)	150888	19	0
BLP68-Lib1	LB	2015	no	MiSeq (screening)	325714	196	0.1
BLP68-Lib2	LB	2015	no	MiSeq (screening)	327834	137	0
BLP9LF-Lib1	LB	2015	no	MiSeq (screening)	413342	72	0
BLP9LF-Lib2	LB	2015	no	MiSeq (screening)	1963101	372	0
BUCH2	PB	24/03/17	yes	NextSeq (genome)	41902633	13221049	31.6
CBV95	PB	30/11/17	yes	NextSeq (genome)	8245595	5276451	64
CBV95	PB	8-9/07/17	no	NextSeq (screening)	595811	410795	68.9
CCF315-1-Lib1	LB	2015	no	MiSeq (screening)	358754	1538	0.4

Sample	Type of bone	Library preparation	USER	Type of run	Total	Unique	% endogenous
CCF315-1-Lib2	LB	2015	no	MiSeq (screening)	358754	1538	0.4
CCF315-2-fracA	LB	2015	no	MiSeq (screening)	161415	88	0.1
CCF315-2-Lib1	LB	2015	no	MiSeq (screening)	526670	1557	0.3
CCF315-2-Lib2	LB	2015	no	MiSeq (screening)	359448	1322	0.4
CCT99-fracA	LB	2015	no	MiSeq (screening)	111391	183	0.2
CCT99-Lib1	LB	2015	no	MiSeq (screening)	631102	6	0
CCT99-Lib2	LB	2015	no	MiSeq (screening)	250719	19	0
CLR01	TO	07/06/16	no	NextSeq (screening)	23754	4215	17.7
CLR05	TO	07/06/16	no	NextSeq (screening)	81300	1617	2
CLR1	TO	16/11/16	no	NextSeq (screening)	380276	29485	7.8
CLR11	TO	07/06/16	no	NextSeq (screening)	89094	6947	7.8
CLR12	TO	07/06/16	no	NextSeq (screening)	149368	5145	3.4
CLR13	TO	07/06/16	no	NextSeq (screening)	160014	2140	1.3
CLR6	TO	16/11/16	no	NextSeq (screening)	45513	226	0.5
CLR8	TO	16/11/16	no	NextSeq (screening)	370949	144	0
COL11	PB	30/11/17	yes	NextSeq (genome)	11322871	4231690	37.4
COL11	PB	8-9/07/17	no	NextSeq (screening)	284581	135870	47.7
COL153A	PB	30/11/17	yes	NextSeq (genome)	10928453	5076619	46.5
COL153A	PB	8-9/07/17	no	NextSeq (screening)	262175	161169	61.5
COL153i	PB	30/11/17	yes	NextSeq (genome)	6753440	3209015	47.5
COL153i	PB	8-9/07/17	no	NextSeq (screening)	303226	179844	59.3
COL336	PB	8-9/07/17	no	NextSeq (screening)	289633	121352	41.9
CRE20D	PB	24/03/17	yes	NextSeq (genome)	35973524	10442261	29
CROI1-4	PB	8-9/07/17	no	NextSeq (screening)	357333	176170	49.3
CROI11	PB	8-9/07/17	no	NextSeq (screening)	415151	143425	34.5
CROI12-2	PB	8-9/07/17	no	NextSeq (screening)	560037	147983	26.4
ERS1164	PB	30/11/17	yes	NextSeq (genome)	9852764	6362026	64.6
ERS1164	PB	8-9/07/17	no	NextSeq (screening)	473103	307687	65
ERS83-2	PB	8-9/07/17	no	NextSeq (screening)	62696	29324	46.8
ERS86	PB	30/11/17	yes	NextSeq (genome)	13120776	6185307	47.1
ERS86	PB	8-9/07/17	no	NextSeq (screening)	424597	223573	52.7
ERS88	PB	30/11/17	yes	NextSeq (genome)	5974275	3953068	66.2
ERS88	PB	8-9/07/17	no	NextSeq (screening)	120428	88448	73.4
Es1-1-Lib1	LB	03/02/16	no	MiSeq (screening)	271981	13	0
Es1-1-Lib2	LB	03/02/16	no	MiSeq (screening)	379158	26	0
Es1018-16-Lib1	PB	14/03/16	no	NextSeq (screening)	87251	18598	21.3
Es107-3-Lib1	LB	09/02/16	no	MiSeq (screening)	362345	90	0
Es107-3-Lib2	LB	09/02/16	no	MiSeq (screening)	24475	5	0
Es113-3-Lib1	LB	12/10/2015	no	MiSeq (screening)	124893	24	0
Es113-3-Lib2	LB	12/10/2015	no	MiSeq (screening)	286041	44	0
Es129-1-Lib1	LB	03/02/16	no	MiSeq (screening)	299069	246	0.1
Es129-1-Lib2	LB	03/02/16	no	MiSeq (screening)	297542	252	0.1
Es133-1-Lib1	PB	17/03/16	no	NextSeq (screening)	1513	12	0.8
Es159-1-Lib1	TO	15/02/16	no	MiSeq (screening)	257643	29	0
Es159-1-Lib2	TO	15/02/16	no	MiSeq (screening)	289888	27	0
Es168-Lib1	PB	17/03/16	no	NextSeq (screening)	628	49	7.8
Es168-Lib2	PB	17/03/16	no	NextSeq (screening)	86943	11774	13.5
Es178-1-Lib1	LB	09/02/16	no	MiSeq (screening)	212489	31	0
Es178-1-Lib2	LB	09/02/16	no	MiSeq (screening)	410337	67	0
Es2-2-Lib1	LB	09/02/16	no	MiSeq (screening)	237359	40	0
Es2-2-Lib2	LB	09/02/16	no	MiSeq (screening)	293400	49	0
Es278-1-Lib1	PB	24/03/16	no	NextSeq (screening)	97669	33010	33.8
Es37-2-Lib1	TO	15/02/16	no	MiSeq (screening)	41496	53	0.1
Es37-2-Lib2	TO	15/02/16	no	MiSeq (screening)	287790	330	0.1
Es42-3-Lib1	PB	17/03/16	no	NextSeq (screening)	236	3	1.3
Es42-3-Lib2	PB	17/03/16	no	NextSeq (screening)	47597	3583	7.5
Es454-1-Lib1	PB	28/03/16	no	NextSeq (screening)	184652	24414	13.2
Es48-8C-Lib1	PB	03/02/16	no	MiSeq (screening)	31300	9090	29
Es48-8-Lib2	PB	03/02/16	no	MiSeq (screening)	273397	79031	28.9
Es48bis-Lib1	PB	28/03/16	no	NextSeq (screening)	111594	17444	15.6
Es537-11-Lib1	LB	10/12/2015	no	MiSeq (screening)	84042	25	0
Es537-11-Lib2	LB	10/12/2015	no	MiSeq (screening)	357254	416	0.1
Es603-1-Lib1	LB	10/12/2015	no	MiSeq (screening)	99735	16	0
Es603-1-Lib2	LB	10/12/2015	no	MiSeq (screening)	160599	27	0
Es626-27-Lib1	PB	24/03/16	no	NextSeq (screening)	228489	57993	25.4
Es637-1-Lib1	PB	17/03/16	no	NextSeq (screening)	2974	817	27.5

Sample	Type of bone	Library preparation	USER	Type of run	Total	Unique	% endogenous
Es637-1-Lib2	PB	17/03/16	no	NextSeq (screening)	95498	11999	12.6
Es656-1-Lib1	LB	10/12/2015	no	MiSeq (screening)	88308	32	0
Es656-1-Lib2	LB	10/12/2015	no	MiSeq (screening)	128894	37	0
Es72-1-Lib1	LB	09/02/16	no	MiSeq (screening)	288207	21	0
Es72-1-Lib2	LB	09/02/16	no	MiSeq (screening)	24	0	0
Es74-1-Lib1	TO	15/02/16	no	MiSeq (screening)	206535	18	0
Es74-1-Lib2	TO	15/02/16	no	MiSeq (screening)	252933	16	0
Es75-5-Lib1	PB	24/03/16	no	NextSeq (screening)	1	0	0
Es769-1-Lib1	LB	10/12/2015	no	MiSeq (screening)	79531	14	0
Es769-1-Lib2	LB	10/12/2015	no	MiSeq (screening)	122594	17	0
Es791-1-Lib1	PB	17/03/16	no	NextSeq (screening)	5967	130	2.2
Es90-9-Lib1	PB	28/03/16	no	NextSeq (screening)	150028	7897	5.3
Es97-1	PB	24/03/17	yes	NextSeq (genome)	22392628	10021521	44.8
Es97-1-Lib1	PB	28/03/16	no	NextSeq (screening)	154989	39807	25.7
EUG1	PB	20/09/16	no	NextSeq (screening)	985764	43385	4.4
EUG10	PB	20/09/16	no	NextSeq (screening)	862940	25394	2.9
EUG11	PB	24/03/17	yes	NextSeq (genome)	37709440	5384029	14.3
EUG11	PB	20/09/16	no	NextSeq (screening)	974382	157051	16.1
EUG12	PB	28/09/16	no	NextSeq (screening)	1018367	137026	13.5
EUG2	PB	20/09/16	no	NextSeq (screening)	6955	28	0.4
EUG8	PB	20/09/16	no	NextSeq (screening)	893109	137368	15.4
GOX287	PB	8-9/07/17	no	NextSeq (screening)	198932	88321	44.4
Jeb8	PB	30/11/17	yes	NextSeq (genome)	8738720	4555585	52.1
MAND1175	PB	8-9/07/17	no	NextSeq (screening)	296891	73522	24.8
MAZ16	PB	19/07/16	no	NextSeq (screening)	767164	22325	2.9
MDV188-2-1-Lib1	LB	2015	no	MiSeq (screening)	276109	366	0.1
MDV188-2-1-Lib2	LB	2015	no	MiSeq (screening)	376773	588	0.2
MDV188-3-Lib1	LB	2015	no	MiSeq (screening)	410821	17	0
MDV188-3-Lib2	LB	2015	no	MiSeq (screening)	370144	4	0
MDV188-4-Lib1	LB	2015	no	MiSeq (screening)	272815	577	0.2
MDV188-4-Lib2	LB	2015	no	MiSeq (screening)	244076	82	0
MDV188-6-1-Lib1	LB	2015	no	MiSeq (screening)	199809	64	0
MDV188-6-1-Lib2	LB	2015	no	MiSeq (screening)	300586	26	0
MDV188-6-2-Lib1	LB	2015	no	MiSeq (screening)	287366	142	0
MDV188-6-2-Lib2	LB	2015	no	MiSeq (screening)	201202	48	0
MDV188-6-3-Lib1	LB	2015	no	MiSeq (screening)	251547	85	0
MDV188-6-3-Lib2	LB	2015	no	MiSeq (screening)	296774	8	0
MDV189-2-1A	LB	2015	no	MiSeq (screening)	135361	160	0.1
MDV189-2-1-Lib1	LB	2015	no	MiSeq (screening)	388211	5	0
MDV189-2-1-Lib2	LB	2015	no	MiSeq (screening)	365577	53	0
MDV189-2-2-Lib2	LB	2015	no	MiSeq (screening)	295121	86	0
MDV1892-2-Lib1	LB	2015	no	MiSeq (screening)	310829	5	0
MDV248	TO	30/11/17	yes	NextSeq (genome)	18833468	3785925	20.1
MDV248-Lib1	TO	03/02/16	no	MiSeq (screening)	566654	67966	12
MDV248-Lib2	TO	03/02/16	no	MiSeq (screening)	222116	29261	13.2
MDV254-Lib1	LB	2015	no	MiSeq (screening)	360314	52	0
MDV254-Lib2	LB	2015	no	MiSeq (screening)	356453	69	0
MDV272-Lib1	LB	03/02/16	no	MiSeq (screening)	547989	336	0.1
MDV272-Lib2	LB	03/02/2016	no	NextSeq (screening)	35421	10337	29.2
MDV272-Lib2	LB	03/02/16	no	MiSeq (screening)	266064	236	0.1
MDV318-Lib1	LB	13/10/2016	no	MiSeq (screening)	370963	81	0
MDV318-Lib2	LB	13/10/2016	no	MiSeq (screening)	30801	8	0
MDV561-Lib1	LB	2015	no	MiSeq (screening)	483096	87	0

Sample	Type of bone	Library preparation	USER	Type of run	Total	Unique	% endogenous
MDV561-Lib2	LB	2015	no	MiSeq (screening)	224144	52	0
MDV562V-Lib1	LB	2015	no	MiSeq (screening)	482773	51	0
MDV562V-Lib2	LB	2015	no	MiSeq (screening)	532626	54	0
MDV563II-Lib1	LB	2015	no	MiSeq (screening)	518187	91	0
MDV563II-Lib2	LB	2015	no	MiSeq (screening)	516999	118	0
MDV563III-Lib1	LB	2015	no	MiSeq (screening)	147483	218	0.1
MDV563III-Lib2	LB	2015	no	MiSeq (screening)	15285	120	0.8
MDV563IV-Lib1	LB	2015	no	MiSeq (screening)	298030	366	0.1
MDV563IV-Lib2	LB	2015	no	MiSeq (screening)	352296	706	0.2
MDV93-2-Lib1	LB	2015	no	MiSeq (screening)	390401	1123	0.3
MDV93-2-Lib2	LB	2015	no	MiSeq (screening)	380418	1273	0.3
MDV93-3-lib1	LB	2015	no	MiSeq (screening)	609310	390	0.1
MDV93-3-Lib2	LB	2015	no	MiSeq (screening)	521324	285	0.1
MIT10-2032	PB	8-9/07/17	no	NextSeq (screening)	256045	11302	4.4
MIT1030B	PB	8-9/07/17	no	NextSeq (screening)	259486	38536	14.9
MIT1031	PB	8-9/07/17	no	NextSeq (screening)	263414	68404	26
MIT1059	PB	8-9/07/17	no	NextSeq (screening)	173299	16020	9.2
MIT11	PB	8-9/07/17	no	NextSeq (screening)	325279	39952	12.3
MIT1167A	PB	8-9/07/17	no	NextSeq (screening)	213862	38076	17.8
MIT1167B	PB	8-9/07/17	no	NextSeq (screening)	120310	36236	30.1
MIT1167C	PB	8-9/07/17	no	NextSeq (screening)	471332	57624	12.2
Mor4	PB	07/06/16	no	NextSeq (screening)	58603	14136	24.1
Mor5	PB		no	NextSeq (screening)	58789	10921	18.6
Mor54	PB	07/06/16	no	NextSeq (screening)	18607	5397	29
Mor6	PB	24/03/17	yes	NextSeq (genome)	9826636	5524198	56.2
Mor6	PB	20/09/16	no	NextSeq (screening)	1596523	471944	29.6
Mor6	PB	07/06/16	no	NextSeq (screening)	29730	9196	30.9
NHI11-Lib1	LB	13/10/2016	no	MiSeq (screening)	241	8	3.3
NHI11-Lib2	LB	13/10/2016	no	MiSeq (screening)	185706	4	0
NHI13-Lib1	LB	13/10/2016	no	MiSeq (screening)	705514	65532	9.3
NHI13-Lib2	LB	13/10/2016	no	MiSeq (screening)	183597	5	0
NHI1-Lib1	LB	15/02/16	no	MiSeq (screening)	272234	4	0
NHI1-Lib1	LB	15/02/16	no	MiSeq (screening)	38635	1	0
NHI20-Lib1	LB	15/02/16	no	MiSeq (screening)	224150	16	0
NHI20-Lib1	LB	15/02/16	no	MiSeq (screening)	59	0	0
NIED	PB	30/11/17	yes	NextSeq (genome)	6531228	4246651	65
NIED	PB	8-9/07/17	no	NextSeq (screening)	294077	203447	69.2
NOE2	PB	07/06/16	no	NextSeq (screening)	83973	2039	2.4
NOR2B2	PB	8-9/07/17	no	NextSeq (screening)	310069	88657	28.6
NOR2B20	PB	8-9/07/17	no	NextSeq (screening)	338651	20243	6
NOR2B6	PB	30/11/17	yes	NextSeq (genome)	9782279	5628522	57.5
NOR2B6	PB	8-9/07/17	no	NextSeq (screening)	382363	213408	55.8
NOR3-10	PB	8-9/07/17	no	NextSeq (screening)	357273	87713	24.6
NOR3-13	PB	8-9/07/17	no	NextSeq (screening)	41749	20291	48.6
NOR3-15	PB	30/11/17	yes	NextSeq (genome)	8036606	3692295	45.9
NOR3-15	PB	8-9/07/17	no	NextSeq (screening)	242573	138041	56.9
NOR3-6	PB	30/11/17	yes	NextSeq (genome)	8043152	5307105	66
NOR36	PB	8-9/07/17	no	NextSeq (screening)	342255	243079	71
NOR4	PB	30/11/17	yes	NextSeq (genome)	8007537	3423937	42.8
NOR4	PB	8-9/07/17	no	NextSeq (screening)	283746	159849	56.3
NOR5-2	PB	8-9/07/17	no	NextSeq (screening)	634209	320181	50.5
OBE1450	PB	28/09/16	no	NextSeq (screening)	773946	358	0
OBE3626-1	PB	24/03/17	yes	NextSeq (genome)	9680492	7301781	75.4
OBE3626-1	PB	28/09/16	no	NextSeq (screening)	539806	380502	70.5
OBE3626-2	PB	07/06/16	no	NextSeq (screening)	75961	12932	17
OBE3628	PB	28/09/16	no	NextSeq (screening)	233314	33143	14.2
OBE3629	PB	28/09/16	no	NextSeq (screening)	203255	115	0.1
OBE3722	PB	30/11/17	yes	NextSeq (genome)	8491464	5263626	62
OBE3722	PB	28/09/16	no	NextSeq (screening)	538436	269629	50.1
PECH10	PB	8-9/07/17	no	NextSeq (screening)	336203	18908	5.6
PECH3	PB	8-9/07/17	no	NextSeq (screening)	365528	71301	19.5
PECH5	PB	30/11/17	yes	NextSeq (genome)	2449026	1664074	67.9
PECH5	PB	8-9/07/17	no	NextSeq (screening)	244801	184865	75.5
PECH8	PB	30/11/17	yes	NextSeq (genome)	8100956	5063541	62.5
PECH8	PB	8-9/07/17	no	NextSeq (screening)	123370	89167	72.3
PECH9	PB	8-9/07/17	no	NextSeq (screening)	391557	88544	22.6

Sample	Type of bone	Library preparation	USER	Type of run	Total	Unique	% endogenous
PEI10	PB	30/11/17	yes	NextSeq (genome)	16177053	2097428	13
PEI10	PB	8-9/07/17	no	NextSeq (screening)	155270	24999	16.1
PEI2	PB	8-9/07/17	no	NextSeq (screening)	235450	15376	6.5
PER1150	PB	8-9/07/17	no	NextSeq (screening)	198089	53395	27
PER1150B	PB	8-9/07/17	no	NextSeq (screening)	280970	73064	26
PER1150C	PB	30/11/17	yes	NextSeq (genome)	8630223	5189611	60.1
PER1150C	PB	8-9/07/17	no	NextSeq (screening)	249782	160950	64.4
PER26	PB	8-9/07/17	no	NextSeq (screening)	148173	65899	44.5
PER3023	PB	30/11/17	yes	NextSeq (genome)	8978029	5343431	59.5
PER3023	PB	8-9/07/17	no	NextSeq (screening)	67763	42599	62.9
PER3123	PB	30/11/17	yes	NextSeq (genome)	7504991	4421014	58.9
PER3123	PB	8-9/07/17	no	NextSeq (screening)	207316	130978	63.2
PER503	PB	30/11/17	yes	NextSeq (genome)	9685192	5558692	57.4
PER503	PB	8-9/07/17	no	NextSeq (screening)	183060	114230	62.4
PEY163	PB	30/11/17	yes	NextSeq (genome)	8162983	4130458	50.6
PEY163	PB		no	NextSeq (screening)	314431	169049	53.8
PEY53	PB	30/11/17	yes	NextSeq (genome)	8641790	4853700	56.2
PEY53	PB	8-9/07/17	no	NextSeq (screening)	291179	164005	56.3
PEY73	PB	8-9/07/17	no	NextSeq (screening)	859029	400670	46.6
Pey74	PB	07/06/16	no	NextSeq (screening)	140931	17646	12.5
Pir1	PB	19/07/16	no	NextSeq (screening)	586102	69988	11.9
Pir2	PB	19/07/16	no	NextSeq (screening)	1536572	234397	15.3
Pir3	PB	19/07/16	no	NextSeq (screening)	520301	167222	32.1
PIR3037AB	PB	30/11/17	yes	NextSeq (genome)	8517422	2266752	26.6
PIR3037AB	PB	8-9/07/17	no	NextSeq (screening)	162893	41435	25.4
PIR3116B	PB	30/11/17	yes	NextSeq (genome)	9878183	3785048	38.3
PIR3116B	PB	8-9/07/17	no	NextSeq (screening)	219116	81323	37.1
PIR3117	PB	8-9/07/17	no	NextSeq (screening)	124489	30016	24.1
Pir4	PB	24/03/17	yes	NextSeq (genome)	13430343	3411025	25.4
Pir4	PB	19/07/16	no	NextSeq (screening)	65650	23119	35.2
Pir6	PB	24/03/17	yes	NextSeq (genome)	7767384	3317293	42.7
Pir6	PB	19/07/16	no	NextSeq (screening)	630381	302187	47.9
Pir7	PB	19/07/16	no	NextSeq (screening)	715869	108154	15.1
PON1	PB	8-9/07/17	no	NextSeq (screening)	314448	78353	24.9
PSS4170	PB	24/03/17	yes	NextSeq (genome)	42093382	12642125	30
PSS4693	PB	24/03/17	yes	NextSeq (genome)	40108567	23039209	57.4
PT2	PB	30/11/17	yes	NextSeq (genome)	7504094	3933011	52.4
PT2	PB	8-9/07/17	no	NextSeq (screening)	258428	142456	55.1
PT7	PB	8-9/07/17	no	NextSeq (screening)	285055	105751	37.1
QUIN234	PB	30/11/17	yes	NextSeq (genome)	13092518	5103721	39
QUIN234	PB	8-9/07/17	no	NextSeq (screening)	306849	112181	36.6
QUIN58	PB	30/11/17	yes	NextSeq (genome)	9454812	3732016	39.5
QUIN58	PB	8-9/07/17	no	NextSeq (screening)	359725	206115	57.3
QUIN59	PB	8-9/07/17	no	NextSeq (screening)	240791	57055	23.7
Red1041-1160	TO	28/09/16	no	NextSeq (screening)	600218	261	0
Red1047	TO	28/09/16	no	NextSeq (screening)	654650	226	0
Red1058-1519	TO	28/09/16	no	NextSeq (screening)	744188	73	0
Red1095	TO	28/09/16	no	NextSeq (screening)	937292	111	0
Red1161-1515	TO	28/09/16	no	NextSeq (screening)	712949	158	0
Red1161-1528	TO	28/09/16	no	NextSeq (screening)	700169	18	0
Red1161-64B	PB	07/06/16	no	NextSeq (screening)	24753	1899	7.7
RIX15	PB	30/11/17	yes	NextSeq (genome)	7260014	4922963	67.8
RIX15	PB	8-9/07/17	no	NextSeq (screening)	149464	102595	68.6
RIX2	PB	30/11/17	yes	NextSeq (genome)	12363728	5944314	48.1
RIX2	PB	8-9/07/17	no	NextSeq (screening)	238292	118004	49.5
RIX3	PB	8-9/07/17	no	NextSeq (screening)	186686	24162	12.9
RIX4	PB	30/11/17	yes	NextSeq (genome)	8379379	5851228	69.8
RIX4	PB	8-9/07/17	no	NextSeq (screening)	172269	118771	68.9
RIX8	PB	8-9/07/17	no	NextSeq (screening)	221070	98113	44.4
ROS100	PB	16/11/16	no	NextSeq (screening)	832879	387574	46.5
ROS102	PB	24/03/17	yes	NextSeq (genome)	7036232	4971229	70.7
ROS102	PB	16/11/16	no	NextSeq (screening)	545980	364844	66.8
ROS108	PB	7/7/16	no	NextSeq (screening)	85635	33486	39.1
ROS111	PB	16/11/16	no	NextSeq (screening)	543982	253094	46.5
ROS31	PB	7/7/16	no	NextSeq (screening)	783365	391499	50
ROS36	PB	16/11/16	no	NextSeq (screening)	541058	367809	68

