Archaeogenetics of Western Europe: the transition from the Mesolithic to the Neolithic

Alessandro Fichera

Main Supervisor: **Dr Maria Pala**

Co-Supervisors: **Prof Martin Richards, Dr Ceiridwen Edwards**

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Biological and Geographical Sciences
School of Applied Sciences
University of Huddersfield

University of HUDDERSFIELD
Inspiring global professionals

To my Grandfather

Copyright Statement

I. The author of this thesis (including any appendices and/ or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

II. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

III. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

Acknowledgements

I would like to thank my main supervisor, Dr Maria Pala, for her constant supervision and support during my years of PhD, all the things I learned from you, are and will be extremely useful for my career. I would like to thank as well, my two co-supervisors, Prof Martin Richards and Dr Ceiridwen Edwards, your vast knowledge on various subjects played an essential role on the realization of this work.

I thank the Leverhulme Trust Doctoral Scholarship for the funding of my PhD.

I would like to thank the collaborators who provided the samples analysed in this thesis: the late Rebecca Miller and Prof John Stewart for the ancient samples from Belgium, and Dr Mait Metspalu and his team (University of Tartu) for the modern mitogenomes.

I want to thank Prof David Reich and his team at the University of Harvard for collaborating in the sequencing of the ancient samples.

I thank Dr Pedro Soares for the help regarding haplogroup HV and his helpful support.

A special thank goes to the Human Genetics research team at the University of Pavia for providing modern samples analysed in this thesis, introducing me to the subject and believing in me for all these years, in particular: Prof Antonio Torroni, Dr Anna Olivieri, Dr Alessandro Achilli and Dr Francesca Gandini.

My immense gratitude goes to my biological family for all the sacrifices they did to bring me here and for the support during all these years. A sincere thanks goes to the new family I found here at the University of Huddersfield, in particular: Marina, Joao, George, Gonzalo, Bobby, Simao, Rita, Pierre, Bernardo, Rohan, Fabiola, Jennifer, Marisa, Katharina, Suhail, Dulce and many others. You helped in making my time here beautiful and unforgettable.

I would like to thank again Francesca Gandini, I could not ask for a better friend, advisor and colleague. You have been by my side for the last 8 years, in good and bad times.

I have so many good memories and will never forget you. Thank you zia!

Abstract

The transition from hunting to farming started in the Fertile Crescent of the Near East, about 12 thousand years ago (kya). During the following millennia, farming spread across Europe largely due to migrations of people from a source in western Anatolia. The aim of this thesis is to investigate and assess the relative contribution of local hunter-gatherers and dispersing farming groups in a region of Western Europe where the archaeological evidence suggests potential complexity. In order to do so, two parallel approaches were carried out: i) the study of human remains from three archaeological sites in Belgium; and ii) a broader phylogeographic analysis of modern mitochondrial DNA sequences belonging to haplogroup HV.

Here I report the first genome-wide analysis of one Mesolithic and 32 Middle to Late Neolithic Belgian individuals. The Mesolithic individual was largely similar to other Western European Mesolithic and Late Palaeolithic samples. However, within the Neolithic group I observed two genetic clusters. The first cluster appears to be the result of an admixture between local Mesolithic hunter-gatherers and Neolithic farmers of Anatolian descent. However, the Mesolithic component was much larger than seen to date in other west European Neolithic samples, with a possible sex bias towards local males carrying Y-chromosome haplogroup I and dispersing females. The second, less numerous genome-wide cluster revealed admixture from a Pontic-Caspian Steppe related population, further indicated by the presence of Y-chromosome R1b-M269.

The phylogeographic analysis of modern mitochondrial haplogroup HV confirmed an Upper Palaeolithic Near Eastern origin. The new findings suggest an early introduction of several HV lineages into the north coast of the Mediterranean from the Late Glacial onwards, which increased during the Neolithic. In particular, the Mediterranean area appears to have served as a reservoir of HV lineages and as a source of later migrations in both the Neolithic and the Bronze Age.

CONTENTS

List of Figures

List of Tables

List of Abbreviations

ABS: Abri Sandron
AD: *anno Domini*
aDNA: ancient deoxyribonucleic acid
AMH: Anatomically modern humans
BA: Bronze Age
bp: base pairs
BP: before present
Bst: *Bacillus stearothermophilus*
BWA: Burrows-Wheeler Aligner
CA: Copper Age
CHG: Caucasus hunter-gatherers
CRS: Cambridge reference sequence
CTR: control region
CV: cross-validation
ddH_2O: double distilled water
D-loop: displacement loop
DNA: deoxyribonucleic acid
dNTP: deoxyribonucleotide triphosphate
EB: elution buffer
EBA: Early Bronze Age
EBT: elution buffer + tween
EDTA: ethylenediaminetetraacetic acid
EHG: eastern hunter-gatherers
EN: Early Neolithic
FA: Founder Analysis
GATK: Genome Analysis Toolkit
GMF: Grotte du Mont Falise
HG: hunter-gatherers
hg19: Human genome build version 19
IBD: Identity by descent
IGV: Integrative Genomic Viewer
Indel: Insertion or deletion of one or more nucleotides
Kb: kilobases
Kya: thousand of years ago
LaH: La Hoguette
LBA: Late Bronze Age
LBK: Linearbandkeramik
LCT: lactase
LGM: Last Glacial Maximum
LLM: lower left molar
LM: lower molar
LN: Late Neolithic
LRJ: Lincombian-Ranisian-Jerzmanowician
LRM: lower right molar
MBA: Middle Bronze Age
ML: maximum likelihood
MN: Middle Neolithic
MSY: male-specific region of Y chromosome
MtDNA: mitochondrial deoxyribonucleic acid
NE: North-East
NGS: next-generation sequencing
Np: nucleotide position
NW: North-West
PAML: Phylogenetic Analysis by Maximum Likelihood