Sample	Type of bone	Library preparation	USER	Type of run	Total	Unique	% endogenous
ROS42	PB	7/7/16	no	NextSeq (screening)	854191	480152	56.2
ROS43	PB	16/11/16	no	NextSeq (screening)	461362	169883	36.8
ROS45	PB	24/03/17	yes	NextSeq (genome)	11762977	8298657	70.5
ROS45	PB	16/11/16	no	NextSeq (screening)	624610	402698	64.5
ROS47	PB	16/11/16	no	NextSeq (screening)	529500	241425	45.6
ROS62	PB	7/7/16	no	NextSeq (screening)	941827	359087	38.1
ROS66	PB	16/11/16	no	NextSeq (screening)	284617	59494	20.9
ROS78	PB	24/03/17	yes	NextSeq (genome)	16758668	12314751	73.5
ROS78	PB	16/11/16	no	NextSeq (screening)	493375	301316	61.1
ROS82	PB	24/03/17	yes	NextSeq (genome)	7317897	3669225	50.1
ROS82	PB	7/7/16	no	NextSeq (screening)	812162	419981	51.7
ROS86	PB	16/11/16	no	NextSeq (screening)	332800	91819	27.6
Sad1	PB	8-9/07/17	no	NextSeq (screening)	352080	200759	57
Sad4	PB	8-9/07/17	no	NextSeq (screening)	827696	262549	31.7
SAU1	PB	30/11/17	yes	NextSeq (genome)	13462124	83110	0.6
Schw432	PB	24/03/17	yes	NextSeq (genome)	14496411	5217845	36
SCHW432	PB	07/06/16	no	NextSeq (screening)	74379	21934	29.5
Schw72-15	PB	24/03/17	yes	NextSeq (genome)	14204935	7746223	54.5
Schw72-15	PB	20/09/16	no	NextSeq (screening)	80374	36025	44.8
Schw72-15	PB	07/06/16	no	NextSeq (screening)	39932	15443	38.7
Schw72-16	PB		no	NextSeq (screening)	75427	13905	18.4
Scwh72-17	PB	20/09/16	no	NextSeq (screening)	625965	252061	40.3
Schw9	PB	07/06/16	no	NextSeq (screening)	47327	2975	6.3
SCPG2	PB	8-9/07/17	no	NextSeq (screening)	365171	139810	38.3
SCPG79	PB	8-9/07/17	no	NextSeq (screening)	258271	53622	20.8
VAS524	PB	8-9/07/17	no	NextSeq (screening)	319206	68594	21.5
VAS557	PB	8-9/07/17	no	NextSeq (screening)	443922	30140	6.8
VAS75	PB	8-9/07/17	no	NextSeq (screening)	386808	171740	44.4
VAS792	PB	8-9/07/17	no	NextSeq (screening)	552752	286465	51.8
VIGN203	PB	8-9/07/17	no	NextSeq (screening)	438543	237827	54.2
VIGN3052	PB	8-9/07/17	no	NextSeq (screening)	247097	110760	44.8
WET370-1	PB	30/11/17	yes	NextSeq (genome)	9699916	3818672	39.4
WET370-1	PB	8-9/07/17	no	NextSeq (screening)	345149	186643	54.1
WET429	PB	8-9/07/17	no	NextSeq (screening)	156365	72834	46.6

PB: petrous bone, LB: long bone, TO: tooth.

Supplementary Table 4:Summary of mitochondrial capture results and haplogroup assignment for ancient French individuals

Sample ID	Label	Unique reads	Coverage (X)	Haplogroup	Phymer score	Contamination (%)
Berg02-2	Brg	18541	83	W5b	0.972	0+/-0
BERG103-2	Brg	10318	46	X2b-226	0.999	0+/-0
BERG157-1	Brg	17096	77	U8a1	0.998	0+/-0
BERG157-2	Brg	32106	144	J1c1b	0.999	0+/-0
Berg157-4	Brg	27397	123	X2b	0.995	0+/-0
BERG157-5	Brg	31682	143	J2b1a	0.998	0+/-0
BERG157-6	Brg	31775	143	J1c1b1	0.998	0+/-0
BERG157-7	Brg	31344	141	U5b1c	0.998	0+/-0
BERG157-9	Brg	32687	147	K1a3a	0.998	0+/-0
Berg34-1	Brg	14149	64	K1a2a	0.998	0+/-0
BERG57	Brg	5001	23	H5	0.983	0+/-0
BERG746	Brg	16091	72	U5b2a3	1.000	0+/-0
BERG79	Brg	18205	82	K1a2c	0.997	0.005+/-0.005
BERG02-3	Brg_low	8010	36	R	0.979	0.015+/-0.015
BERG118-1	Brg_low	9298	42	K1	0.982	0+/-0
BERG157-3	Brg	28911	130	K1a-195	0.995	0+/-0
BERG61	Brg_low	7986	36	R	0.977	0+/-0
Berg02-5	Brg_rel	19419	87	W5b	0.998	0+/-0
BUCH2	Buc	22154	100	K1b1b1	0.997	0+/-0
BUCH3	Buc	14017	63	K1b1a1	0.999	0+/-0
BUCH4	Buc	8704	39	H1c	0.982	0+/-0
BUCH6	Buc	23848	107	K1a4	0.999	0.12+/-0.005
Es1018-16	Esc	27331	123	X2b-226	0.998	0+/-0
Es133-1	Esc	12158	55	K1a4a1	0.995	0+/-0
Es168	Esc	30552	137	K1a4a1b	0.999	0+/-0
Es278-1	Esc	27657	124	[H10-16093	0.998	0+/-0
Es42-3	Esc	30076	135	T2b	0.998	0+/-0
Es454-1	Esc	29854	134	K1a-195	0.997	0+/-0
Es48-8	Esc	29936	135	T2b3c	0.999	0+/-0
Es48bis	Esc	29426	132	K1a3a	0.998	0+/-0
Es626-27	Esc	31895	144	K1a4a1	0.998	0.025+/-0.005
Es637-1	Esc	30960	139	K1a4a1	0.998	0+/-0
Es75-5	Esc	25868	116	K1a4a1	0.999	0+/-0
Es791-1	Esc	23411	105	U5b1c	1.000	0+/-0
Es90-9	Esc	11425	51	H1e1a	0.995	0+/-0
Es97-1	Esc	30207	136	K1a-195	0.999	0+/-0
BIS130	FBA	6983	31	K2b1b	0.996	0.135+/-0.005
BIS159	FBA	7405	33	U5a1a1	0.999	0+/-0
BIS382	FBA	9025	41	U2e2a4	0.998	0+/-0
EUG11	FBA	12083	54	V	0.999	0+/-0
FAD9	FBA	20180	91	K1b1a1	0.999	0+/-0
MAND1175	FBA	6786	31	X2b-226	0.998	0.02+/-0.01
MIT1030B	FBA	5023	23	J2a1a1	0.999	0+/-0
MIT1031	FBA	3836	17	J1b1a1	0.994	0+/-0
MIT1167A	FBA	5278	24	T2e	0.996	0+/-0
MIT1167C	FBA	5904	27	H1e1a	0.999	0+/-0
NIED	FBA	15983	72	U5b2b	0.997	0.075+/-0.005
OBE3626-2	FBA	4629	21	V	0.992	0+/-0
OBE3722	FBA	6116	28	I4a	0.998	0+/-0
PIR3116B	FBA	10374	47	H2a1	0.999	0+/-0
PSS4170	FBA	12456	56	H	0.995	0+/-0
QUIN234	FBA	9570	43	H3	0.999	0+/-0
QUIN58	FBA	12815	58	H1ah	0.998	0.03+/-0.005
QUIN59	FBA	4830	22	N1a1a1a	0.978	0+/-0
RIX15	FBA	11612	52	T2b19	0.999	0.05+/-0.005
RIX4	FBA	9442	42	K1a4a1a	0.998	0.13+/-0.005
VIGN3052	FBA	11507	52	T2e	0.996	0+/-0
VTQ14	FBA	16784	76	I2	0.999	0+/-0
BIS388-1	FBA	4749	21	U2e1	0.975	0+/-0
BIS388-2	FBA	6058	27	X2b-226	0.998	0+/-0
BIS385	FBA	3098	14	T2c1d1	0.960	0+/-0
MIT11	FBA	2684	12	K1c	0.953	0+/-0
PIR3117	FBA	3455	16	K2a5	0.984	0+/-0
VIGN203	FBA	2290	10	U5b1b1-@16192	0.977	0+/-0
MIT1167B	FBA_rel	6173	28	T2e	0.995	0+/-0
BERG351-2	FIA	9471	43	K1a-195	0.998	0+/-0
BERG643-2	FIA	16711	75	K1c1	0.999	0+/-0
CLR05	FIA	3810	17	H-16291	0.972	0.51+/-0.03
CLR1	FIA	8343	38	H3-16311	0.998	0.035+/-0.01
CLR11	FIA	18218	82	U8	0.997	0+/-0

CLR12	FIA	7106	32	H1	0.996	0.09+/-0.04
CLR13	FIA	11176	50	W5b	0.995	0.08+/-0.02
NOR2B2	FIA	5977	27	J1c-16261	0.996	0.185+/-0.005
PECH10	FIA	7356	33	H1t	0.998	0+/-0
PECH3	FIA	5227	24	V	0.998	0+/-0
PECH5	FIA	12956	58	H1ab	0.998	0.04+/-0.005
PECH8	FIA	15011	68	J1c3j	0.998	0.015+/-0.005
PECH9	FIA	11363	51	J1c5	0.997	0.04+/-0.005
Pey74	FIA	23853	107	H-152	0.997	0+/-0
PIR3037AB	FBA	3770	17	U4c1a	0.980	0+/-0
PER1150B	HG	3989	18	U5b1-16189-@16192	0.987	0.055+/-0.01
PER1150C	HG	8880	40	U5b1-16189	0.997	0+/-0
PER26	HG	6223	28	U5b1	0.997	0.07+/-0.005
PER3023	HG	12136	55	U5b1-16189-@16192	0.993	0+/-0
PER3123	HG	11097	50	U5b1-16189	0.997	0.115+/-0.005
PER503	HG	13141	59	U5b1-16189	0.997	0+/-0
BLP10	Mch	20968	94	H1	0.999	0+/-0
BRE445B	Mch	20741	93	T2c1e	0.998	0+/-0
BRE445C	Mch	7834	35	V	0.993	0.1+/-0.015
BRE445FK	Mch	7527	34	U5b1	0.998	0+/-0
BRE445HI	Mch	32136	145	K1b1a1	0.998	0+/-0
CBV95	Mch	12862	58	J2a1a1	0.999	0.105+/-0.005
CHEP2	Mch	16906	76	H1	0.999	0+/-0
CHEP3	Mch	21398	96	U5b1b2	0.999	0+/-0
ISM1	Mch	25272	114	X2b-226	0.999	0+/-0
ISM2	Mch	25863	116	U5b1-16189	0.998	0+/-0
Mairy4249	Mch	11975	54	K1a3a	0.994	0+/-0
PSS1	Mch	27227	123	T2b-152	0.997	0+/-0
PSS2042	Mch	13735	62	U5b2c	0.995	0+/-0
PSS2323	Mch	11579	52	K1a2	0.999	0+/-0
PSS282	Mch	19284	87	U5b1-16189-@16192	0.997	0+/-0
PSS3072	Mch	16971	76	U5a2d	0.999	0+/-0
PSS4693	Mch	26542	119	U5b3b	1.000	0+/-0
Recy1	Mch	21536	97	T2b	0.998	0+/-0
VAC1061	Mch	11426	51	U5b1-16189-@16192	0.998	0+/-0
VAC4155-1	Mch	9877	44	J1c1b1	1.000	0+/-0
BRE445A	Mch_rel	11761	53	K1b1a1	0.998	0+/-0
CHEP4	Mch_rel	23116	104	U5b1b2	0.999	0.015+/-0.005
VAC4155-2	Mch_rel	9715	44	J1c1b1	0.999	0.095+/-0.015
BRE447	NFEN	8560	39	V	0.997	0.105+/-0.02
LARZ4	NFEN	21497	97	J1c3	0.999	0+/-0
MDV248	NFEN	19463	88	K1a2	0.999	0+/-0
ORC1	NFEN	16582	75	J1c1b1	1.000	0+/-0
Schw72-16	NFEN	17373	78	H1	0.999	0+/-0
Scwh72-17	NFEN	19406	87	J1c2	0.998	0+/-0
Mor5	NFEN	6500	29	W5b	0.937	0+/-0
Schw72-15	NFEN	21847	98	T2f	0.999	0+/-0
ROS102	Ros	22503	101	K1a1	0.997	0+/-0
Ros115	Ros	27170	122	T2b	0.999	0.005+/-0.005
Ros25	Ros	24594	111	U8b1b1	0.998	0+/-0
Ros26	Ros	10997	49	T2b3-151	1.000	0+/-0
ROS31	Ros	15370	69	T2b3-151	0.999	0+/-0
Ros40B	Ros	30307	136	H3-152	0.999	0+/-0
ROS45	Ros	18673	84	H5u	0.998	0+/-0
ROS47	Ros	23933	108	H1c	0.999	0+/-0
Ros57	Ros	30178	136	T2f	0.998	0.015+/-0.005
ROS62	Ros	11462	52	J1c2	0.997	0+/-0
ROS66	Ros	11640	52	K1a	0.993	0+/-0
Ros7	Ros	24342	110	J1c2	0.999	0+/-0
ROS78	Ros	24660	111	U8b1b1	0.998	0+/-0
ROS82	Ros	29391	132	T2b	0.998	0+/-0
Ros85	Ros	7427	33	X2b4	0.997	0+/-0
ROS86	Ros	19946	90	J2b1a	0.998	0+/-0
Ros88	Ros	12354	56	K2a9	0.997	0+/-0
Ros91	Ros	16748	75	K2a	0.999	0+/-0
Ros96	Ros	5294	24	K2b1	0.988	0.255+/-0.02
ROS42	Ros	6551	29	K1b1	0.978	0+/-0
BOU1	SFLNCA	25993	117	U2e1c1	0.999	0.1+/-0.005
BOU10	SFLNCA	12546	56	K1a-195	0.999	0+/-0
BOU11	SFLNCA	32184	145	T2c1d-152	0.998	0.02+/-0.005
BOU12	SFLNCA	9111	41	H-195	0.981	0+/-0
BOU15	SFLNCA	15441	69	V17	0.998	0.105+/-0.005
BOU4	SFLNCA	25346	114	K1b1a	0.998	0.04+/-0.005
BOU6	SFLNCA	12879	58	U5b2b3a	0.999	0+/-0

BOU7	SFLNCA	17535	79	H10-16093	0.998	0.07+/-0.005
BOU9	SFLNCA	29110	131	K1a-195	0.999	0+/-0
EUG12	SFLNCA	10365	47	J1c8a	0.997	0+/-0
EUG2	SFLNCA	12650	57	U5b1d1	0.999	0+/-0
EUG5	SFLNCA	5199	23	H1ak1	0.994	0+/-0
EUG8	SFLNCA	6965	31	T2c1d-152	0.997	0+/-0
FAD1	SFLNCA	14694	66	J1c1b	0.999	0+/-0
FAD10	SFLNCA	21404	96	H	0.999	0+/-0
FAD12	SFLNCA	10699	48	H	0.999	0+/-0
FAD2	FBA	18870	85	H1-16189	0.998	0+/-0
FAD3	SFLNCA	13653	61	H3	0.998	0+/-0
MIT1155	SFLNCA	4714	21	U5b1-16189	0.961	0+/-0
MIT83A	SFLNCA	4378	20	K1b1a	0.966	0+/-0
BER36	SFMN	8275	37	K1b1a	0.996	0+/-0
BER37	SFMN	17692	80	U3a1	0.999	0+/-0
BER49	SFMN	11702	53	J1c1b	0.999	0+/-0
CRE11A	SFMN	14802	67	X2b-226	0.999	0+/-0
CRE11B	SFMN	15443	69	J1c3	0.999	0+/-0
CRE11C	SFMN	14794	67	J2b1a2	1.000	0+/-0
CRE14	SFMN	20148	91	H4a1a	1.000	0+/-0
CRE20-B	SFMN	21272	96	H3	0.999	0+/-0
CRE20-D	SFMN	17307	78	U5b1c	0.999	0+/-0
CRE21	SFMN	13089	59	K1a-195	0.997	0+/-0
CRE29	SFMN	16137	73	K1a1b1	0.999	0+/-0
CRE3	SFMN	15268	69	U3a1	0.999	0+/-0
CRE30	SFMN	6072	27	H3	0.989	0+/-0
CRE34	SFMN	8839	40	V	0.998	0+/-0
CRE5B	SFMN	13167	59	H3c	0.999	0+/-0
CRE5C	SFMN	14705	66	H1a	0.998	0+/-0
CRE5D	SFMN	14250	64	J1c3	0.999	0+/-0
Cx13	SFMN	11253	51	U5b1-16189	0.999	0+/-0
Cx161	SFMN	12331	55	H1	0.999	0+/-0
Cx165	SFMN	18043	81	H1e1a	0.999	0+/-0
Cx19	SFMN	21767	98	K1a4a1	0.998	0+/-0
Cx20	SFMN	7471	34	H4a1a	0.993	0+/-0
Pir2	SFMN	15750	71	J1c3	1.000	0+/-0
Pir3	SFMN	15904	72	K1a4a1	0.995	0+/-0
Pir4	SFMN	17978	81	V	0.999	0+/-0
Pir7	SFMN	16060	72	X2b	0.997	0+/-0
SP1	SFMN	7177	32	V	0.987	0+/-0
SP249	SFMN	9248	42	J2b1a2	0.999	0+/-0
SP395	SFMN	18084	81	K1a-195	0.995	0+/-0
VT5-7	SFMN	20040	90	X2b-226	0.999	0+/-0
VT7	SFMN	5453	25	HV0-195	0.996	0+/-0
VTP4-3	SFMN	21298	96	U5b2b	0.999	0+/-0
CRE20-A	SFMN	8640	39	K1a4a1	0.968	0+/-0
CRE7-A	SFMN	7371	33	J2b1a	0.978	0+/-0
Cx166	SFMN	3717	17	J1c3	0.970	0+/-0

Supplementary Table 5: Summary of Y chromosome SNPs capture results and haplogroup assignment for ancient French individuals

Sample	Period	Capture # SNPs	# derived alleles	Haplogroup	Shotgun # SNPs	# derived alleles	Haplogroup
Berg157-4	Brg	25	2	E1b1b			
BERG157-1	Brg	41	3	I2			
BERG103-2	Brg	94	4	IJK			
BERG79	Brg	46	6	E1b1b1a1			
BERG157-3	Brg	124	11	E1b1b1a1			
BERG157-5	Brg	222	12	I2			
BERG157-9	Brg	259	13	E1b1b1a1			
BERG157-7	Brg	296	17	I2	5300	133	I2a1b
BERG61	Brg	14	1	IJK			
BUCH2	Buc	115	4	undetermined	7422	207	H2a1
PSS282	Buc	115	7	E1b1a1a1a1c2c			
Es48b	Esc	111	7	I			
Es42-3	Esc	143	9	I			
Es637-1II	Esc	2	1	R1b1			
Es637	Esc	3	1	R1b1			
PER3123	HG	144	5	I	2754	72	I
PSS3072	Mch	87	8	I			
BLP10	Mch	146	8	I	3971	90	I
PSS4170	NFBA	74	3	R1b1	7529	231	R1b1a1a2a2a
BIS3882	NFBA	108	4	R1b1a1a			
RIX15	NFBA	241	6	undetermined	3238	98	R1b1a1a2a1a
BIS130	NFBA	106	1	undetermined	2462	67	R1b1a1a2
CBV95	NFBB	213	12	R1b1a1	3251	97	R1b1a1a2
Schw72-15	NFEN	119	2	C1a2			
Scwh72-17	NFEN	86	3	G2a			
MOR6	NFEN	46	5	G2a			
LARZ4	NFEN	134	8	H2			
MDV248	NFEN	67	3	undetermined	2466	65	H2a1
NOR2B2	NFIA	93	3	G2a			
ROS66	Ros	46	2	IJK			
ROS45	Ros	134	5	I2	5227	131	I2a1b
ROS102	Ros	162	6	I	3119	76	I2a1b
ROS47	Ros	206	11	I2			
ROS78	Ros	278	15	I2	7951	208	I2a1b
EUG11	SFBA	114	4	R	3338	111	R1b1a2a
MAND1175	SFBA	69	7	R1b1			
PIR3116B	SFBA	100	9	R1b1	1999	50	R1b1a1a2
FAD9	SFBA	103	9	R1b1			
QUIN58	SFBA	230	16	R1b1a1a2	2401	79	R1b1a1a2
CLR11	SFIA	33	2	E1b1b			
Pey74	SFIA	78	3	R			
PECH10	SFIA	63	3	R1b1			
PECH3	SFIA	115	3	R1b1a1a			
CLR1	SFIA	58	3	R1b1a1a2			
PECH8	SFIA	413	20	R1b1a1a2a1a2c1a1a1a1a1	2911	86	R1b1a1a2a1a
PECH5	SFIA	446	24	R1b1a1a2	1025	40	R1b1a1a2a
FAD2	SFLNCA	67	6	R1b1			
FAD3	SFLNCA	88	10	I			
CRE3	SFMN	61	4	I2			
CRE10-1	SFMN	1	1	I			

Supplementary Table 6: Ancient western Eurasian dataset collected from the literature

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I1502	Kompolt-Kigyoser	Hungary	2194-2082	Hungary_BA	CBA	Central_BA	Gamba *et al.* 2014
I1504	Ludas-Varju-Dulo	Hungary	1266-1181	Hungary_BA	CBA	Central_BA	Gamba *et al.* 2014
I1495	Apc-Berekalya I	Hungary	4491-4416	Hungary_EN	CEN	Central_EN	Gamba *et al.* 2014
I1496	Apc-Berekalya I	Hungary	5206-5002	Hungary_EN	CEN	Central_EN	Gamba *et al.* 2014
I1498	Debrecen Tocopart Erdoalja	Hungary	5291-5144	Hungary_EN	CEN	Central_EN	Gamba *et al.* 2014
I1499	Garadna	Hungary	5281-5132	Hungary_EN	CEN	Central_EN	Gamba *et al.* 2014
I1500	Kompolt-Kigyoser	Hungary	5295-5116	Hungary_EN	CEN	Central_EN	Gamba *et al.* 2014
I1505	Polgar Ferenci hat	Hungary	5211-5116	Hungary_EN	CEN	Central_EN	Gamba *et al.* 2014
I1506	Polgar Ferenci hat	Hungary	5306-5229	Hungary_EN	CEN	Central_EN	Gamba *et al.* 2014
I1508	Berettyóújfalu-Morotva-Liget:	Hungary	5713-5643	Hungary_EN	CEN	Central_EN	Gamba *et al.* 2014
I1497	Apc-Berekalya I	Hungary	2903-2806	Central_MN	CLMN	Central_MN	Gamba *et al.* 2014
I1507	Tiszaszolos-Domahaza	Hungary	5781-5715	WHG	HG	Central_HG	Gamba *et al.* 2014
I0585	La Braña -Arintero, Leon	Spain	5990-5740	WHG	HG	Iberia_HG	Olade *et al.* 2014
I0047	Halberstadt-Sonntagsfeld	Germany	2022-1937	Unetice_EBA	CBA	Central_BA	Haak *et al.* 2015
I0115	Esperstedt	Germany	1931-1780	Unetice_EBA	CBA	Central_BA	Haak *et al.* 2015
I0116	Esperstedt	Germany	2118-1961	Unetice_EBA	CBA	Central_BA	Haak *et al.* 2015
I0117	Esperstedt	Germany	2199-2064	Unetice_EBA	CBA	Central_BA	Haak *et al.* 2015
I0164	Quedlinburg VIII	Germany	2012-1919	Unetice_EBA	CBA	Central_BA	Haak *et al.* 2015
I0803	Eulau	Germany	2115-1996	Unetice_EBA	CBA	Central_BA	Haak *et al.* 2015
I0108	Rothenschirmbach	Germany	2497-2436	Bell_Beaker_LN	CBB	Central_LN	Haak *et al.* 2015
I0111	Rothenschirmbach	Germany	2414-2333	Bell_Beaker_LN	CBB	Central_LN	Haak *et al.* 2015
I0112	Quedlinburg XII	Germany	2340-2190	Bell_Beaker_LN	CBB	Central_LN	Haak *et al.* 2015
I0113	Quedlinburg XII	Germany	2290-2130	Bell_Beaker_LN	CBB	Central_LN	Haak *et al.* 2015
I0805	Quedlinburg VII	Germany	2360-2190	Bell_Beaker_LN	CBB	Central_LN	Haak *et al.* 2015
I0806	Quedlinburg VII	Germany	2296-2206	Bell_Beaker_LN	CBB	Central_LN	Haak *et al.* 2015
I0049	Esperstedt	Germany	2454-2291	Corded_Ware_LN	CLN	Central_LN	Haak *et al.* 2015
I0050	Esperstedt	Germany	2800-2050	Corded_Ware_LN	CLN	Not included	Haak *et al.* 2015
I0103	Esperstedt	Germany	2566-2477	Corded_Ware_LN	CLN	Central_LN	Haak *et al.* 2015
I0104	Esperstedt	Germany	2473-2348	Corded_Ware_LN	CLN	Central_LN	Haak *et al.* 2015
I0106	Esperstedt	Germany	2454-2291	Corded_Ware_LN	CLN	Central_LN	Haak *et al.* 2015
I0176	Szemely-Hegyes	Hungary	5210-4940	LBKT_EN	CEN	Central_EN	Haak *et al.* 2015
I0174	Alsónyék-Bátaszék, Mérnöki telep	Hungary	5710-5530	Starcevo_EN	CEN	Central_EN	Haak *et al.* 2015
I0170	Oberwiederstedt, Arschkerbe West	Germany	4475-3950	Gatersleben_EN	CEN	Not included	Haak *et al.* 2015
I0019	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	5500-4800	LBK_EN	CEN	Not included	Haak *et al.* 2015
I0020	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	5500-4800	LBK_EN	CEN	Not included	Haak *et al.* 2015
I0021	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	5500-4800	LBK_EN	CEN	Not included	Haak *et al.* 2015
I0022	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	5500-4800	LBK_EN	CEN	Central_EN	Haak *et al.* 2015
I0023	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	5500-4800	LBK_EN	CEN	Not included	Haak *et al.* 2015
I0024	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	5500-4800	LBK_EN	CEN	Not included	Haak *et al.* 2015
I0025	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	5500-4800	LBK_EN	CEN	Central_EN	Haak *et al.* 2015

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I0026	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	5500-4800	LBK_EN	CEN	Central_EN	Haak *et al.* 2015
I0027	Viesenhäuser Hof, Stuttgart-Mühlhausen	Germany	5500-4800	LBK_EN	CEN	Not included	Haak *et al.* 2015
I0100	Halberstadt-Sonntagsfeld	Germany	5032-4946	LBK_EN	CEN	Central_EN	Haak *et al.* 2015
I0101	Derenburg-Meerenstieg II	Germany	5500-4775	LBK_EN	CEN	Not included	Haak *et al.* 2015
I0796	Karsdorf	Germany	5500-4775	LBK_EN	CEN	Not included	Haak *et al.* 2015
I0797	Karsdorf	Germany	5500-4775	LBK_EN	CEN	Central_EN	Haak *et al.* 2015
I0820	Halberstadt-Sonntagsfeld	Germany	5298-5247	LBK_EN	CEN	Not included	Haak *et al.* 2015
I0054	Unterwiederstedt	Germany	5222-5022	LBK_EN	CEN	Central_EN	Haak *et al.* 2015
I0051	Alberstedt	Germany	2494-2344	Alberstedt_LN	CLN	Not included	Haak *et al.* 2015
I0118	Alberstedt	Germany	2459-2345	Alberstedt_LN	CLN	Central_LN	Haak *et al.* 2015
I0058	Benzingerode-Heimburg	Germany	2283-2146	BenzigerodeHeimburg_LN	CLN	Central_LN	Haak *et al.* 2015
I0059	Benzingerode-Heimburg	Germany	2286-2153	BenzigerodeHeimburg_LN	CLN	Central_LN	Haak *et al.* 2015
I0551	Salzmünde-Schiebzig	Germany	3400-3025	Salzmünde_MN	CLMN	Central_MN	Haak *et al.* 2015
I0172	Esperstedt	Germany	3360-3086	Esperstedt_MN	CLMN	Central_MN	Haak *et al.* 2015
I0807	Esperstedt	Germany	3887-3797	Baalberge_MN	CLMN	Central_MN	Haak *et al.* 2015
I0011	Motala	Sweden	5898-5531	Motala_HG	HG	Central_HG	Haak *et al.* 2015
I0012	Motala	Sweden	5898-5531	Motala_HG	HG	Central_HG	Haak *et al.* 2015
I0013	Motala	Sweden	5898-5531	Motala_HG	HG	Central_HG	Haak *et al.* 2015
I0015	Motala	Sweden	5898-5531	Motala_HG	HG	Central_HG	Haak *et al.* 2015
I0017	Motala	Sweden	5898-5531	Motala_HG	HG	Central_HG	Haak *et al.* 2015
I0409	Els Trocs	Spain	5311-5218	Spain_EN	IEN	Iberia_EN	Haak *et al.* 2015
I0410	Els Trocs	Spain	5178-5066	Spain_EN	IEN	Iberia_EN	Haak *et al.* 2015
I0412	Els Trocs	Spain	5310-5206	Spain_EN	IEN	Iberia_EN	Haak *et al.* 2015
I0413	Els Trocs	Spain	5177-5068	Spain_EN	IEN	Iberia_EN	Haak *et al.* 2015
I0404	La Mina	Spain	3900-3600	Spain_MN	IMN	Not included	Haak *et al.* 2015
I0405	La Mina	Spain	3900-3600	Spain_MN	IMN	Iberia_MN	Haak *et al.* 2015
I0406	La Mina	Spain	3900-3600	Spain_MN	IMN	Iberia_MN	Haak *et al.* 2015
I0407	La Mina	Spain	3900-3600	Spain_MN	IMN	Iberia_MN	Haak *et al.* 2015
I0408	La Mina	Spain	3900-3600	Spain_MN	IMN	Iberia_MN	Haak *et al.* 2015
I0231	Ekaterinovka, Southern Steppe, Samara	Russia	2910-2875	Yamnaya	SBA	Steppe_BA	Haak *et al.* 2015
I0355	Belogor'e I, Volga River, Samara	Russia	3500-2700	Yamnaya	SBA	Not included	Haak *et al.* 2015
I0357	Lopatino I, Sok River, Samara	Russia	3090-2910	Yamnaya	SBA	Steppe_BA	Haak *et al.* 2015
I0428	Lopatino I, Sok River, Samara	Russia	3090-2910	Yamnaya	SBA	Not included	Haak *et al.* 2015
I0429	Lopatino I, Sok River, Samara	Russia	3339-2917	Yamnaya	SBA	Steppe_BA	Haak *et al.* 2015
I0438	Luzkhi I, Samara River, Samara	Russia	3021-2635	Yamnaya	SBA	Steppe_BA	Haak *et al.* 2015
I0439	Lopatino I, Sok River, Samara	Russia	3305-2925	Yamnaya	SBA	Steppe_BA	Haak *et al.* 2015
I0441	Kurmanaevka III, Buzuluk, Samara	Russia	3010-2622	Yamnaya	SBA	Steppe_BA	Haak *et al.* 2015
I0443	Lopatino II, Sok River, Samara	Russia	3500-2700	Yamnaya	SBA	Steppe_BA	Haak *et al.* 2015
I0444	Kutuluk I, Kutuluk River, Samara	Russia	3335-2881	Yamnaya	SBA	Steppe_BA	Haak *et al.* 2015
I0370	Ishkinovka I, Eastern Orenburg, Pre-Ural steppe, Samara	Russia	3300-2700	Yamnaya_Samara	SBA	Steppe_BA	Haak *et al.* 2015
I0099	Halberstadt-Sonntagsfeld	Germany	1113-1021	Halberstadt_LBA	CBA	Central_LN	Haak *et al.* 2015

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I0804	Eulau	Germany	2131-1982	Unetice_EBA	CBA	Central_LN	Haak *et al.* 2015
I0060	Rothenschirmbach	Germany	2428-2149	Bell_Beaker_LN	CBB	Central_LN	Haak *et al.* 2015
I0795	Karsdorf	Germany	5207-5070	LBK_EN	CEN	Central_EN	Haak *et al.* 2015
I0550	Karsdorf	Germany	2564-2475	Karsdorf_LN	CLN	Central_LN	Haak *et al.* 2015
I0171	Benzingerode-Heimburg	Germany	2204-2136	BenzigerodeHeimburg_LN	CLN	Central_LN	Haak *et al.* 2015
I0559	Quedlinburg IX	Germany	3645-3537	Baalberge_MN	CLMN	Central_MN	Haak *et al.* 2015
I0162	Oberwiederstedt 3, Schrammhöhe	Germany	4625-4250	Rössen_EN	CEMN	Not included	Haak *et al.* 2015
I0163	Oberwiederstedt 3, Schrammhöhe	Germany	4582-4407	Rössen_EN	CEMN	Not included	Haak *et al.* 2015
I0165	Oberwiederstedt 3, Schrammhöhe	Germany	4625-4250	Rössen_EN	CEMN	Not included	Haak *et al.* 2015
I0166	Halberstadt-Sonntagsfeld	Germany	5500-4775	Rössen_EN	CEMN	Not included	Haak *et al.* 2015
I0552	Salzmünde-Schiebzig	Germany	3400-3025	Salzmünde_MN	CLMN	Not included	Haak *et al.* 2015
I0554	Salzmünde-Schiebzig	Germany	3237-3171	Salzmünde_MN	CLMN	Not included	Haak *et al.* 2015
I0798	Salzmünde-Schiebzig	Germany	4172-4089	Schöningen_MN	CLMN	Not included	Haak *et al.* 2015
I0799	Salzmünde-Schiebzig	Germany	4100-3950	Schöningen_MN	CLMN	Not included	Haak *et al.* 2015
I0175	Bátaszék-Lajvérpuszta	Hungary	4900-4500	Lengyel_Neolithic	CEMN	Not included	Haak *et al.* 2015
I0212	Halle-Queis	Germany	3944-3852	Baalberge_MN	CLMN	Not included	Haak *et al.* 2015
I0556	Quedlinburg VII 2	Germany	3950-3400	Baalberge_MN	CLMN	Not included	Haak *et al.* 2015
I0557	Quedlinburg VII 2	Germany	3950-3400	Baalberge_MN	CLMN	Not included	Haak *et al.* 2015
I0560	Quedlinburg IX	Germany	3640-3510	Baalberge_MN	CLMN	Not included	Haak *et al.* 2015
I0808	Halle-Queis	Germany	3950-3400	Baalberge_MN	CLMN	Not included	Haak *et al.* 2015
I0822	Halle-Queis	Germany	3630-3581	Baalberge_MN	CLMN	Not included	Haak *et al.* 2015
I0548	Benzingerode	Germany	3104-2919	Bernburg_MN	CLMN	Not included	Haak *et al.* 2015
I0549	Benzingerode	Germany	3101-2919	Bernburg_MN	CLMN	Not included	Haak *et al.* 2015
I0014	Motala	Sweden	5898-5531	Motala_HG	HG	Not included	Haak *et al.* 2015
I0016	Motala	Sweden	5898-5531	Motala_HG	HG	Not included	Haak *et al.* 2015
I0126	Kutuluk III, Kutuluk River, Samara	Russia	2867-2484	Poltavka	SBA	Steppe_BA	Mathieson *et al.* 2015
I0371	Grachevka II, Sok River, Samara	Russia	2875-2580	Poltavka	SBA	Steppe_BA	Mathieson *et al.* 2015
I0374	Nikolaevka III, Samara River, Samara	Russia	2800-2200	Poltavka	SBA	Steppe_BA	Mathieson *et al.* 2015
I0246	Utyevka VI, Samara_River, Samara	Russia	2469-1928	Potapovka	SBA	Steppe_BA	Mathieson *et al.* 2015
I0232	Novoselki, Northern Forest, Samara	Russia	1850-1200	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0234	Rozhdestveno I, Samara Steppes, Samara	Russia	1850-1600	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0235	Rozhdestvenno I, Samara Steppes, Samara	Russia	1850-1600	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0358	Spiridonovka IV, Samara River, Samara	Russia	1913-1629	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0359	Spiridonovka IV, Samara River, Samara	Russia	1850-1200	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0360	Spiridonovka IV, Samara River, Samara	Russia	1850-1200	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0361	Spiridonovka IV, Samara River, Samara	Russia	1850-1200	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0707	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I0708	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I0709	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I0723	Menteşe	Turkey	6400-5600	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I0724	Menteşe	Turkey	6400-5600	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I0726	Menteşe	Turkey	6400-5600	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I0727	Menteşe	Turkey	6400-5600	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I0736	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I0744	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I0745	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I0746	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1096	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1097	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1098	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1099	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1100	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1101	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1102	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1103	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1579	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1580	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1581	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1583	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1585	Barcın	Turkey	6500-6200	Anatolia_Neolithic	ANE	Anatolia_N	Mathieson *et al.* 2015
I1546	Benzingerode-Heimburg	Germany	2500-2050	Bell_Beaker_LN	CBB	Central_LN	Mathieson *et al.* 2015
I1549	Benzingerode-Heimburg	Germany	2500-2050	Bell_Beaker_LN	CBB	Central_LN	Mathieson *et al.* 2015
I1532	Esperstedt	Germany	2500-2050	Central_LNBA	CLN	Central_LN	Mathieson *et al.* 2015
I1534	Esperstedt	Germany	2500-2050	Central_LNBA	CLN	Central_LN	Mathieson *et al.* 2015
I1536	Esperstedt	Germany	2500-2050	Central_LNBA	CLN	Central_LN	Mathieson *et al.* 2015
I1538	Esperstedt	Germany	2500-2050	Central_LNBA	CLN	Central_LN	Mathieson *et al.* 2015
I1539	Esperstedt	Germany	2625-2291	Central_LNBA	CLN	Central_LN	Mathieson *et al.* 2015
I1540	Esperstedt	Germany	2500-2050	Central_LNBA	CLN	Central_LN	Mathieson *et al.* 2015
I1542	Esperstedt	Germany	2500-2050	Central_LNBA	CLN	Central_LN	Mathieson *et al.* 2015
I1544	Esperstedt	Germany	2500-2050	Central_LNBA	CLN	Central_LN	Mathieson *et al.* 2015
I0581	El Mirador Cave, Atapuerca, Burgos	Spain	3010-2975	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1271	El Mirador Cave, Atapuerca, Burgos	Spain	2880-2630	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1272	El Mirador Cave, Atapuerca, Burgos	Spain	2880-2830	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1276	El Mirador Cave, Atapuerca, Burgos	Spain	2880-2630	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1277	El Mirador Cave, Atapuerca, Burgos	Spain	2830-2820	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1280	El Mirador Cave, Atapuerca, Burgos	Spain	2880-2630	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1281	El Mirador Cave, Atapuerca, Burgos	Spain	2895-2855	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1282	El Mirador Cave, Atapuerca, Burgos	Spain	2880-2630	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1284	El Mirador Cave, Atapuerca, Burgos	Spain	2880-2630	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1300	El Mirador Cave, Atapuerca, Burgos	Spain	2880-2630	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I1303	El Mirador Cave, Atapuerca, Burgos	Spain	2880-2630	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I1314	El Mirador Cave, Atapuerca, Burgos	Spain	2880-2630	Iberia_Chalcolithic	ILNCA	Iberia_LN	Mathieson *et al.* 2015
I0440	Lopatino II, Sok River, Samara	Russia	2885-2665	Poltavka	SBA	Steppe_BA	Mathieson *et al.* 2015
I0418	Utyevka IV, Samara River, Samara	Russia	2125-2044	Potapovka	SBA	Steppe_BA	Mathieson *et al.* 2015
I0419	Utyevka VI, Samara River, Samara	Russia	2200-1900	Potapovka	SBA	Steppe_BA	Mathieson *et al.* 2015
I0422	Barinovka I, Samara River, Samara	Russia	1850-1200	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0423	Barinovka I, Samara River, Samara	Russia	1850-1200	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0424	Uvarovka I, Samara River, Samara	Russia	1850-1600	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0430	Spiridonovka II, Samara River, Samara	Russia	1850-1600	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
I0431	Spiridonovka II, Samara River, Samara	Russia	1850-1600	Srubnaya	SBA	Steppe_BA	Mathieson *et al.* 2015
CB13	Cova Bonica	Spain			IEN	Iberia_EN	Olalde *et al.* 2015
G21	Almonda Cave	Spain			IEN	Not included	Olalde *et al.* 2015
I1290	Ganj Dareh	Iran	8179-7613	Iran_N	Not included	Iran_N	Lazaridis *et al.* 2016
I1944	Ganj Dareh	Iran	8000-7700	Iran_N	Not included	Iran_N	Lazaridis *et al.* 2016
I1945	Ganj Dareh	Iran	8000-7700	Iran_N	Not included	Iran_N	Lazaridis *et al.* 2016
I1949	Ganj Dareh	Iran	8241–7962	Iran_N	Not included	Iran_N	Lazaridis *et al.* 2016
I1951	Ganj Dareh	Iran	8202-7681	Iran_N	Not included	Iran_N	Lazaridis *et al.* 2016
I0867	Motza	Israel	7300-6200	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1414	'Ain Ghazal	Jordan	8300-7900	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1415	'Ain Ghazal	Jordan	8197-7653	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1679	'Ain Ghazal	Jordan	6900-6800	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1699	'Ain Ghazal	Jordan	6800-6700	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1700	'Ain Ghazal	Jordan	8300-7900	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1701	'Ain Ghazal	Jordan	7750-7569	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1704	'Ain Ghazal	Jordan	7446-7058	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1707	'Ain Ghazal	Jordan	7722-7541	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1709	'Ain Ghazal	Jordan	8300-7900	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1710	'Ain Ghazal	Jordan	7733-7526	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I1727	'Ain Ghazal	Jordan	8300-7900	Levant_N	Not included	Levant_N	Lazaridis *et al.* 2016
I0861	Raqefet Cave	Israel	11840-9760	Natufian	Not included	Levant_HG	Lazaridis *et al.* 2016
I1069	Raqefet Cave	Israel	11840-9760	Natufian	Not included	Levant_HG	Lazaridis *et al.* 2016
I1072	Raqefet Cave	Israel	11840-9760	Natufian	Not included	Levant_HG	Lazaridis *et al.* 2016
I1685	Raqefet Cave	Israel	11840-9760	Natufian	Not included	Levant_HG	Lazaridis *et al.* 2016
I1687	Raqefet Cave	Israel	11520-11110	Natufian	Not included	Levant_HG	Lazaridis *et al.* 2016
I1690	Raqefet Cave	Israel	11840-9760	Natufian	Not included	Levant_HG	Lazaridis *et al.* 2016
BerryAuBac1		France	5370-5220	Mesolithic	HG	Not included	Posth *et al.* 2016
Bockstein		Germany	-	Mesolithic	HG	Not included	Posth *et al.* 2016
	CuiryLesChaudardes1	France	6410-6100	Mesolithic	HG	Not included	Posth *et al.* 2016
Falkenstein		Germany	-	Mesolithic	HG	Not included	Posth *et al.* 2016
Felsdach		Germany	-	Mesolithic	HG	Not included	Posth *et al.* 2016
HohlensteinStadel		Germany	-	Mesolithic	HG	Not included	Posth *et al.* 2016

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
Iboussieres25-1		France	10140-10240	Mesolithic	HG	Not included	Posth *et al.* 2016
Iboussieres31-2		France	10140-10240	Mesolithic	HG	Not included	Posth *et al.* 2016
Iboussieres39		France	10090-9460	Mesolithic	HG	Not included	Posth *et al.* 2016
LesCloseaux3		France	8290-7610	Mesolithic	HG	Not included	Posth *et al.* 2016
Loschbour		Luxembourg	-	Mesolithic	HG	Not included	Posth *et al.* 2016
	MareuilLesMeaux1	France	8280-9360	Mesolithic	HG	Not included	Posth *et al.* 2016
Ofnet		England	-	Mesolithic	HG	Not included	Posth *et al.* 2016
Rochedane		France	11140-10880	Mesolithic	HG	Not included	Posth *et al.* 2016
I2355	Pusztataskony-Ledence I.	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2357	Pusztataskony-Ledence I.	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2375	Tiszadob-Ó-Kenéz	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2377	Tiszadob-Ó-Kenéz	Hungary	5208-4942	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2378	Hejőkürt-Lidl logisztikai központ	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2379	Hejőkürt-Lidl logisztikai központ	Hungary	5209-4912	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2380	Mezőkövesd-Mocsolyás	Hungary	5500-5300	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2382	Mezőkövesd-Mocsolyás	Hungary	5500-5300	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2384	Hajdúnánás-Eszlári út	Hungary	5302-5057	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2744	Cegléd, site 4/2	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2745	Cegléd, site 4/3	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I3535	Hajdúnánás-Eszlári út	Hungary	5221-5000	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I3537	Mezőkövesd-Mocsolyás	Hungary	5481-5361	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I4186	Ebes-Sajtgyár	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I4187	Ebes-Zsongvölgy	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I4188	Polgár-Piócás	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I4199	Tiszadob-Ó-Kenéz	Hungary	5300-4900	ALPc	CEN	Central_EN	Lipson *et al.* 2017
I2794	Törökszentmiklós, road 4, site 3	Hungary	5706-5541	Koros_EN	CEN	Central_EN	Lipson *et al.* 2017
I4971	Tiszaszőlős-Domaháza	Hungary	5736-5547	Koros_EN	CEN	Central_EN	Lipson *et al.* 2017
I1882	Budakeszi, Szőlőskert-Tangazdaság	Hungary	5300-4900	LBKT_MN	CEN	Central_EN	Lipson *et al.* 2017
I1904	Bátaszék-Lajvér	Hungary	5208-4948	LBKT_MN	CEN	Central_EN	Lipson *et al.* 2017
I2739	Alsónyék-Bátaszék, site 11	Hungary	5309-5074	LBKT_MN	CEN	Central_EN	Lipson *et al.* 2017
I3536	Enese elkerülő, Kóny, Proletár-dülő, M85, Site 2	Hungary	5300-4900	LBKT_MN	CEN	Central_EN	Lipson *et al.* 2017
I4181	Bölcske-Gyűrűsvölgy	Hungary	5300-4900	LBKT_MN	CEN	Central_EN	Lipson *et al.* 2017
I4196	Budakeszi, Szőlőskert-Tangazdaság	Hungary	5300-4900	LBKT_MN	CEN	Central_EN	Lipson *et al.* 2017
I1876	Alsónyék-Bátaszék, Mérnöki telep	Hungary	5641-5547	Starcevo_EN	CEN	Central_EN	Lipson *et al.* 2017
I1878	Alsónyék-Bátaszék, Mérnöki telep	Hungary	5832-5667	Starcevo_EN	CEN	Central_EN	Lipson *et al.* 2017
I1880	Lánycsók, Gata-Csatola	Hungary	5800-5500	Starcevo_EN	CEN	Central_EN	Lipson *et al.* 2017
I0046	Halberstadt-Sonntagsfeld	Germany	5206-5004	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I0048	Halberstadt-Sonntagsfeld	Germany	5206-5052	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I0659	Halberstadt-Sonntagsfeld	Germany	5079-4997	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I0821	Halberstadt-Sonntagsfeld	Germany	5034-4942	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I0018	Viesenhaeuser Hof, Stuttgart-Muehlhausen	Germany	5310-5076	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I0057	Halberstadt-Sonntagsfeld	Germany	5219-5021	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I1550	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2005	Halberstadt-Sonntagsfeld	Germany	5295-5057	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2008	Halberstadt-Sonntagsfeld	Germany	5212-4992	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2014	Halberstadt-Sonntagsfeld	Germany	5199-4857	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2016	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2017	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2020	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2021	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2022	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2026	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2029	Halberstadt-Sonntagsfeld	Germany	5295-5057	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2030	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2032	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2036	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2037	Halberstadt-Sonntagsfeld	Germany	5210-5002	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2038	Halberstadt-Sonntagsfeld	Germany	5500-4850	LBK_EN	CEN	Central_EN	Lipson *et al.* 2017
I2788	Abony, Turjányos-dűlő	Hungary	3909-3651	Protoboleraz_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2789	Abony, Turjányos-dűlő	Hungary	3800-3600	Protoboleraz_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2790	Abony, Turjányos-dűlő	Hungary	3762-3636	Protoboleraz_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2791	Abony, Turjányos-dűlő	Hungary	3658-3384	Protoboleraz_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2366	Budakalász-Luppa csárda	Hungary	3340-2945	Baden_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2367	Budakalász-Luppa csárda	Hungary	3332-2929	Baden_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2369	Budakalász-Luppa csárda	Hungary	3367-3103	Baden_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2370	Alsónémedi	Hungary	3346-2945	Baden_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2371	Alsónémedi	Hungary	3359-3098	Baden_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2752	Balatonlelle-Felső-Gamász	Hungary	3600-2850	Baden_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2753	Balatonlelle-Felső-Gamász	Hungary	3332-2929	Baden_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2763	Vörs	Hungary	3300-2850	Baden_LCA	CCA	Central_LN	Lipson *et al.* 2017
I2785	Vámosgyörk MHAT telep	Hungary	3600-2850	Baden_LCA	CCA	Central_LN	Lipson *et al.* 2017
I1907	Enese elkerülő, Kóny, Proletár-dülö, M85, Site 2	Hungary	4333-4072	Balaton_Lasinja_MCA	CCA	Central_LN	Lipson *et al.* 2017
I1908	Keszthely-Fenékpuszta, Pusztaszentegyházi-dűlő	Hungary	4300-3900	Balaton_Lasinja_MCA	CCA	Central_LN	Lipson *et al.* 2017
I1909	Lánycsók, Csata-alja	Hungary	4232-4046	Balaton_Lasinja_MCA	CCA	Central_LN	Lipson *et al.* 2017
I2394	Veszprém Jutasi út	Hungary	4339-4237	Balaton_Lasinja_MCA	CCA	Central_LN	Lipson *et al.* 2017
I4189	Alsónyék, site 11	Hungary	4300-3900	Balaton_Lasinja_MCA	CCA	Central_LN	Lipson *et al.* 2017
I2783	Nemesnádudvar-Papföld, M9/7 lh.	Hungary	4228-3963	Hunyadihalom_MCA	CCA	Central_LN	Lipson *et al.* 2017
I2793	Törökszentmiklós, road 4, site 3	Hungary	4444-4257	Tiszapolgar_Bodrogkeresztur_ECA	CCA	Central_LN	Lipson *et al.* 2017
I2353	Pusztataskony-Ledence I.	Hungary	4500-4000	Tiszapolgar_ECA	CCA	Central_LN	Lipson *et al.* 2017
I2354	Pusztataskony-Ledence I.	Hungary	4500-4000	Tiszapolgar_ECA	CCA	Central_LN	Lipson *et al.* 2017
I2395	Pusztataskony-Ledence I.	Hungary	4500-4000	Tiszapolgar_ECA	CCA	Central_LN	Lipson *et al.* 2017

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I2012	Halberstadt-Sonntagsfeld	Germany	4600-4300	Roessen_MN	CEMN	Central_MN	Lipson *et al.* 2017
I0800	Salzmünde-Schiebzig	Germany	3334-3262	Salzmünde_MN	CLMN	Central_MN	Lipson *et al.* 2017
I0802	Salzmünde-Schiebzig	Germany	3400-3025	Salzmünde_MN	CLMN	Central_MN	Lipson *et al.* 2017
I1899	Veszprém Jutasi út	Hungary	4800-4500	Lengyel_LN	CEMN	Central_LN	Lipson *et al.* 2017
I1901	Veszprém Jutasi út	Hungary	4936-4742	Lengyel_LN	CEMN	Central_LN	Lipson *et al.* 2017
I1902	Felsőörs-Bárókert	Hungary	4800-4500	Lengyel_LN	CEMN	Central_LN	Lipson *et al.* 2017
I1903	Bátaszék-Lajvér	Hungary	4800-4500	Lengyel_LN	CEMN	Central_LN	Lipson *et al.* 2017
I1905	Csabdi-Télizöldes	Hungary	4826-4602	Lengyel_LN	CEMN	Central_LN	Lipson *et al.* 2017
I1906	Csabdi-Télizöldes	Hungary	4800-4500	Lengyel_LN	CEMN	Central_LN	Lipson *et al.* 2017
I2352	Veszprém Jutasi út	Hungary	4800-4500	Lengyel_LN	CEMN	Central_LN	Lipson *et al.* 2017
I1890	Fajsz-Garadomb	Hungary	5100-4750	Sopot_LN	CEMN	Central_LN	Lipson *et al.* 2017
I1891	Fajsz-Garadomb	Hungary	5195-4842	Sopot_LN	CEMN	Central_LN	Lipson *et al.* 2017
I1893	Alsónyék-Elkerülő 2. site	Hungary	5030-4848	Sopot_LN	CEMN	Central_LN	Lipson *et al.* 2017
I4183	Szemely-Hegyes	Hungary	4904-4709	Sopot_LN	CEMN	Central_LN	Lipson *et al.* 2017
I4184	Szemely-Hegyes	Hungary	4930-4715	Sopot_LN	CEMN	Central_LN	Lipson *et al.* 2017
I4185	Alsónyék-Elkerülő 2. site	Hungary	5016-4838	Sopot_LN	CEMN	Central_LN	Lipson *et al.* 2017
I2358	Pusztataskony-Ledence I.	Hungary	5000-4500	Tisza_LN	CEMN	Central_LN	Lipson *et al.* 2017
I2387	Hódmezővásárhely-Kökénydomb Vörös tanya	Hungary	5000-4500	Tisza_LN	CEMN	Central_LN	Lipson *et al.* 2017
I2746	Vésztő-Mágor	Hungary	5000-4500	Tisza_LN	CEMN	Central_LN	Lipson *et al.* 2017
I1887	Versend-Gilencsa	Hungary	5400-5000	Vinca_MN	CEMN	Central_MN	Lipson *et al.* 2017
I1894	Versend-Gilencsa	Hungary	5400-5000	Vinca_MN	CEMN	Central_MN	Lipson *et al.* 2017
I1895	Szederkény-Kukorica-dülö	Hungary	5321-5081	Vinca_MN	CEMN	Central_MN	Lipson *et al.* 2017
I1896	Szederkény-Kukorica-dülö	Hungary	5320-5080	Vinca_MN	CEMN	Central_MN	Lipson *et al.* 2017
I1563	Blatterhole Cave	Germany	3704-3117	Blatterhohle_MN	CLMN	Central_MN	Lipson *et al.* 2017
I1565	Blatterhole Cave	Germany	4038-3532	Blatterhohle_MN	CLMN	Central_MN	Lipson *et al.* 2017
I1593	Blatterhole Cave	Germany	3958-3344	Blatterhohle_MN	CLMN	Central_MN	Lipson *et al.* 2017
I1594	Blatterhole Cave	Germany	3337-3024	Blatterhohle_MN	CLMN	Central_MN	Lipson *et al.* 2017
I1972	El Prado de Pancorbo, Burgos	Spain	4827-4692	Iberia_EN	IEN	Iberia_EN	Lipson *et al.* 2017
I2199	El Prado de Pancorbo, Burgos	Spain	5216-5031	Iberia_EN	IEN	Iberia_EN	Lipson *et al.* 2017
I1838	Las Yurdinas II, Alava, Basque Country	Spain	3354-2943	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I1843	Alto de la Huesera, Alava, Basque country	Spain	3092-2877	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I1975	La Chabola de la Hechicera, Alava, Basque country	Spain	3090-2894	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I1976	El Sotillo, Alava, Basque Country	Spain	2571-2347	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I1981	Alto de la Huesera, Alava, Basque country	Spain	3014-2891	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I2467	El Sotillo, Alava, Basque Country	Spain	2481-2212	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I2473	El Sotillo, Alava, Basque Country	Spain	2916-2714	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I3269	Las Yurdinas II, Alava, Basque Country	Spain	3350-2750	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I3270	Las Yurdinas II, Alava, Basque Country	Spain	3350-2750	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I3271	Las Yurdinas II, Alava, Basque Country	Spain	3350-2750	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I3272	La Chabola de la Hechicera, Alava, Basque country	Spain	3263-2903	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I3273	La Chabola de la Hechicera, Alava, Basque country	Spain	3627-3363	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I3276	Alto de la Huesera, Alava, Basque country	Spain	3092-2918	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I3277	Alto de la Huesera, Alava, Basque country	Spain	3100-2850	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I5838	El Mirador Cave, Atapuerca, Burgos	Spain	2900-2346	Iberia_CA	ILNCA	Iberia_LN	Lipson *et al.* 2017
I4884	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	1882-1745	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I4892	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	1881-1701	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I5035	Velké Přílepy	Czech Republic	2300-1900	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I5037	Moravská Nová Ves	Czech Republic	2300-1900	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I5042	Moravská Nová Ves	Czech Republic	2300-1900	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I5043	Moravská Nová Ves	Czech Republic	2300-1900	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I5044	Moravská Nová Ves	Czech Republic	2300-1900	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I7195	Prague 5, Jinonice, Zahradnictví	Czech Republic	2200-1700	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I7196	Prague 5, Jinonice, Zahradnictví	Czech Republic	2200-1700	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I7197	Prague 5, Jinonice, Zahradnictví	Czech Republic	2200-1700	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I7198	Prague 5, Jinonice, Zahradnictví	Czech Republic	2200-1700	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I7199	Prague 5, Jinonice, Zahradnictví	Czech Republic	2200-1700	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I7200	Prague 5, Jinonice, Zahradnictví	Czech Republic	2200-1700	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I7201	Prague 5, Jinonice, Zahradnictví	Czech Republic	2200-1700	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I7202	Prague 5, Jinonice, Zahradnictví	Czech Republic	2200-1700	Czech_EBA	CBA	Central_BA	Olalde *et al.* 2018
I4070	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	1880-1657	Netherlands_BA	CBA	Central_BA	Olalde *et al.* 2018
I4071	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	1883-1664	Netherlands_BA	CBA	Central_BA	Olalde *et al.* 2018
I7040	Szigetszentmiklós-Üdülősor	Hungary	2500-2200	Hungary_BA	CBA	Central_BA	Olalde *et al.* 2018
I7042	Szigetszentmiklós-Üdülősor	Hungary	2500-2200	Hungary_BA	CBA	Central_BA	Olalde *et al.* 2018
I7043	Szigetszentmiklós-Üdülősor	Hungary	2500-2200	Hungary_BA	CBA	Central_BA	Olalde *et al.* 2018
I6531	Dzielnica	Poland	2286-2038	Poland_BA	CBA	Central_BA	Olalde *et al.* 2018
I6537	Racibórz-Stara Wieś	Poland	2290-2041	Poland_BA	CBA	Central_BA	Olalde *et al.* 2018
I6579	Iwiny	Poland	2335-2046	Poland_BA	CBA	Central_BA	Olalde *et al.* 2018
I4885	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	2289-2143	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I4886	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	2205-2042	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I4888	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	2190-2029	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I4889	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	2281-2062	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I4891	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	2281-2062	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I4895	Prague 5, Jinonice, Butovická Street	Czech Republic	2273-2047	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I4896	Prague 5, Jinonice, Butovická Street	Czech Republic	2288-2142	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I4945	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	2291-2144	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I4946	Prague 5, Jinonice, Butovická Street	Czech Republic	2296-2146	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I5024	Kněževes	Czech Republic	2278-2032	BK_Czech_CZE	CBB	Not included	Olalde *et al.* 2018
I5025	Kněževes	Czech Republic	2500-1900	BK_Czech_CZE	CBB	Not included	Olalde *et al.* 2018
I5514	Prague 5, Jinonice, Butovická Street	Czech Republic	2500-2000	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I5666	Lochenice	Czech Republic	2500-1900	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I6468	Velké Přílepy	Czech Republic	2500-1900	BK_Czech_CZE	CBB	Not included	Olalde *et al.* 2018
I6476	Lovosice II	Czech Republic	2500-1900	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I6480	Velké Přílepy	Czech Republic	2500-1900	BK_Czech_CZE	CBB	Not included	Olalde *et al.* 2018
I7205	Radovesice	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7210	Radovesice	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7211	Radovesice	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7213	Radovesice	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7249	Brandýsek	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7250	Brandýsek	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7251	Brandýsek	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7269	Brandýsek	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7275	Brandýsek	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7276	Brandýsek	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7278	Brandýsek	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7281	Prague 5 - Malá Ohrada	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7282	Radovesice	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7286	Radovesice	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7287	Radovesice	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7289	Radovesice	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I7290	Radovesice	Czech Republic	2500-2200	BK_Czech_CZE	CBB	Central_LN	Olalde *et al.* 2018
I3588	Alburg-Lerchenhaid, Spedition Häring, Stkr. Straubing, Bavaria	Germany	2300-2150	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I3589	Alburg-Lerchenhaid, Spedition Häring, Stkr. Straubing, Bavaria	Germany	2300-2150	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I3590	Alburg-Lerchenhaid, Spedition Häring, Stkr. Straubing, Bavaria	Germany	2335-2140	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I3592	Alburg-Lerchenhaid, Spedition Häring, Stkr. Straubing, Bavaria	Germany	2457-2203	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I3594	Alburg-Lerchenhaid, Spedition Häring, Stkr. Straubing, Bavaria	Germany	2300-2150	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I3600	Alburg-Lerchenhaid, Spedition Häring, Stkr. Straubing, Bavaria	Germany	2300-2150	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I3601	Alburg-Lerchenhaid, Spedition Häring, Stkr. Straubing, Bavaria	Germany	2300-2150	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I3602	Alburg-Lerchenhaid, Spedition Häring, Stkr. Straubing, Bavaria	Germany	2300-2150	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I3607	Künzing-Bruck, Lkr. Deggendorf, Bavaria	Germany	2350-2250	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I4249	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I4250	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2433-2149	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5014	Manching-Oberstimm, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5017	Augsburg Sportgelände, Augsburg, Bavaria	Germany	2460-2206	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5019	Künzing-Bruck, Lkr. Deggendorf, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5020	Landau an der Isar, Bavaria	Germany	2457-2205	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5021	Osterhofen-Altenmarkt, Bavaria	Germany	2571-2341	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5022	Osterhofen-Altenmarkt, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5023	Osterhofen-Altenmarkt, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5519	Augsburg Sportgelände, Augsburg, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5520	Augsburg Sportgelände, Augsburg, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5521	Augsburg Sportgelände, Augsburg, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I5523	Landau an der Isar, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5524	Landau an der Isar, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5525	Landau an der Isar, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5527	Manching-Oberstimm, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5529	Osterhofen-Altenmarkt, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5531	Weichering, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Central_LN	Olalde *et al.* 2018
I5655	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5658	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5659	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5661	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5663	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5833	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5834	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I6590	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I6591	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I6624	Irlbach, County of Straubing-Bogen, Bavaria	Germany	2500-2000	BK_Germany_BAV	CBB	Not included	Olalde *et al.* 2018
I5836	Worms-Herrnsheim, Rhineland-Palatinate	Germany	2500-2000	BK_Germany_Wor	CBB	Central_LN	Olalde *et al.* 2018
I2364	Budapest-Békásmegyer	Hungary	2468-2063	BK_Hungary_HUN	CBB	Central_LN	Olalde *et al.* 2018
I5015	Budapest-Békásmegyer	Hungary	2466-2210	BK_Hungary_HUN	CBB	Central_LN	Olalde *et al.* 2018
I7044	Szigetszentmiklós-Üdülősor	Hungary	2500-2200	BK_Hungary_HUN	CBB	Central_LN	Olalde *et al.* 2018
I7045	Szigetszentmiklós-Üdülősor	Hungary	2500-2200	BK_Hungary_HUN	CBB	Central_LN	Olalde *et al.* 2018
I2365	Budapest-Békásmegyer	Hungary	2464-2207	BK_Hungary_HUN2	CBB	Central_LN	Olalde *et al.* 2018
I3528	Budakalász, Csajerszke (M0 Site 12)	Hungary	2559-2301	BK_Hungary_HUN2	CBB	Central_LN	Olalde *et al.* 2018
I3529	Budakalász, Csajerszke (M0 Site 12)	Hungary	2500-2200	BK_Hungary_HUN2	CBB	Central_LN	Olalde *et al.* 2018
I4178	Szigetszentmiklós,Felső Ürge-hegyi dűlő	Hungary	2500-2200	BK_Hungary_HUN2	CBB	Central_LN	Olalde *et al.* 2018
I2741	Szigetszentmiklós,Felső Ürge-hegyi dűlő	Hungary	2457-2153	BK_Hungary_Sfu1	CBB	Central_LN	Olalde *et al.* 2018
I2786	Szigetszentmiklós,Felső Ürge-hegyi dűlő	Hungary	2458-2205	BK_Hungary_Sfu2	CBB	Central_LN	Olalde *et al.* 2018
I2787	Szigetszentmiklós,Felső Ürge-hegyi dűlő	Hungary	2457-2201	BK_Hungary_Sfu3	CBB	Central_LN	Olalde *et al.* 2018
I4067	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	1946-1690	BK_Netherlands_Tui	CBB	Central_LN	Olalde *et al.* 2018
I4068	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	2131-1951	BK_Netherlands_Tui	CBB	Central_LN	Olalde *et al.* 2018
I4069	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	2188-1887	BK_Netherlands_Tui	CBB	Central_LN	Olalde *et al.* 2018
I4073	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	2195-1905	BK_Netherlands_Tui	CBB	Central_LN	Olalde *et al.* 2018
I4074	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	2278-1915	BK_Netherlands_Tui	CBB	Central_LN	Olalde *et al.* 2018
I4075	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	2118-1937	BK_Netherlands_Tui	CBB	Central_LN	Olalde *et al.* 2018
I4076	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	1882-1750	BK_Netherlands_Tui	CBB	Central_LN	Olalde *et al.* 2018
I5748	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	2579-2233	BK_Netherlands_Tui	CBB	Central_LN	Olalde *et al.* 2018

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I5750	De Tuithoorn, Oostwoud, Noord-Holland	The Netherlands	2300-1900	BK_Netherlands_Tui	CBB	Central_LN	Olalde *et al.* 2018
I6534	Kornice	Poland	2456-2149	BK_Poland_POL	CBB	Central_LN	Olalde *et al.* 2018
I6538	Strachów	Poland	2008-1765	BK_Poland_POL	CBB	Central_LN	Olalde *et al.* 2018
I6580	Jordanów Śląski	Poland	2300-2150	BK_Poland_POL	CBB	Central_LN	Olalde *et al.* 2018
I6582	Kornice	Poland	2343-2057	BK_Poland_POL	CBB	Central_LN	Olalde *et al.* 2018
I6583	Żerniki Wielkie	Poland	2289-2050	BK_Poland_POL	CBB	Central_LN	Olalde *et al.* 2018
I4251	Samborzec	Poland	2431-2150	BK_Poland_Sam	CBB	Central_LN	Olalde *et al.* 2018
I4252	Samborzec	Poland	2285-2138	BK_Poland_Sam	CBB	Central_LN	Olalde *et al.* 2018
I4253	Samborzec	Poland	2456-2207	BK_Poland_Sam	CBB	Central_LN	Olalde *et al.* 2018
I5755	Sion-Petit-Chasseur, Dolmen XI	Switzerland	2469-1984	BK_Switzerland_Sio	CBB	Central_LN	Olalde *et al.* 2018
I5757	Sion-Petit-Chasseur, Dolmen XI	Switzerland	2469-1984	BK_Switzerland_Sio	CBB	Central_LN	Olalde *et al.* 2018
I5759	Sion-Petit-Chasseur, Dolmen XI	Switzerland	2469-1984	BK_Switzerland_Sio	CBB	Central_LN	Olalde *et al.* 2018
I7272	Brandýsek	Czech Republic	2900-2200	Corded_Ware_Czech	CLN	Central_LN	Olalde *et al.* 2018
I7279	Brandýsek	Czech Republic	2900-2200	Corded_Ware_Czech	CLN	Central_LN	Olalde *et al.* 2018
I7280	Brandýsek	Czech Republic	2900-2200	Corded_Ware_Czech	CLN	Central_LN	Olalde *et al.* 2018
I5117	Mezőcsát-Hörcsögös	Hungary	3400-3000	Hungary_LCA	CCA	Central_LN	Olalde *et al.* 2018
I5118	Mezőcsát-Hörcsögös	Hungary	3400-3000	Hungary_LCA	CCA	Central_LN	Olalde *et al.* 2018
I5119	Mezőcsát-Hörcsögös	Hungary	3400-3000	Hungary_LCA	CCA	Central_LN	Olalde *et al.* 2018
I4893	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	4449-4348	Czech_MN	CEMN	Central_MN	Olalde *et al.* 2018
I4894	Prague 8, Kobylisy, Ke Stírce Street	Czech Republic	4488-4368	Czech_MN	CEMN	Central_MN	Olalde *et al.* 2018
I2421	Hasting Hill, Sunderland, Tyne and Wear, England	Great Britain	1930-1759	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2457	Amesbury Down, Wiltshire, England	Great Britain	2467-2030	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2460	Amesbury Down, Wiltshire, England	Great Britain	2022-1827	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2461	Porton Down, Wiltshire, England	Great Britain	2500-2140	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2462	East Kent Access (Phase II), Thanet, Kent, England	Great Britain	2128-1891	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2463	East Kent Access (Phase II), Thanet, Kent, England	Great Britain	1925-1700	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2464	Boscombe Airfield, Wiltshire, England	Great Britain	1745-1614	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2597	Amesbury Down, Wiltshire, England	Great Britain	2269-2033	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2602	East Kent Access (Phase II), Thanet, Kent, England	Great Britain	1892-1699	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2604	Barton Stacey, Hampshire, England	Great Britain	2265-2031	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2609	Hexham Golf Course, Northumberland, England	Great Britain	2022-1771	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2610	Summerhill, Blaydon, Tyne and Wear, England	Great Britain	1935-1745	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2612	Hasting Hill, Sunderland, Tyne and Wear, England	Great Britain	2464-2208	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2618	Reaverhill, Barrasford, Northumberland, England	Great Britain	2134-1950	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I5373	Carsington Pasture Cave, Brassington, Derbyshire, England	Great Britain	2194-1980	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I5377	River Thames Skulls, Mortlake, London, England	Great Britain	1894-1694	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I5441	Neale's Cave, Paington, Devon, England	Great Britain	1938-1744	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I6680	Low Hauxley, Northumberland, England	Great Britain	1876-1625	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I7635	Windmill Fields, Stockton-on-Tees, North Yorkshire, England	Great Britain	2193-1978	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I7638	Barrow Hills, Radley, Oxfordshire, England	Great Britain	2286-1778	England_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I5383	Raven Scar Cave, Ingleton, North Yorkshire, England	Great Britain	1090-900	England_LBA	GBA	UK_BA	Olalde *et al.* 2018
I2458	Amesbury Down, Wiltshire, England	Great Britain	1492-1287	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I2639	Amesbury Down, Wiltshire, England	Great Britain	1600-1430	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I3082	Canada Farm, Sixpenny Handley, Dorset, England	Great Britain	1498-1394	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7568	Over Narrows, Needingworth Quarry, England	Great Britain	1600-1300	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7570	Baston and Langtoft, South Lincolnshire, England	Great Britain	1735-1531	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7571	Over Narrows, Needingworth Quarry, England	Great Britain	1448-1259	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7572	Over Narrows, Needingworth Quarry, England	Great Britain	1510-1302	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7573	Turners Yard, Fordham, Cambridgeshire, England	Great Britain	1642-1507	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7574	Clay Farm, Cambridgeshire, England	Great Britain	1406-1230	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7575	Biddenham Loop, Bedfordshire, England	Great Britain	1263-1043	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7576	Biddenham Loop, Bedfordshire, England	Great Britain	1206-1003	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7577	Biddenham Loop, Bedfordshire, England	Great Britain	1386-1123	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7578	Biddenham Loop, Bedfordshire, England	Great Britain	1225-1013	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7580	Biddenham Loop, Bedfordshire, England	Great Britain	1256-1018	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7626	Biddenham Loop, Bedfordshire, England	Great Britain	1395-1130	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7627	Biddenham Loop, Bedfordshire, England	Great Britain	1410-1222	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7628	Biddenham Loop, Bedfordshire, England	Great Britain	1208-978	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I7640	Clay Farm, Cambridgeshire, England	Great Britain	1430-1282	England_MBA	GBA	UK_BA	Olalde *et al.* 2018
I2567	Dryburn Bridge, East Lothian, Scotland	Great Britain	2275-1884	Scotland_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2569	Eweford Cottages, East Lothian, Scotland	Great Britain	2139-1915	Scotland_CA_EBA	GBA	UK_BA	Olalde *et al.* 2018
I2981	Stenchme, Lop Ness, Orkney, Scotland	Great Britain	1950-1496	Scotland_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I3132	Covesea Cave 2, Moray, Scotland	Great Britain	2118-1887	Scotland_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I5515	Aberdour Road, Dunfermline, Fife, Scotland	Great Britain	2034-1775	Scotland_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I5516	Doune, Perth and Kinross, Scotland	Great Britain	1866-1615	Scotland_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2859	Covesea Caves, Moray, Scotland	Great Britain	910-809	Scotland_LBA	GBA	UK_BA	Olalde *et al.* 2018
I2860	Covesea Cave 2, Moray, Scotland	Great Britain	969-815	Scotland_LBA	GBA	UK_BA	Olalde *et al.* 2018
I2861	Covesea Cave 2, Moray, Scotland	Great Britain	976-828	Scotland_LBA	GBA	UK_BA	Olalde *et al.* 2018
I3130	Covesea Caves, Moray, Scotland	Great Britain	977-829	Scotland_LBA	GBA	UK_BA	Olalde *et al.* 2018
I2573	Longniddry, Evergreen House, Coast Road, East Lothian, Scotland	Great Britain	1500-1301	Scotland_MBA	GBA	UK_BA	Olalde *et al.* 2018
I2653	Longniddry, Evergreen House, Coast Road, East Lothian, Scotland	Great Britain	1500-1300	Scotland_MBA	GBA	UK_BA	Olalde *et al.* 2018
I2654	Longniddry, Evergreen House, Coast Road, East Lothian, Scotland	Great Britain	1500-1300	Scotland_MBA	GBA	UK_BA	Olalde *et al.* 2018
I2655	Pabay Mor, Lewis, Western Isles, Scotland	Great Britain	1441-1272	Scotland_MBA	GBA	UK_BA	Olalde *et al.* 2018
I1775	Great Orme Mines, Llandudno, North Wales	Great Britain	1730-1532	Wales_CA_EBA	GBA	UK_LN	Olalde *et al.* 2018
I2574	North Face Cave, Llandudno, North Wales	Great Britain	1414-1227	Wales_MBA	GBA	UK_BA	Olalde *et al.* 2018
I1767	Windmill Fields, Stockton-on-Tees, North Yorkshire,England	Great Britain	2200-1979	BK_England_NOR	GBB	UK_LN	Olalde *et al.* 2018
I1770	Staxton Beacon, Staxton,England	Great Britain	2400-1600	BK_England_NOR	GBB	UK_LN	Olalde *et al.* 2018
I5382	Windmill Fields, Stockton-on-Tees, North Yorkshire,England	Great Britain	2278-1981	BK_England_NOR	GBB	UK_LN	Olalde *et al.* 2018
I6679	Low Hauxley, Northumberland, England	Great Britain	2124-1890	BK_England_NOR	GBB	UK_LN	Olalde *et al.* 2018
I2417	Amesbury Down, Wiltshire, England	Great Britain	2500-2140	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I2418	Amesbury Down, Wiltshire, England	Great Britain	2455-2200	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2443	Yarnton, Oxfordshire, England	Great Britain	2284-2028	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2445	Yarnton, Oxfordshire, England	Great Britain	2396-1929	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2446	Yarnton, Oxfordshire, England	Great Britain	2455-2139	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2447	Yarnton, Oxfordshire, England	Great Britain	2115-1910	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2450	Abingdon Spring Road cemetery, Oxfordshire, England	Great Britain	2461-2146	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2452	Dairy Farm, Willington, England	Great Britain	2276-1980	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2453	West Deeping, Lincolnshire, England	Great Britain	2288-2040	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2454	Over Narrows, Needingworth Quarry, England	Great Britain	2198-1983	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2455	Over Narrows, Needingworth Quarry, England	Great Britain	2125-1911	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2459	Amesbury Down, Wiltshire, England	Great Britain	2455-2150	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2565	Amesbury Down, Wiltshire, England	Great Britain	2457-2147	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2566	Amesbury Down, Wiltshire, England	Great Britain	2204-2035	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2598	Amesbury Down, Wiltshire, England	Great Britain	2135-1953	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I3255	Trumpington Meadows, Cambridge, England	Great Britain	2135-1949	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I3256	Trumpington Meadows, Cambridge, England	Great Britain	2203-2029	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I4950	Central Flying School, Upavon, Wiltshire, England	Great Britain	2500-1800	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I4951	Flying School, Netheravon, Wiltshire, England	Great Britain	2500-1800	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I5376	River Thames Skulls, Syon Reach, London, England	Great Britain	2454-2141	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I5379	Canada Farm, Sixpenny Handley, Dorset, England	Great Britain	2469-2296	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I5512	Flying School, Netheravon, Wiltshire, England	Great Britain	2500-1800	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I5513	Nr. Ablington, Figheldean, England	Great Britain	2500-1800	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I6774	Ditchling Road, Brighton, Sussex, England	Great Britain	2287-2044	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I6775	Wick Barrow, Stogursey, Somerset, England	Great Britain	2400-2000	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I6777	Wilsford Down, Wilsford-cum-Lake G.54	Great Britain	2500-1900	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I6778	Wilsford Down, Wilsford-cum-Lake G.52	Great Britain	2500-1900	BK_England_SOU	GBB	UK_LN	Olalde *et al.* 2018
I2416	Amesbury Down, Wiltshire, England	Great Britain	2455-2151	BK_England_SOUout	GBB	UK_LN	Olalde *et al.* 2018
I5385	Achavanich, Wick, Highland, Scotland	Great Britain	2455-2147	BK_Scotland_Ach	GBB	UK_LN	Olalde *et al.* 2018
I2568	Dryburn Bridge, East Lothian, Scotland	Great Britain	2286-2038	BK_Scotland_ELO	GBB	UK_LN	Olalde *et al.* 2018
I5471	Thurston Mains, Innerwick, East Lothian, Scotland	Great Britain	2266-2025	BK_Scotland_ELO	GBB	UK_LN	Olalde *et al.* 2018
I5367	Sorisdale, Coll, Argyll and Bute, Scotland	Great Britain	2467-2232	BK_Scotland_Sor	GBB	UK_LN	Olalde *et al.* 2018
I0518	Banbury Lane, Northamptonshire, England	Great Britain	3360-3100	England_N	GMN	UK_MN	Olalde *et al.* 2018
I0519	Banbury Lane, Northamptonshire, England	Great Britain	3360-3100	England_N	GMN	UK_MN	Olalde *et al.* 2018
I0520	Banbury Lane, Northamptonshire, England	Great Britain	3360-3100	England_N	GMN	UK_MN	Olalde *et al.* 2018
I2605	Eton Rowing Course, Buckinghamshire, England	Great Britain	3631-2944	England_N	GMN	UK_MN	Olalde *et al.* 2018
I2606	Eton Rowing Course, Buckinghamshire, England	Great Britain	3330-2900	England_N	GMN	UK_MN	Olalde *et al.* 2018
I3068	Carsington Pasture Cave, Brassington, Derbyshire, England	Great Britain	3692-3522	England_N	GMN	UK_MN	Olalde *et al.* 2018
I4949	Nr. Millbarrow, Winterbourne Monkton, Wiltshire, England	Great Britain	3629-3376	England_N	GMN	UK_MN	Olalde *et al.* 2018
I5366	Cissbury Flint Mine, Worthing, West Sussex, England	Great Britain	3643-3383	England_N	GMN	UK_MN	Olalde *et al.* 2018
I5374	Totty Pot, Cheddar, Somerset, England	Great Britain	2830-2460	England_N	GMN	UK_MN	Olalde *et al.* 2018
I6750	Fussell's Lodge, Salisbury, Wiltshire, England	Great Britain	3755-3660	England_N	GMN	UK_MN	Olalde *et al.* 2018

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I6751	Fussell's Lodge, Salisbury, Wiltshire, England	Great Britain	3755-3660	England_N	GMN	UK_MN	Olalde *et al.* 2018
I6759	Lesser Kelco Cave, Giggleswick Scar, North Yorkshire, England	Great Britain	3650-3522	England_N	GMN	UK_MN	Olalde *et al.* 2018
I6761	Whitehawk, Brighton, Sussex, England	Great Britain	3650-3400	England_N	GMN	UK_MN	Olalde *et al.* 2018
I6762	Upper Swell, Chipping Norton, Gloucestershire, England	Great Britain	4000-3300	England_N	GMN	UK_MN	Olalde *et al.* 2018
I2630	Isbister, Orkney, Scotland	Great Britain	2580-2463	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2631	Quoyness, Orkney, Scotland	Great Britain	3097-2906	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2634	Tulach an t'Sionnach, Highland, Scotland	Great Britain	3703-3534	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2635	Tulloch of Assery A, Highland, Scotland	Great Britain	3652-3389	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2636	Holm of Papa Westray North, Orkney, Scotland	Great Britain	3519-3361	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2637	Holm of Papa Westray North, Orkney, Scotland	Great Britain	3629-3370	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2650	Holm of Papa Westray North, Orkney, Scotland	Great Britain	3638-3380	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2651	Holm of Papa Westray North, Orkney, Scotland	Great Britain	3360-3098	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2657	Macarthur Cave, Oban, Argyll and Bute, Scotland	Great Britain	3951-3780	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2659	Distillery Cave, Oban, Argyll and Bute, Scotland	Great Britain	3761-3643	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2660	Distillery Cave, Oban, Argyll and Bute, Scotland	Great Britain	3513-3352	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2691	Distillery Cave, Oban, Argyll and Bute, Scotland	Great Britain	3700-3639	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2796	Point of Cott, Orkney, Scotland	Great Britain	3705-3535	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2932	Isbister, Orkney, Scotland	Great Britain	2570-2347	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2933	Isbister, Orkney, Scotland	Great Britain	3010-2885	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2934	Isbister, Orkney, Scotland	Great Britain	3338-3022	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2935	Isbister, Orkney, Scotland	Great Britain	3335-3011	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2977	Isbister, Orkney, Scotland	Great Britain	3008-2763	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2978	Isbister, Orkney, Scotland	Great Britain	3335-3023	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2979	Isbister, Orkney, Scotland	Great Britain	3333-2941	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2980	Point of Cott, Orkney, Scotland	Great Britain	3360-3101	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I2988	Clachaig, Arran, North Ayrshire, Scotland	Great Britain	3516-3361	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I3041	Raschoille Cave, Oban, Argyll and Bute, Scotland	Great Britain	3942-3037	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I3085	Isbister, Orkney, Scotland	Great Britain	3338-3026	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I3133	Raschoille Cave, Oban, Argyll and Bute, Scotland	Great Britain	3631-3377	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I3134	Raschoille Cave, Oban, Argyll and Bute, Scotland	Great Britain	3633-3377	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I3135	Raschoille Cave, Oban, Argyll and Bute, Scotland	Great Britain	3640-3383	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I3136	Raschoille Cave, Oban, Argyll and Bute, Scotland	Great Britain	3520-3365	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I3137	Raschoille Cave, Oban, Argyll and Bute, Scotland	Great Britain	3800-3200	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I3138	Raschoille Cave, Oban, Argyll and Bute, Scotland	Great Britain	3263-2923	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I5370	Raschoille Cave, Oban, Argyll and Bute, Scotland	Great Britain	4000-3300	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I5371	Raschoille Cave, Oban, Argyll and Bute, Scotland	Great Britain	4000-3300	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I7554	Unstan, Orkney, Scotland	Great Britain	3366-3103	Scotland_N	GMN	UK_MN	Olalde *et al.* 2018
I5358	Rhos Ddigre, Llanarmon-yn-Iâl, Denbighshire, Wales	Great Britain	3080-2904	Wales_N	GMN	UK_MN	Olalde *et al.* 2018
I5359	Tinkinswood, Cardiff, Glamorgan, Wales	Great Britain	3800-3600	Wales_N	GMN	UK_MN	Olalde *et al.* 2018
I1970	Verdelha dos Ruivos, District of Lisbon	Portugal	2700-2300	BK_Portugal_POR	IBB	Iberia_LN	Olalde *et al.* 2018

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I4229	Cova da Moura, Torres Vedras	Portugal	2336-2063	BK_Portugal_POR	IBB	Iberia_LN	Olalde *et al.* 2018
I6467	Verdelha dos Ruivos, District of Lisbon	Portugal	2700-2300	BK_Portugal_POR	IBB	Iberia_LN	Olalde *et al.* 2018
I0459	Arroyal I, Burgos	Spain	2600-2200	BK_Spain_BUR1	IBB	Iberia_LN	Olalde *et al.* 2018
I0460	Arroyal I, Burgos	Spain	2460-2209	BK_Spain_BUR1	IBB	Iberia_LN	Olalde *et al.* 2018
I0461	Arroyal I, Burgos	Spain	2455-2201	BK_Spain_BUR2	IBB	Iberia_LN	Olalde *et al.* 2018
I0462	Arroyal I, Burgos	Spain	2566-2345	BK_Spain_BUR2	IBB	Iberia_LN	Olalde *et al.* 2018
I5665	Virgazal, Tablada de Rudrón, Burgos	Spain	2280-1984	BK_Spain_BUR2	IBB	Iberia_LN	Olalde *et al.* 2018
I0257	Paris Street, Cerdanyola, Barcelona	Spain	2572-2348	BK_Spain_Cer	IBB	Iberia_LN	Olalde *et al.* 2018
I0258	Paris Street, Cerdanyola, Barcelona	Spain	2850-2250	BK_Spain_Cer	IBB	Iberia_LN	Olalde *et al.* 2018
I0260	Paris Street, Cerdanyola, Barcelona	Spain	2850-2250	BK_Spain_Cer	IBB	Iberia_LN	Olalde *et al.* 2018
I0261	Paris Street, Cerdanyola, Barcelona	Spain	2850-2250	BK_Spain_Cer	IBB	Iberia_LN	Olalde *et al.* 2018
I0262	Paris Street, Cerdanyola, Barcelona	Spain	2850-2250	BK_Spain_Cer	IBB	Iberia_LN	Olalde *et al.* 2018
I0263	Paris Street, Cerdanyola, Barcelona	Spain	2850-2250	BK_Spain_Cer	IBB	Iberia_LN	Olalde *et al.* 2018
I0823	Paris Street, Cerdanyola, Barcelona	Spain	2850-2250	BK_Spain_Cer	IBB	Iberia_LN	Olalde *et al.* 2018
I0825	Paris Street, Cerdanyola, Barcelona	Spain	2474-2298	BK_Spain_Cer	IBB	Iberia_LN	Olalde *et al.* 2018
I0826	Paris Street, Cerdanyola, Barcelona	Spain	2834-2482	BK_Spain_Cer	IBB	Iberia_LN	Olalde *et al.* 2018
I4245	Camino de las Yeseras, Madrid	Spain	2460-2291	BK_Spain_MAD1	IBB	Iberia_LN	Olalde *et al.* 2018
I4247	Camino de las Yeseras, Madrid	Spain	2464-2210	BK_Spain_MAD1	IBB	Iberia_LN	Olalde *et al.* 2018
I6475	La Magdalena, Madrid	Spain	2500-2000	BK_Spain_MAD1	IBB	Iberia_LN	Olalde *et al.* 2018
I6542	Camino de las Yeseras, Madrid	Spain	2500-1750	BK_Spain_MAD1	IBB	Iberia_LN	Olalde *et al.* 2018
I6584	Humanejos, Madrid	Spain	2500-2000	BK_Spain_MAD1	IBB	Iberia_LN	Olalde *et al.* 2018
I6587	Humanejos, Madrid	Spain	2500-2000	BK_Spain_MAD1	IBB	Iberia_LN	Olalde *et al.* 2018
I6472	La Magdalena, Madrid	Spain	2500-2000	BK_Spain_MAD2	IBB	Iberia_LN	Olalde *et al.* 2018
I6539	Humanejos, Madrid	Spain	2500-2000	BK_Spain_MAD2	IBB	Iberia_LN	Olalde *et al.* 2018
I6588	Humanejos, Madrid	Spain	2500-2000	BK_Spain_MAD2	IBB	Iberia_LN	Olalde *et al.* 2018
I6623	Camino de las Yeseras, Madrid	Spain	1956-1743	BK_Spain_MAD2	IBB	Iberia_LN	Olalde *et al.* 2018
I6471	La Magdalena, Madrid	Spain	2500-2000	BK_Spain_Mag1	IBB	Iberia_LN	Olalde *et al.* 2018
I6543	Camino de las Yeseras, Madrid	Spain	2479-1945	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6596	Humanejos, Madrid	Spain	2900-2300	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6604	Camino de las Yeseras, Madrid	Spain	2127-1905	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6605	Camino de las Yeseras, Madrid	Spain	2479-2287	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6608	Camino de las Yeseras, Madrid	Spain	2020-1768	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6609	Camino de las Yeseras, Madrid	Spain	2436-2143	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6612	Camino de las Yeseras, Madrid	Spain	2479-1945	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6613	Camino de las Yeseras, Madrid	Spain	2479-1945	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6617	Humanejos, Madrid	Spain	2900-2300	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6628	Humanejos, Madrid	Spain	2900-2300	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6629	Humanejos, Madrid	Spain	2900-2300	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6630	Humanejos, Madrid	Spain	2900-2300	C_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I0453	Camino del Molino, Caravaca, Murcia	Spain	2458-2147	SE_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I0455	Camino del Molino, Caravaca, Murcia	Spain	2904-2667	SE_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018

Sample ID	Site	Country	Dates (years BCE)	Published label	PCA mitogenomes label	PCA genomes label	Reference
I0456	Camino del Molino, Caravaca, Murcia	Spain	2920-2340	SE_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I0457	Camino del Molino, Caravaca, Murcia	Spain	2920-2340	SE_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I6601	Bolores, Estremadura	Portugal	2800-2600	SW_Iberia_CA	ILNCA	Iberia_LN	Olalde *et al.* 2018
I1390	Sierentz, Les Villas d'Aurele, Haut-Rhin	France	2480-2210 [2480	BK_France_HAR	NFBB	Not included	Olalde *et al.* 2018
I1391	Rouffach, Haut-Rhin	France	2396-2060	BK_France_HAR	NFBB	Not included	Olalde *et al.* 2018
I1392	Hégenheim Necropole, Haut-Rhin	France	2833-2475	BK_France_Heg	NFBB	Not included	Olalde *et al.* 2018
I1381	Mondelange, PAC de la Sente, Moselle	France	2400-1900	BK_France_Mon	NFBB	Not included	Olalde *et al.* 2018
I1382	Mondelange, PAC de la Sente, Moselle	France	2434-2135	BK_France_Mon	NFBB	Not included	Olalde *et al.* 2018
I1388	Marlens, Sur les Barmes, Haute-Savoie	France	2455-2134	BK_France_Mar	SFBB	South_France_LN	Olalde *et al.* 2018
I2575	La Fare, Forcalquier	France	2475-2210	BK_France_AHP	SFBB	South_France_LN	Olalde *et al.* 2018
I3874	Villard, Lauzet-Ubaye	France	2200-2035	BK_France_AHP	SFBB	South_France_LN	Olalde *et al.* 2018
I3875	Villard, Lauzet-Ubaye	France	2133-1946	BK_France_AHP	SFBB	South_France_LN	Olalde *et al.* 2018
I4303	Clos de Roque, Saint Maximin-la-Sainte-Baume	France	4778-4586	France_MLN	SFMN	South_France_EMN	Olalde *et al.* 2018
I4304	Clos de Roque, Saint Maximin-la-Sainte-Baume	France	4787-4589	France_MLN	SFMN	South_France_EMN	Olalde *et al.* 2018
I4305	Clos de Roque, Saint Maximin-la-Sainte-Baume	France	4825-4616	France_MLN	SFMN	South_France_EMN	Olalde *et al.* 2018
I4308	Collet Redon, La Couronne-Martigues	France	3501-3112	France_MLN	SFMN	South_France_LMN	Olalde *et al.* 2018
V18_atp005	El Portalón, Atapuerca	Spain	-	Iberia_EN	IEN	Not included	Valdiosera *et al.* 2018
V18_atp019	El Portalón, Atapuerca	Spain	-	Iberia_EN	IEN	Not included	Valdiosera *et al.* 2018
V18_mur	Murciélagos de Zuheros, Andalusia	Spain	-	Iberia_EN	IEN	Not included	Valdiosera *et al.* 2018
atp002	El Portalón, Atapuerca	Spain	-	Iberia_CA	ILNCA	Not included	Valdiosera *et al.* 2018
atp016	El Portalón, Atapuerca	Spain	-	Iberia_CA	ILNCA	Not included	Valdiosera *et al.* 2018
atp12-1420	El Portalón, Atapuerca	Spain	-	Iberia_CA	ILNCA	Not included	Valdiosera *et al.* 2018
por002	El Portalón, Atapuerca	Spain	-	Iberia_CA	ILNCA	Not included	Valdiosera *et al.* 2018
por004	El Portalón, Atapuerca	Spain	-	Iberia_CA	ILNCA	Not included	Valdiosera *et al.* 2018
c40331	Cueva de los Cuarenta, Andalusia	Spain	-	Iberia_LN	ILNCA	Not included	Valdiosera *et al.* 2018
san216	San Quílez, Basque country	Spain	-	Iberia_LN	ILNCA	Not included	Valdiosera *et al.* 2018

Supplementary Table 7: Sex assignment based on shotgun sequencing data for ancient French individuals

Sample	Nseqs	NchrY+NchrX	NchrY	R_y	SE	95% CI	Assignment
BERG02-3	10857	381	13	0.0341	0.0093	0.0159-0.0523	consistent with XX but not XY
BERG79	7736	299	4	0.0134	0.0066	0.0004-0.0264	consistent with XX but not XY
CCF315-2-Lib1	1453	80	1	0.0125	0.0124	-0.0118-0.0368	consistent with XX but not XY
CLR13	1978	60	1	0.0167	0.0165	-0.0157-0.0491	consistent with XX but not XY
Es637-1-Lib2	10664	460	6	0.013	0.0053	0.0027-0.0234	consistent with XX but not XY
MOR4	12916	529	5	0.0095	0.0042	0.0012-0.0177	consistent with XX but not XY
BERG157-1	11355	316	27	0.0854	0.0157	0.0546-0.1163	consistent with XY but not XX
BERG157-3	21029	617	50	0.081	0.011	0.0595-0.1026	consistent with XY but not XX
BERG61	7288	161	14	0.087	0.0222	0.0434-0.1305	consistent with XY but not XX
CLR01	27690	741	60	0.081	0.01	0.0613-0.1006	consistent with XY but not XX
CLR11	6393	182	22	0.1209	0.0242	0.0735-0.1682	consistent with XY but not XX
CLR12	4765	162	11	0.0679	0.0198	0.0292-0.1066	consistent with XY but not XX
Es1018-16-Lib1	16656	464	34	0.0733	0.0121	0.0496-0.097	consistent with XY but not XX
Es42-3-Lib2	3127	80	9	0.1125	0.0353	0.0433-0.1817	consistent with XY but not XX
Es48bis-Lib1	15592	438	43	0.0982	0.0142	0.0703-0.126	consistent with XY but not XX
Es90-9-Lib1	6998	161	15	0.0932	0.0229	0.0483-0.1381	consistent with XY but not XX
Es97-1-Lib1	35867	897	77	0.0858	0.0094	0.0675-0.1042	consistent with XY but not XX
EUG1	38933	939	86	0.0916	0.0094	0.0731-0.11	consistent with XY but not XX
EUG11	142598	3744	297	0.0793	0.0044	0.0707-0.088	consistent with XY but not XX
MDV272-Lib2	9502	265	18	0.0679	0.0155	0.0376-0.0982	consistent with XY but not XX
MIT10-2032	10646	274	23	0.0839	0.0168	0.0511-0.1168	consistent with XY but not XX
MIT1059	15062	422	36	0.0853	0.0136	0.0587-0.112	consistent with XY but not XX
MOR54	4939	154	12	0.0779	0.0216	0.0356-0.1203	consistent with XY but not XX
NOR3-13	19335	575	49	0.0852	0.0116	0.0624-0.108	consistent with XY but not XX
OBE3626-2	11769	339	23	0.0678	0.0137	0.0411-0.0946	consistent with XY but not XX
OBE3628	29698	751	57	0.0759	0.0097	0.057-0.0948	consistent with XY but not XX
PECH3	67330	1932	169	0.0875	0.0064	0.0749-0.1001	consistent with XY but not XX
PEI2	14426	416	42	0.101	0.0148	0.072-0.1299	consistent with XY but not XX
Pey74	16123	509	38	0.0747	0.0116	0.0518-0.0975	consistent with XY but not XX
PIR3037AB	38702	1022	88	0.0861	0.0088	0.0689-0.1033	consistent with XY but not XX
ROS66	53420	1349	121	0.0897	0.0078	0.0744-0.1049	consistent with XY but not XX
Att52-1	6486	298	12	0.0403	0.0114	0.0179-0.0626	Not Assigned
BERG643-1	4392	144	7	0.0486	0.0179	0.0135-0.0837	Not Assigned
BLP32-Lib2	116	6	1	0.1667	0.1521	-0.1315-0.4649	Not Assigned
BLP9LF-Lib1	63	2	1	0.5	0.3536	-0.193-1.193	Not Assigned
CLR05	1482	34	2	0.0588	0.0404	-0.0203-0.1379	Not Assigned
CLR6	213	4	1	0.25	0.2165	-0.1744-0.6744	Not Assigned
MDV188-4-Lib1	543	24	1	0.0417	0.0408	-0.0383-0.1216	Not Assigned
MDV563III-Lib2	106	5	1	0.2	0.1789	-0.1506-0.5506	Not Assigned
MDV563IV-Lib2	643	26	2	0.0769	0.0523	-0.0255-0.1794	Not Assigned
MDV93-2-Lib1	1031	15	1	0.0667	0.0644	-0.0596-0.1929	Not Assigned
MOR5	10015	358	17	0.0475	0.0112	0.0255-0.0695	Not Assigned
Pir2	211898	7540	303	0.0402	0.0023	0.0358-0.0446	Not Assigned
Red1161-64B	1762	49	3	0.0612	0.0342	-0.0059-0.1284	Not Assigned
Sad1	188897	7576	250	0.033	0.0021	0.029-0.037	Not Assigned
SCHW432	20145	644	30	0.0466	0.0083	0.0303-0.0629	Not Assigned
SCHW9	2712	76	4	0.0526	0.0256	0.0024-0.1028	Not Assigned
ATT26	417638	20655	52	0.0025	0.0003	0.0018-0.0032	XX
BERG02-2	7284	375	2	0.0053	0.0038	-0.002-0.0127	XX
BERG02-5	11771	524	3	0.0057	0.0033	-0.0007-0.0122	XX
BERG118-1	12330	576	3	0.0052	0.003	-0.0007-0.0111	XX
BERG157-2	465570	22446	53	0.0024	0.0003	0.0017-0.003	XX
BERG157-6	262717	12934	32	0.0025	0.0004	0.0016-0.0033	XX
BES1154	30882	1585	4	0.0025	0.0013	0.0001-0.005	XX
BFM262	203490	9979	26	0.0026	0.0005	0.0016-0.0036	XX
BFM265	157873	7466	21	0.0028	0.0006	0.0016-0.004	XX
BIS358	28250	1360	5	0.0037	0.0016	0.0005-0.0069	XX
BIS385	57081	2653	5	0.0019	0.0008	0.0002-0.0035	XX
BIS3881	31639	1340	2	0.0015	0.0011	-0.0006-0.0036	XX
BLH447	52838	2550	8	0.0031	0.0011	0.001-0.0053	XX
COL11	129235	6349	13	0.002	0.0006	0.0009-0.0032	XX
COL153A	152743	7818	17	0.0022	0.0005	0.0011-0.0032	XX
COL153i	170880	7888	20	0.0025	0.0006	0.0014-0.0036	XX
COL336	114791	5825	9	0.0015	0.0005	0.0005-0.0026	XX
ERS1164	294404	14385	25	0.0017	0.0003	0.0011-0.0024	XX
ERS86	213774	10470	33	0.0032	0.0005	0.0021-0.0042	XX
ERS88	84674	4270	7	0.0016	0.0006	0.0004-0.0029	XX
Es168-Lib2	10572	494	3	0.0061	0.0035	-0.0008-0.0129	XX
Es278-1-Lib1	29750	1362	2	0.0015	0.001	-0.0006-0.0035	XX
Es454-1-Lib1	21539	1026	3	0.0029	0.0017	-0.0004-0.0062	XX

Sample	Nseqs	NchrY+NchrX	NchrY	R_y	SE	95% CI	Assignment
Es48-8C-Lib2	69556	2982	3	0.001	0.0006	-0.0001-0.0021	XX
Es626-27-Lib1	53802	2659	9	0.0034	0.0011	0.0012-0.0056	XX
EUG10	22849	1071	4	0.0037	0.0019	0.0001-0.0074	XX
EUG12	123647	6063	14	0.0023	0.0006	0.0011-0.0035	XX
EUG8	126268	6114	16	0.0026	0.0007	0.0013-0.0039	XX
MIT1030B	36514	1754	4	0.0023	0.0011	0.0-0.0045	XX
MIT1031	64712	3030	4	0.0013	0.0007	0.0-0.0026	XX
MIT11	37659	1776	6	0.0034	0.0014	0.0007-0.0061	XX
MIT1167A	36081	1759	5	0.0028	0.0013	0.0004-0.0053	XX
MIT1167B	34452	1674	6	0.0036	0.0015	0.0007-0.0064	XX
MIT1167C	54556	2504	4	0.0016	0.0008	0.0-0.0032	XX
MOR6	431704	19784	53	0.0027	0.0004	0.002-0.0034	XX
NHI13-Lib1	57492	2511	2	0.0008	0.0006	-0.0003-0.0019	XX
NIED	192767	9328	14	0.0015	0.0004	0.0007-0.0023	XX
NOR2B6	201729	9277	29	0.0031	0.0006	0.002-0.0043	XX
NOR3-10	83178	3802	12	0.0032	0.0009	0.0014-0.0049	XX
NOR3-15	132291	5967	19	0.0032	0.0007	0.0018-0.0046	XX
NOR36	231643	11428	25	0.0022	0.0004	0.0013-0.003	XX
NOR4	153070	7733	22	0.0028	0.0006	0.0017-0.004	XX
OBE3722	250059	11344	26	0.0023	0.0004	0.0014-0.0032	XX
PECH9	82585	4216	9	0.0021	0.0007	0.0007-0.0035	XX
PEI10	23332	1155	1	0.0009	0.0009	-0.0008-0.0026	XX
PER1150	50417	2441	6	0.0025	0.001	0.0005-0.0044	XX
PER1150B	69281	3143	7	0.0022	0.0008	0.0006-0.0039	XX
PER1150C	151685	7248	21	0.0029	0.0006	0.0017-0.0041	XX
PER26	62848	3193	6	0.0019	0.0008	0.0004-0.0034	XX
PER3023	40606	1919	4	0.0021	0.001	0.0-0.0041	XX
PER503	107526	5198	18	0.0035	0.0008	0.0019-0.0051	XX
PEY53	155471	7446	20	0.0027	0.0006	0.0015-0.0039	XX
PEY73	380749	18555	41	0.0022	0.0003	0.0015-0.0029	XX
Pir1	62378	2866	13	0.0045	0.0013	0.0021-0.007	XX
Pir3	150293	7246	38	0.0052	0.0008	0.0036-0.0069	XX
Pir4	20567	968	2	0.0021	0.0015	-0.0008-0.0049	XX
Pir6	274345	13357	38	0.0028	0.0005	0.0019-0.0037	XX
Pir7	96000	4674	20	0.0043	0.001	0.0024-0.0062	XX
PT2	134598	6816	15	0.0022	0.0006	0.0011-0.0033	XX
PT7	97720	4967	13	0.0026	0.0007	0.0012-0.004	XX
QUIN234	105857	5269	11	0.0021	0.0006	0.0009-0.0033	XX
QUIN59	53353	2481	6	0.0024	0.001	0.0005-0.0044	XX
RIX3	22792	1037	3	0.0029	0.0017	-0.0004-0.0062	XX
RIX4	113791	5502	10	0.0018	0.0006	0.0007-0.0029	XX
RIX8	92924	4406	11	0.0025	0.0008	0.001-0.004	XX
ROS108	30301	1401	2	0.0014	0.001	-0.0005-0.0034	XX
ROS111	233175	10965	27	0.0025	0.0005	0.0015-0.0034	XX
ROS31	359611	16925	39	0.0023	0.0004	0.0016-0.003	XX
ROS36	340834	16196	34	0.0021	0.0004	0.0014-0.0028	XX
ROS42	438470	20718	60	0.0029	0.0004	0.0022-0.0036	XX
ROS43	155191	7219	29	0.004	0.0007	0.0026-0.0055	XX
ROS82	383833	18234	47	0.0026	0.0004	0.0018-0.0033	XX
ROS86	82816	3949	11	0.0028	0.0008	0.0011-0.0044	XX
ATT13	84358	2510	238	0.0948	0.0058	0.0834-0.1063	XY
BERG157-5	312868	8486	732	0.0863	0.003	0.0803-0.0922	XY
BERG157-7	357588	9555	839	0.0878	0.0029	0.0821-0.0935	XY
BERG157-9	296909	8069	733	0.0908	0.0032	0.0846-0.0971	XY
Berg34-1	174416	4393	394	0.0897	0.0043	0.0812-0.0981	XY
Berg34-2	249697	6478	641	0.099	0.0037	0.0917-0.1062	XY
BES1096B	83606	2282	214	0.0938	0.0061	0.0818-0.1057	XY
BES1248	144977	4001	353	0.0882	0.0045	0.0794-0.097	XY
BES1249	121528	3230	333	0.1031	0.0054	0.0926-0.1136	XY
BIS130	81605	1954	177	0.0906	0.0065	0.0779-0.1033	XY
BIS159	94253	2470	223	0.0903	0.0058	0.079-0.1016	XY
BIS382	113472	3107	315	0.1014	0.0054	0.0908-0.112	XY
BIS3882	74881	2137	198	0.0927	0.0063	0.0804-0.1049	XY
BLP10	40817	1072	107	0.0998	0.0092	0.0819-0.1178	XY
CBV95	389306	11123	1019	0.0916	0.0027	0.0863-0.097	XY
ERS83-2	27686	759	89	0.1173	0.0117	0.0944-0.1401	XY
MAND1175	69681	1930	185	0.0959	0.0067	0.0827-0.109	XY
MDV248-Lib1	63823	1702	154	0.0905	0.007	0.0769-0.1041	XY
NOR2B2	84065	2269	206	0.0908	0.006	0.079-0.1026	XY
NOR2B20	18963	529	54	0.1021	0.0132	0.0763-0.1279	XY
NOR5-2	304014	7675	717	0.0934	0.0033	0.0869-0.0999	XY
OBE3626-1	351025	8918	782	0.0877	0.003	0.0818-0.0936	XY
PECH10	17961	513	52	0.1014	0.0133	0.0752-0.1275	XY

Sample	Nseqs	NchrY+NchrX	NchrY	R_y	SE	95% CI	Assignment
PECH5	177234	4770	436	0.0914	0.0042	0.0832-0.0996	XY
PECH8	85473	2352	209	0.0889	0.0059	0.0774-0.1004	XY
PER3123	125067	3466	309	0.0892	0.0048	0.0797-0.0986	XY
PEY163	160624	4359	424	0.0973	0.0045	0.0885-0.1061	XY
PIR3116B	76258	2130	192	0.0901	0.0062	0.078-0.1023	XY
PIR3117	28103	768	74	0.0964	0.0106	0.0755-0.1172	XY
QUIN58	195976	5396	515	0.0954	0.004	0.0876-0.1033	XY
Red1161-1528	13	1	1	1.0	0.0	1.0-1.0	XY
RIX15	98011	2761	287	0.1039	0.0058	0.0926-0.1153	XY
RIX2	111694	3182	288	0.0905	0.0051	0.0805-0.1005	XY
ROS100	355330	9214	812	0.0881	0.003	0.0823-0.0939	XY
ROS102	329491	8815	770	0.0874	0.003	0.0815-0.0932	XY
ROS45	367704	9509	769	0.0809	0.0028	0.0754-0.0864	XY
ROS47	221603	6046	545	0.0901	0.0037	0.0829-0.0974	XY
ROS62	330507	9021	825	0.0915	0.003	0.0855-0.0974	XY
ROS78	277794	7582	730	0.0963	0.0034	0.0896-0.1029	XY
Sad4	249191	7052	710	0.1007	0.0036	0.0937-0.1077	XY
Schw72-15	32262	864	87	0.1007	0.0102	0.0806-0.1208	XY
SCHW72-17	224897	6120	526	0.0859	0.0036	0.0789-0.093	XY
WET370-1	178097	4836	440	0.091	0.0041	0.0829-0.0991	XY

Supplementary Table 8: Summary of the result obtained through shotgun sequencing on 58 ancient French individuals

Label ind.	Label PCA and D statistics	Label f3 statistics	Date BP	Date BC	Total	Mapped Unique	Genome coverage	% SNPs	mtDNA	Y chr	Sext	Contam. 1	Contam. 2*
PER1150C	South_France_HG	Mesolitic (Southern) France		7177-7057	8630223	5189611	0.130	0.184	U5b1-16189		XX		
PER3023	South_France_HG	Mesolitic (Southern) France		7177-7057	8978029	5343431	0.134	0.181	U5b3e		XX		
PER3123	South_France_HG	Mesolitic (Southern) France		7177-7057	7504991	4421014	0.111	0.157	U5b1-16189	I	XY	6.61E-05	6.61E-05
PER503	South_France_HG	Mesolitic (Southern) France		7177-7057BC	9685192	5558692	0.139	0.191	U5b1f1a		XX		
Sch72-15	North_France_EN	Early Neolithic (Northern) France	6340 +/-40BP	5383-5220 cal	14204935	7746223	0.194	0.240	T2f	C1a2b	XY	6.61E-05	6.61E-05
Schw432	North_France_EN	Early Neolithic (Northern) France	6130 +/-40BP	5211-4962 cal	14496411	5217845	0.130	0.166	X2b		XX		
MDV248	North_France_EN	Early Neolithic (Northern) France			18833468	3785925	0.095	0.144	K1a12	H2a1	XY	6.61E-05	6.61E-05
Mor6	North_France_EN	Early Neolithic (Northern) France Outlier		5200-5000	9826636	5524198	0.138	0.172	H3q1		XX		
ROS102	North_France_MN	Middle Neolithic France		4700-4500	7036232	4971229	0.124	0.171	K1a1a	I2a1b	XY	6.61E-05	0.02230484
ROS45	North_France_MN	Middle Neolithic France	5833 +/-29BP	4730-4655	11762977	8298657	0.207	0.270	H5u	I2a1b	XY	0.001978663	6.61E-05
ROS78	North_France_MN	Middle Neolithic France		4700-4500	16758668	12314751	0.308	0.387	U8b1b1	I2a1b	XY	0.000970655	0.00442525
ROS82	North_France_MN	Middle Neolithic France	5816 +/-28BP	4725-4640	7317897	3669225	0.092	0.112	T2f7		XX		
BUCH2	North_France_MN	Middle Neolithic France		4400-4200	41902633	13221049	0.331	0.355	U5b1-16189	H2a1	XY	0.01606251	0.009249083
CRE20D	South_France_MN	Middle Neolithic France		4400-4200	35973524	10442261	0.261	0.272	H-16291		XX		
Pir4	South_France_MN	Middle Neolithic France	5494 +/-49 BP	4450-4173	13430343	3411025	0.085	0.085	V10a		XX		
BERG02-2	North_France_MN	Middle Neolithic France	5135 +/-35BP	4037-3803	27870147	11918249	0.298	0.330	W5b		XX		
BERG157-2	North_France_MN	Middle Neolithic France		4300-3900	14072235	10250332	0.256	0.313	J1c1b		XX		
BERG157-7	North_France_MN	Middle Neolithic France	5335 +/-35BP	4315-4048	13849797	8395221	0.210	0.256	U5b1c	I2a1b	XY	0.00075613	6.61E-05
BLP10	North_France_MN	Middle Neolithic France			21195766	7100652	0.178	0.210	H1e	I	XY	6.61E-05	6.61E-05

Label ind.	Label PCA and D statistics	Label f3 statistics	Date BP	Date BC	Total	Mapped Unique	Genome coverage	% SNPs	mtDNA	Y chr	Sext	Contam. 1	Contam. 2*
Es97-1	North_France_MN	Middle Neolithic France	4230-3879 BCE	4230-3879	2239262	1002152 1	0.251	0.292	K1a26	I	XY	0.005714 086	6.61E-05
PSS4693	North_France_MN	Middle Neolithic France	4650 +/- 45BP	3626-3351	4010856 7	2303920 9	0.576	0.568	U5b3b		XX		
CBV95	North_France_LN	Bell Beaker Aisne	3970 +/- 30BP	2574-2452	8245595	5276451	0.132	0.179	J2a1a1	R1b1a1a2	XY	0.008331 817	0.01744 863
PEI10	South_France_LN	Bell Beaker Aude	4200 +- 30 BP	2894-2678	1617705 3	2097428	0.052	0.064			XX		
BIS130	North_France_BA	Bronze Age France		2000-1800	8544634	3805484	0.095	0.149	K2b1b	R1b1a1a2	XY	6.61E-05	6.61E-05
BIS385	North_France_BA	Bronze Age France		2000-1800	1056730 9	6594540	0.165	0.204	T2c1d2a		XX		
EUG11	South_France_BA	Bronze Age France	3580 +/- 30BP	2028-1878	3770944 0	5384029	0.135	0.164	V9	R1b1a2a	XY	0.012013 05	0.01037 555
NIED	North_France_BA	Bronze Age France		914-810	6531228	4246651	0.106	0.144	U5b2b		XX		
OBE3626-1	North_France_BA	Bronze Age France	3505 ± 35 BP	1926-1701	9680492	7301781	0.183	0.251	R1b	R1b1a1a2	XY	0.000568 833	6.61E-05
OBE3722	North_France_BA	Bronze Age France		2000-1700	8491464	5263626	0.132	0.192	I5		XX		
PIR3037AB	North_France_BA	Bronze Age France	3663 +/- 43BP		8517422	2266752	0.057	0.070	U4c1	R1b1a1a2	XY	0.059825 9	6.61E-05
PIR3116B	South_France_BA	Bronze Age France	3790 +- 30 BP	2336-2135	9878183	3785048	0.095	0.121	H2a1e	R1b1a1a2	XY	0.023224 47	6.61E-05
Pir6	South_France_BA	Bronze Age France	*3356+/- 43BP*		*7767384*	*3317293*	*0.083*	0.085			*XX*		
PSS4170	South_France_BA	Bronze Age France	3690 +/- 30BP	2196-1977	4209338 2	1264212 5	0.316	0.368	H25	R1b1a1a2 a2a	XY	0.002240 537	0.00784 9955
QUIN234	South_France_BA	Bronze Age France		2100-1200	1309251 8	5103721	0.128	0.152	H3m		XX		
QUIN58	South_France_BA	Bronze Age France		2100-1200	9454812	3732016	0.093	0.132	H1ah	R1b1a1a2	XY	6.61E-05	6.61E-05
RIX15	North_France_BA	Bronze Age France		1700-1600	7260014	4922963	0.123	0.176	T2b19	R1b1a1a2 a1a	XY	6.61E-05	6.61E-05
RIX2	North_France_BA	Bronze Age France		1700-1600	1236372 8	5944314	0.149	0.204	T2c1d-152	R1b1a1a2 a1a	XY	0.005517 262	0.01306 447
RIX4	North_France_BA	Bronze Age France	3415+- 40BP	1750-1630	8379379	5851228	0.146	0.213	K1a4a1a1		XX		
ATT26	North_France_IA	Bronze Age France		400-100	5545147	3688046	0.092	0.134	H8		XX		
BES1248	South_France_IA	Iron Age France		600-400	7907786	4948778	0.124	0.162	J1c8a2	I1	XY	0.009846 313	6.61E-05
BFM265	North_France_IA	Iron Age France		250-190	1156366 5	6524970	0.163	0.212	U2e1b2		XX		

Label ind.	Label PCA and D statistics	Label f3 statistics	Date BP	Date BC	Total	Mapped Unique	Genome coverage	% SNPs	mtDNA	Y chr	Sext	Contam. 1	Contam. 2*
COL11	North_France_IA	Iron Age France			11322871	4231690	0.106	0.144	H2b		XX		
COL153A	North_France_IA	Iron Age France		360-320	10928453	5076619	0.127	0.158	K1a26		XX		
COL153i	North_France_IA	Iron Age France		360-320	6753440	3209015	0.080	0.114	H1q		XX		
ERS1164	North_France_IA	Iron Age France		400-100	9852764	6362026	0.159	0.203	H2a2a1		XX		
ERS86	North_France_IA	Iron Age France		400-100	13120776	6185307	0.155	0.196	HV0		XX		
ERS88	North_France_IA	Iron Age France		400-100	5974275	3953068	0.099	0.137	U5a1g		XX		
Jeb8	North_France_IA	Iron Age France		600-400	8738720	4555585	0.114	0.144	J1c8a		XX		
NOR2B6	North_France_IA	Iron Age France		800-700	9782279	5628522	0.141	0.207	K1a2a		XX		
NOR3-15	North_France_IA	Iron Age France		800-700	8036606	3692295	0.092	0.132	J1c2o		XX		
NOR3-6	North_France_IA	Iron Age France		620-500	8043152	5307105	0.133	0.173	H7d		XX		
NOR4	North_France_IA	Iron Age France		620-480	8007537	3423937	0.086	0.121	U5a1a1		XX		
PECH5	South_France_IA	Iron Age France		600-300 BC	2449026	1664074	0.042	0.064	H1e1b1a	R1b1a1a2a	XY	6.61E-05	6.61E-05
PECH8	South_France_IA	Iron Age France		600-300 BC	8100956	5063541	0.127	0.169	J1c2m	R1b1a1a2a1a	XY	0.01627477	0.006399301
PEY163	South_France_IA	Iron Age France		400-200	8162983	4130458	0.103	0.134	W1g	R1b1a1a2	XY	0.01169737	6.61E-05
PEY53	South_France_IA	Iron Age France		400-200	8641790	4853700	0.121	0.155	J1c11a		XX		
PT2	South_France_IA	Iron Age France		450-400	7504094	3933011	0.098	0.138	J1c1b1		XX		
WET370-1	North_France_IA	Iron Age France Outlier		480-300	9699916	3818672	0.095	0.131	H1c5a'	H2a1	XY	6.61E-05	6.61E-05

Supplementary Table 9:Values of D(Mbuti, ancient French individual; WHG, WHG) with combinations of Loschbour, La Braña and KO1 as WHG.

Outgroup	Ancient French individual	WHG	WHG2	D value	std.err	Z score	#SNPs
Mbuti	BERG02-2	Loschbour	KO1	-0.0157	0.009522	-1.646	153054
Mbuti	BERG02-2	La Braña	KO1	0.0197	0.009328	2.111	147373
Mbuti	BERG02-2	La Braña	Loschbour	0.0323	0.008301	3.889	192871
Mbuti	BERG157-2	Loschbour	KO1	-0.0192	0.009455	-2.036	145826
Mbuti	BERG157-2	La Braña	KO1	0.0044	0.009483	0.466	140492
Mbuti	BERG157-2	La Braña	Loschbour	0.0239	0.008926	2.678	183224
Mbuti	BERG157-7	Loschbour	KO1	-0.0066	0.010184	-0.647	117860
Mbuti	BERG157-7	La Braña	KO1	0.0306	0.010348	2.958	113540
Mbuti	BERG157-7	La Braña	Loschbour	0.0299	0.009369	3.189	150018
Mbuti	BLP10	Loschbour	KO1	0.0006	0.011121	0.057	98438
Mbuti	BLP10	La Braña	KO1	0.0353	0.009995	3.533	94849
Mbuti	BLP10	La Braña	Loschbour	0.0298	0.009788	3.044	123286
Mbuti	BUCH2	Loschbour	KO1	0.0043	0.008964	0.485	161589
Mbuti	BUCH2	La Braña	KO1	0.0230	0.008929	2.574	155809
Mbuti	BUCH2	La Braña	Loschbour	0.0181	0.008334	2.169	207854
Mbuti	CRE20D	Loschbour	KO1	-0.0023	0.010010	-0.226	123521
Mbuti	CRE20D	La Braña	KO1	0.0080	0.010041	0.793	119062
Mbuti	CRE20D	La Braña	Loschbour	0.0101	0.008871	1.142	159205
Mbuti	Es97-1	Loschbour	KO1	-0.0088	0.009652	-0.915	136097
Mbuti	Es97-1	La Braña	KO1	-0.0083	0.009764	-0.853	131115
Mbuti	Es97-1	La Braña	Loschbour	0.0029	0.008486	0.343	170939
Mbuti	MDV248	Loschbour	KO1	0.0020	0.012026	0.166	69209
Mbuti	MDV248	La Braña	KO1	-0.0003	0.012096	-0.025	66742
Mbuti	MDV248	La Braña	Loschbour	-0.0077	0.011135	-0.692	84572
Mbuti	Mor6	Loschbour	KO1	0.0040	0.011869	0.340	79526
Mbuti	Mor6	La Braña	KO1	0.0197	0.011238	1.751	76566
Mbuti	Mor6	La Braña	Loschbour	0.0188	0.010693	1.763	100794
Mbuti	O18_I4303	Loschbour	KO1	0.0032	0.007509	0.420	366609
Mbuti	O18_I4303	La Braña	KO1	0.0140	0.007348	1.903	355193
Mbuti	O18_I4303	La Braña	Loschbour	0.0129	0.007045	1.825	404925
Mbuti	O18_I4304	Loschbour	KO1	-0.0010	0.007012	-0.141	391378
Mbuti	O18_I4304	La Braña	KO1	0.0177	0.006886	2.565	378791
Mbuti	O18_I4304	La Braña	Loschbour	0.0175	0.006959	2.521	445827
Mbuti	O18_I4305	Loschbour	KO1	-0.0047	0.007124	-0.667	378283
Mbuti	O18_I4305	La Braña	KO1	0.0145	0.007054	2.053	366343
Mbuti	O18_I4305	La Braña	Loschbour	0.0171	0.007166	2.380	429001
Mbuti	PER1150C	Loschbour	KO1	-0.1043	0.011460	-9.105	89894
Mbuti	PER1150C	La Braña	KO1	0.0001	0.011313	0.012	86651
Mbuti	PER1150C	La Braña	Loschbour	0.0961	0.010953	8.777	107743
Mbuti	PER3023	Loschbour	KO1	-0.0974	0.012135	-8.027	86567
Mbuti	PER3023	La Braña	KO1	-0.0011	0.011867	-0.095	83468
Mbuti	PER3023	La Braña	Loschbour	0.0906	0.011391	7.950	106074
Mbuti	PER3123	Loschbour	KO1	-0.0952	0.012169	-7.827	78221
Mbuti	PER3123	La Braña	KO1	-0.0083	0.011896	-0.698	75409
Mbuti	PER3123	La Braña	Loschbour	0.0904	0.011307	7.998	92275
Mbuti	PER503	Loschbour	KO1	-0.0850	0.010898	-7.801	93320
Mbuti	PER503	La Braña	KO1	-0.0019	0.011249	-0.170	89968
Mbuti	PER503	La Braña	Loschbour	0.0835	0.010921	7.644	112049
Mbuti	Pir4	Loschbour	KO1	-0.0022	0.015994	-0.141	37522
Mbuti	Pir4	La Braña	KO1	0.0038	0.016665	0.230	36192
Mbuti	Pir4	La Braña	Loschbour	0.0009	0.013434	0.065	49565
Mbuti	PSS4693	Loschbour	KO1	-0.0173	0.007622	-2.267	260544
Mbuti	PSS4693	La Braña	KO1	-0.0009	0.007726	-0.122	251019
Mbuti	PSS4693	La Braña	Loschbour	0.0142	0.006934	2.044	331941
Mbuti	ROS102	Loschbour	KO1	0.0019	0.011835	0.163	80364
Mbuti	ROS102	La Braña	KO1	0.0165	0.011775	1.404	77467
Mbuti	ROS102	La Braña	Loschbour	0.0193	0.010447	1.843	100152
Mbuti	ROS45	Loschbour	KO1	-0.0032	0.009902	-0.324	126727
Mbuti	ROS45	La Braña	KO1	0.0138	0.009552	1.449	121975
Mbuti	ROS45	La Braña	Loschbour	0.0110	0.008886	1.244	157920
Mbuti	ROS78	Loschbour	KO1	0.0171	0.008191	2.085	180764
Mbuti	ROS78	La Braña	KO1	0.0225	0.008383	2.685	174201

Outgroup	Ancient French individual	WHG	WHG2	D value	std.err	Z score	#SNPs
Mbuti	ROS78	La Braña	Loschbour	0.0052	0.007917	0.662	226447
Mbuti	ROS82	Loschbour	KO1	0.0083	0.013734	0.606	51390
Mbuti	ROS82	La Braña	KO1	0.0177	0.013693	1.294	49460
Mbuti	ROS82	La Braña	Loschbour	0.0003	0.011974	0.025	65484
Mbuti	Sch72-15	Loschbour	KO1	-0.0044	0.010240	-0.426	110514
Mbuti	Sch72-15	La Braña	KO1	0.0056	0.010382	0.536	106378
Mbuti	Sch72-15	La Braña	Loschbour	0.0050	0.009222	0.538	140202
Mbuti	Schw432	Loschbour	KO1	0.0222	0.011779	1.883	77566
Mbuti	Schw432	La Braña	KO1	0.0150	0.011799	1.268	74734
Mbuti	Schw432	La Braña	Loschbour	-0.0054	0.011137	-0.481	97254

Supplementary Table 10: Values of D(Mbuti, ancient French individual; Early Neolithic farmer, Early Neolithic farmer) with combinations of Hungary_EN, LBK_EN, LBKT_EN and Iberia_EN as Early Neolithic farmers.

Outrgroup	AncientFrench	EarlyNeolithic1	EarlyNeolithic2	Dvalue	std.err	Zscore	#SNPs
Mbuti	ATT26	Iberia_EN	LBK_EN	-0.0057	0.004478	-1.274	82580
Mbuti	BERG02-2	Iberia_EN	LBK_EN	-0.0036	0.003259	-1.107	203299
Mbuti	BERG157-2	Iberia_EN	LBK_EN	-0.0033	0.003364	-0.990	192937
Mbuti	BERG157-7	Iberia_EN	LBK_EN	-0.002	0.003445	-0.586	158062
Mbuti	BES1248	Iberia_EN	LBK_EN	-0.0022	0.004018	-0.541	99805
Mbuti	BFM265	Iberia_EN	LBK_EN	0.0004	0.003552	0.123	130855
Mbuti	BIS130	Iberia_EN	LBK_EN	0.0024	0.003961	0.614	91794
Mbuti	BIS385	Iberia_EN	LBK_EN	-0.0094	0.003784	-2.477	125561
Mbuti	BLP10	Iberia_EN	LBK_EN	-0.003	0.003786	-0.803	129735
Mbuti	BUCH2	Iberia_EN	LBK_EN	-0.0066	0.003264	-2.027	219021
Mbuti	CBV95	Iberia_EN	LBK_EN	-0.0003	0.003808	-0.084	110425
Mbuti	COL11	Iberia_EN	LBK_EN	-0.0045	0.004312	-1.044	88657
Mbuti	COL153A	Iberia_EN	LBK_EN	-0.0032	0.004139	-0.765	97422
Mbuti	COL153i	Iberia_EN	LBK_EN	0.0017	0.004546	0.385	70405
Mbuti	CRE20D	Iberia_EN	LBK_EN	-0.0103	0.003418	-3.009	167730
Mbuti	ERS86	Iberia_EN	LBK_EN	-0.0033	0.003812	-0.857	120713
Mbuti	ERS88	Iberia_EN	LBK_EN	-0.0073	0.004348	-1.677	84332
Mbuti	ERS1164	Iberia_EN	LBK_EN	0.0031	0.003726	0.842	125139
Mbuti	Es97-1	Iberia_EN	LBK_EN	-0.0008	0.00334	-0.239	179953
Mbuti	EUG11	Iberia_EN	LBK_EN	-0.0058	0.004044	-1.422	101012
Mbuti	Jeb8	Iberia_EN	LBK_EN	-0.0004	0.004388	-0.088	88632
Mbuti	MDV248	Iberia_EN	LBK_EN	0.0078	0.004403	1.782	88873
Mbuti	Mor6	Iberia_EN	LBK_EN	-0.0098	0.004058	-2.418	106251
Mbuti	NIED	Iberia_EN	LBK_EN	-0.0097	0.004101	-2.355	88844
Mbuti	NOR2B6	Iberia_EN	LBK_EN	-0.006	0.003697	-1.626	127623
Mbuti	NOR3-6	Iberia_EN	LBK_EN	-0.0023	0.003854	-0.584	106939
Mbuti	NOR3-15	Iberia_EN	LBK_EN	-0.0039	0.004332	-0.895	81341
Mbuti	NOR4	Iberia_EN	LBK_EN	-0.0002	0.00474	-0.047	74580
Mbuti	OBE3626-1	Iberia_EN	LBK_EN	-0.0003	0.003529	-0.094	154952
Mbuti	OBE3722	Iberia_EN	LBK_EN	-0.0051	0.003773	-1.357	118578
Mbuti	PECH5	Iberia_EN	LBK_EN	-0.0072	0.005718	-1.253	39367
Mbuti	PECH8	Iberia_EN	LBK_EN	-0.0053	0.003903	-1.368	104301
Mbuti	PEI10	Iberia_EN	LBK_EN	-0.0023	0.006092	-0.371	39458
Mbuti	PEY53	Iberia_EN	LBK_EN	-0.0073	0.004141	-1.768	95276
Mbuti	PEY163	Iberia_EN	LBK_EN	0.0004	0.004324	0.093	82828
Mbuti	Pir4	Iberia_EN	LBK_EN	-0.0075	0.005656	-1.328	52299
Mbuti	Pir6	Iberia_EN	LBK_EN	-0.0051	0.005405	-0.941	52528
Mbuti	PIR3037AB	Iberia_EN	LBK_EN	-0.0022	0.005634	-0.387	43373
Mbuti	PIR3116B	Iberia_EN	LBK_EN	-0.0071	0.004675	-1.511	74595
Mbuti	PSS4170	Iberia_EN	LBK_EN	-0.0009	0.003062	-0.295	227129
Mbuti	PSS4693	Iberia_EN	LBK_EN	-0.0005	0.002772	-0.163	349869
Mbuti	PT2	Iberia_EN	LBK_EN	0.0006	0.004287	0.150	84826
Mbuti	QUIN58	Iberia_EN	LBK_EN	-0.007	0.004444	-1.583	81306
Mbuti	QUIN234	Iberia_EN	LBK_EN	-0.007	0.004223	-1.664	93532
Mbuti	RIX2	Iberia_EN	LBK_EN	-0.0059	0.003809	-1.560	125685
Mbuti	RIX4	Iberia_EN	LBK_EN	-0.001	0.003551	-0.268	131506
Mbuti	RIX15	Iberia_EN	LBK_EN	0	0.00409	-0.010	108496
Mbuti	ROS45	Iberia_EN	LBK_EN	0.0007	0.003364	0.202	166261
Mbuti	ROS78	Iberia_EN	LBK_EN	0.0008	0.003107	0.269	238415
Mbuti	ROS82	Iberia_EN	LBK_EN	0.0071	0.004837	1.468	69045
Mbuti	ROS102	Iberia_EN	LBK_EN	-0.0058	0.004221	-1.383	105311
Mbuti	Sch72-15	Iberia_EN	LBK_EN	0.0094	0.00359	2.612	147779
Mbuti	Schw432	Iberia_EN	LBK_EN	0.0068	0.004355	1.571	102336
Mbuti	WET370-1	Iberia_EN	LBK_EN	-0.001	0.004611	-0.219	80599
Mbuti	O18_I3874	Iberia_EN	LBK_EN	-0.0007	0.002398	-0.311	368141
Mbuti	O18_I3875	Iberia_EN	LBK_EN	0.0028	0.004389	0.646	152190
Mbuti	O18_I4303	Iberia_EN	LBK_EN	-0.0089	0.002673	-3.347	420640
Mbuti	O18_I4304	Iberia_EN	LBK_EN	-0.0074	0.002676	-2.760	464109
Mbuti	O18_I4305	Iberia_EN	LBK_EN	-0.0079	0.002586	-3.045	446286
Mbuti	PECH5	Iberia_EN	Hungary_EN	-0.009	0.005912	-1.524	38931
Mbuti	PECH8	Iberia_EN	Hungary_EN	-0.0048	0.003937	-1.214	102911
Mbuti	PEI10	Iberia_EN	Hungary_EN	-0.0024	0.006316	-0.373	38814
Mbuti	PEY53	Iberia_EN	Hungary_EN	-0.0034	0.004409	-0.772	93449
Mbuti	PEY163	Iberia_EN	Hungary_EN	0.0015	0.004334	0.354	80929
Mbuti	Pir4	Iberia_EN	Hungary_EN	-0.0065	0.006051	-1.069	50574
Mbuti	Pir6	Iberia_EN	Hungary_EN	-0.0086	0.005894	-1.467	51120
Mbuti	PIR3037AB	Iberia_EN	Hungary_EN	0.0028	0.005862	0.471	42714
Mbuti	PIR3116B	Iberia_EN	Hungary_EN	-0.0053	0.00468	-1.136	73384

Outrgroup	AncientFrench	EarlyNeolithic1	EarlyNeolithic2	Dvalue	std.err	Zscore	#SNPs
Mbuti	PSS4170	Iberia_EN	Hungary_EN	0.0055	0.003222	1.697	220814
Mbuti	PSS4693	Iberia_EN	Hungary_EN	-0.002	0.002879	-0.711	339893
Mbuti	PT2	Iberia_EN	Hungary_EN	0	0.004489	0.003	83392
Mbuti	QUIN58	Iberia_EN	Hungary_EN	-0.0103	0.004495	-2.294	79937
Mbuti	QUIN234	Iberia_EN	Hungary_EN	-0.004	0.004391	-0.908	91609
Mbuti	RIX2	Iberia_EN	Hungary_EN	-0.0048	0.003843	-1.257	123314
Mbuti	RIX4	Iberia_EN	Hungary_EN	-0.0049	0.003577	-1.364	129488
Mbuti	RIX15	Iberia_EN	Hungary_EN	-0.0008	0.004186	-0.198	105869
Mbuti	ROS45	Iberia_EN	Hungary_EN	-0.0019	0.0035	-0.535	162231
Mbuti	ROS78	Iberia_EN	Hungary_EN	-0.0053	0.003329	-1.594	232615
Mbuti	ROS82	Iberia_EN	Hungary_EN	0.0114	0.00512	2.225	67123
Mbuti	ROS102	Iberia_EN	Hungary_EN	-0.0064	0.004434	-1.437	102909
Mbuti	Sch72-15	Iberia_EN	Hungary_EN	0.0048	0.003653	1.306	143620
Mbuti	Schw432	Iberia_EN	Hungary_EN	0.0029	0.004469	0.658	99741
Mbuti	WET370-1	Iberia_EN	Hungary_EN	0.0026	0.004586	0.558	79538
Mbuti	O18_I4303	Iberia_EN	Hungary_EN	-0.0017	0.002643	-0.652	419105
Mbuti	O18_I4304	Iberia_EN	Hungary_EN	-0.0018	0.002671	-0.679	461415
Mbuti	O18_I4305	Iberia_EN	Hungary_EN	-0.0034	0.002659	-1.290	443837
Mbuti	O18_I3874	Iberia_EN	Hungary_EN	0.0043	0.002501	1.716	367338
Mbuti	O18_I3875	Iberia_EN	Hungary_EN	0.0075	0.004059	1.848	151486
Mbuti	ATT26	Iberia_EN	Hungary_EN	-0.0065	0.004622	-1.399	81184
Mbuti	BERG02-2	Iberia_EN	Hungary_EN	-0.0012	0.003503	-0.338	197693
Mbuti	BERG157-2	Iberia_EN	Hungary_EN	-0.0088	0.003468	-2.524	187819
Mbuti	BERG157-7	Iberia_EN	Hungary_EN	-0.0026	0.003611	-0.717	153597
Mbuti	BES1248	Iberia_EN	Hungary_EN	-0.0003	0.004311	-0.064	97752
Mbuti	BFM265	Iberia_EN	Hungary_EN	0.0006	0.003895	0.144	128182
Mbuti	BIS130	Iberia_EN	Hungary_EN	0.004	0.004037	0.996	90921
Mbuti	BIS385	Iberia_EN	Hungary_EN	-0.0029	0.004015	-0.714	123122
Mbuti	BLP10	Iberia_EN	Hungary_EN	-0.0089	0.003895	-2.293	126750
Mbuti	BUCH2	Iberia_EN	Hungary_EN	-0.0084	0.00348	-2.405	212687
Mbuti	CBV95	Iberia_EN	Hungary_EN	-0.0028	0.004078	-0.694	108536
Mbuti	COL11	Iberia_EN	Hungary_EN	-0.0048	0.004472	-1.075	87210
Mbuti	COL153A	Iberia_EN	Hungary_EN	-0.0063	0.004417	-1.430	95849
Mbuti	COL153i	Iberia_EN	Hungary_EN	-0.0024	0.004658	-0.509	69753
Mbuti	CRE20D	Iberia_EN	Hungary_EN	-0.0115	0.003561	-3.222	162898
Mbuti	ERS86	Iberia_EN	Hungary_EN	-0.0008	0.004037	-0.203	118770
Mbuti	ERS88	Iberia_EN	Hungary_EN	-0.0023	0.004493	-0.509	83036
Mbuti	ERS1164	Iberia_EN	Hungary_EN	0.0031	0.003931	0.792	122716
Mbuti	Es97-1	Iberia_EN	Hungary_EN	-0.0057	0.003334	-1.720	175289
Mbuti	EUG11	Iberia_EN	Hungary_EN	-0.0023	0.004346	-0.520	97956
Mbuti	Jeb8	Iberia_EN	Hungary_EN	-0.0004	0.004572	-0.094	87448
Mbuti	MDV248	Iberia_EN	Hungary_EN	0.0066	0.004407	1.488	87159
Mbuti	Mor6	Iberia_EN	Hungary_EN	-0.0134	0.00432	-3.097	103335
Mbuti	NIED	Iberia_EN	Hungary_EN	-0.0095	0.004365	-2.180	87616
Mbuti	NOR2B6	Iberia_EN	Hungary_EN	-0.0013	0.003754	-0.356	125999
Mbuti	NOR3-6	Iberia_EN	Hungary_EN	-0.0006	0.004022	-0.145	105181
Mbuti	NOR3-15	Iberia_EN	Hungary_EN	-0.0084	0.004534	-1.853	80325
Mbuti	NOR4	Iberia_EN	Hungary_EN	0.0002	0.004765	0.052	73372
Mbuti	OBE3626-1	Iberia_EN	Hungary_EN	-0.0005	0.003625	-0.126	151568
Mbuti	OBE3722	Iberia_EN	Hungary_EN	-0.0009	0.003908	-0.218	116526
Mbuti	PECH5	Iberia_EN	LBKT_EN	-0.0089	0.008887	-0.998	36850
Mbuti	PECH8	Iberia_EN	LBKT_EN	-0.0179	0.005323	-3.355	97301
Mbuti	PEI10	Iberia_EN	LBKT_EN	0.0048	0.008514	0.561	36469
Mbuti	PEY53	Iberia_EN	LBKT_EN	-0.0109	0.005997	-1.822	87603
Mbuti	PEY163	Iberia_EN	LBKT_EN	-0.0014	0.006496	-0.209	75568
Mbuti	Pir4	Iberia_EN	LBKT_EN	-0.0153	0.008482	-1.805	46698
Mbuti	Pir6	Iberia_EN	LBKT_EN	-0.0069	0.008084	-0.854	47684
Mbuti	PIR3037AB	Iberia_EN	LBKT_EN	-0.0029	0.007899	-0.369	40215
Mbuti	PIR3116B	Iberia_EN	LBKT_EN	-0.0119	0.006518	-1.821	69167
Mbuti	PSS4170	Iberia_EN	LBKT_EN	-0.0005	0.004269	-0.123	204884
Mbuti	PSS4693	Iberia_EN	LBKT_EN	-0.0053	0.00378	-1.405	315184
Mbuti	PT2	Iberia_EN	LBKT_EN	-0.0012	0.005957	-0.196	78403
Mbuti	QUIN58	Iberia_EN	LBKT_EN	-0.0134	0.006247	-2.140	75048
Mbuti	QUIN234	Iberia_EN	LBKT_EN	-0.0052	0.006252	-0.828	85654
Mbuti	RIX2	Iberia_EN	LBKT_EN	-0.0079	0.005264	-1.494	115861
Mbuti	RIX4	Iberia_EN	LBKT_EN	-0.0066	0.005018	-1.316	121872
Mbuti	RIX15	Iberia_EN	LBKT_EN	-0.0075	0.005801	-1.291	98644
Mbuti	ROS45	Iberia_EN	LBKT_EN	-0.002	0.004666	-0.423	151226
Mbuti	ROS78	Iberia_EN	LBKT_EN	-0.0091	0.004296	-2.109	216428
Mbuti	ROS82	Iberia_EN	LBKT_EN	-0.0041	0.006802	-0.610	62050
Mbuti	ROS102	Iberia_EN	LBKT_EN	-0.008	0.00575	-1.385	95950
Mbuti	Sch72-15	Iberia_EN	LBKT_EN	0.0086	0.005156	1.677	133491

Outrgroup	AncientFrench	EarlyNeolithic1	EarlyNeolithic2	Dvalue	std.err	Zscore	#SNPs
Mbuti	Schw432	Iberia_EN	LBKT_EN	0.0149	0.005743	2.595	93024
Mbuti	WET370-1	Iberia_EN	LBKT_EN	-0.0068	0.006102	-1.119	75049
Mbuti	O18_I4303	Iberia_EN	LBKT_EN	-0.0068	0.003493	-1.960	405909
Mbuti	O18_I4304	Iberia_EN	LBKT_EN	-0.0066	0.003427	-1.922	443623
Mbuti	O18_I4305	Iberia_EN	LBKT_EN	-0.0041	0.00359	-1.148	427233
Mbuti	O18_I3874	Iberia_EN	LBKT_EN	0.0004	0.003379	0.133	357694
Mbuti	O18_I3875	Iberia_EN	LBKT_EN	-0.0037	0.00564	-0.653	145943
Mbuti	ATT26	Iberia_EN	LBKT_EN	-0.0174	0.006285	-2.761	76351
Mbuti	BERG02-2	Iberia_EN	LBKT_EN	-0.0051	0.004527	-1.118	183802
Mbuti	BERG157-2	Iberia_EN	LBKT_EN	-0.0099	0.004578	-2.162	174630
Mbuti	BERG157-7	Iberia_EN	LBKT_EN	-0.0106	0.004933	-2.153	142463
Mbuti	BES1248	Iberia_EN	LBKT_EN	-0.0095	0.005691	-1.677	91608
Mbuti	BFM265	Iberia_EN	LBKT_EN	0.0003	0.005123	0.058	119913
Mbuti	BIS130	Iberia_EN	LBKT_EN	0.0015	0.005648	0.265	86208
Mbuti	BIS385	Iberia_EN	LBKT_EN	-0.0092	0.005527	-1.664	115279
Mbuti	BLP10	Iberia_EN	LBKT_EN	-0.0104	0.00555	-1.882	118037
Mbuti	BUCH2	Iberia_EN	LBKT_EN	-0.0039	0.004323	-0.894	196941
Mbuti	CBV95	Iberia_EN	LBKT_EN	-0.0129	0.005724	-2.256	102034
Mbuti	COL11	Iberia_EN	LBKT_EN	-0.0171	0.006015	-2.848	81958
Mbuti	COL153A	Iberia_EN	LBKT_EN	-0.0091	0.006017	-1.513	90222
Mbuti	COL153i	Iberia_EN	LBKT_EN	-0.0019	0.006356	-0.300	66352
Mbuti	CRE20D	Iberia_EN	LBKT_EN	-0.0133	0.004809	-2.775	150933
Mbuti	ERS86	Iberia_EN	LBKT_EN	-0.0048	0.005159	-0.935	111958
Mbuti	ERS88	Iberia_EN	LBKT_EN	-0.0046	0.006286	-0.738	78213
Mbuti	ERS1164	Iberia_EN	LBKT_EN	0.0089	0.005344	1.664	114963
Mbuti	Es97-1	Iberia_EN	LBKT_EN	-0.0096	0.004725	-2.022	162967
Mbuti	EUG11	Iberia_EN	LBKT_EN	-0.0056	0.005908	-0.954	90597
Mbuti	Jeb8	Iberia_EN	LBKT_EN	-0.0042	0.005989	-0.698	82765
Mbuti	MDV248	Iberia_EN	LBKT_EN	0.0116	0.006201	1.864	81624
Mbuti	Mor6	Iberia_EN	LBKT_EN	-0.0183	0.005855	-3.119	95883
Mbuti	NIED	Iberia_EN	LBKT_EN	-0.002	0.005967	-0.333	82659
Mbuti	NOR2B6	Iberia_EN	LBKT_EN	-0.0069	0.005004	-1.375	119105
Mbuti	NOR3-6	Iberia_EN	LBKT_EN	-0.0082	0.005596	-1.473	99046
Mbuti	NOR3-15	Iberia_EN	LBKT_EN	-0.0068	0.00649	-1.049	75921
Mbuti	NOR4	Iberia_EN	LBKT_EN	-0.0052	0.006626	-0.789	69033
Mbuti	OBE3626-1	Iberia_EN	LBKT_EN	-0.0003	0.004905	-0.055	141434
Mbuti	OBE3722	Iberia_EN	LBKT_EN	-0.0021	0.005293	-0.404	109420