PB: Phosphate-Buffered
PBS: Phosphate-Buffered Saline
PCA: Principal Component Analysis
PCR: polymerase chain reaction
PE: Wash buffer
PEG: polyethylene glycol
rCRS: revised Cambridge reference sequence
ROH: runs of homozygosity
RNA: ribonucleic acid
Rpm: round per minute
RSRS: Reconstructed Sapiens Reference Sequence
SDS: sodium dodecyl sulphate
SE: South-East
SHG: Scandinavian hunter-gatherers
SNP: single-nucleotide polymorphism
TA: transitional complexes
TAW: Trou Al'Wesse
TBE: Tris-borate-EDTA
Tris-HCl: hydroxymethyl aminomethane hydrochloride
tRNA: transfer ribonucleic acid
UDG: uracil-D-glycosylase
USER: Uracil-Specific Excision Reagent
UV: ultraviolet
WHG: western hunter-gatherers

1.1 Brief history of Genomics

Genomics is a major and fast evolving branch of genetics which focuses on the sequencing and analysis of the genome of an organism. However, it is hard to unequivocally pinpoint a single starting point to genomics. Undoubtedly two events could be considered as the origin of modern genomics: in 1944 Avery and colleagues (Avery *et al*. 1944) identified the deoxyribonucleic acid (DNA) as the hereditary material contained inside the cell nucleus, and in 1953, Watson and Crick described its double helix structure (Watson and Crick 1953). The backbone of each strand of DNA, which forms the double helix, is represented by a chain of deoxyriboses linked to each other by phosphodiester links. On each sugar, one of the following bases is attached: adenine (A) or guanine (G) (purines), thymine (T) or cytosine (C) (pyrimidines). The two strands are kept together by the formation of hydrogen bonds between complementary bases: double bonds between T and A, and triple bonds between G and C. The duplication of the DNA is a process regulated by an enzyme called DNA polymerase. It was first discovered in 1956 by Kornberg (Kornberg *et al.* 1956). Inside the cells, there are repair mechanisms which correct possible "errors" introduced by the polymerase during replication. If these errors escape the repair mechanisms, the resulting genetic material is different from the original and this change is called mutation. These discoveries paved the way to everything we know today.

In 1977, Sanger completely sequenced the first genome, bacteriophage φX174, using the dideoxy-sequencing technology (Sanger and Coulson 1975; Sanger *et al*. 1977). Since then, the Sanger method has been the most used sequencing method. However, over the last decade, there has been a progressive shift towards new sequencing methods, generally referred to as next-generation sequencing (NGS). Compared to the "first generation" sequencing, the efficiency of NGS techniques is hugely improved while the costs are reduced.

In 1986 another groundbreaking "invention" drastically improved the world of molecular biology: the polymerase chain reaction (PCR) by Mullis and colleagues (Mullis *et al*. 1986). The PCR is a technique that consists of the repeated copying of a desired region of a DNA

molecule, generating millions or more copies of that segment. The copying is carried out by the purified thermostable DNA polymerase of *Thermus aquaticus*, and relies on thermal cycling: during which the mix of DNA molecules and DNA polymerase is cyclically melted and cooled to promote the replication of the target DNA. The selective replication is achieved thanks to the careful design of primers (short single stranded DNA fragments) containing sequences complementary to the target region. As PCR progresses, the amplified DNA is used as a template for additional cycles of replication, triggering a chain reaction in which the DNA template is exponentially amplified.

1.2 Population Genetics

Mutations are the building blocks of evolution. They can be then, passed to the offspring and represent the primary mechanism to introduce variation in a population. Studying this variation can reveal the history of a species, help tracing its origin and dispersals, and detail how it has adapted to different environments and, in the case of humans, to different lifestyles. Many of the differences between individuals are expressed in the form of isolated base substitutions, also called single-nucleotide polymorphisms (SNPs). Each individual carries an accumulated collection of SNPs, reflecting mutations that occurred in their ancestors during time. These substitutions can be classified in different categories. If a purine is replaced by another purine or a pyrimidine by another pyrimidine the resulting mutation is called transition. If the change is, instead, purine to pyrimidine or vice-versa, the mutation is called transversion. Errors in the sequence can also be represented by insertion or deletion of one or more nucleotides (indels). The site in the genome where these possible variants coexist is called *locus* and the different variants are defined as alleles.

“A population is an interacting and interbreeding group of individuals of the same species inhabiting the same geographical area” (Mayr 1982). A population that has passed through a bottleneck, or has been isolated for a certain amount of time, will show a reduced genetic variation. The genetic evolution of a population, and its gene pool, can be determined by different factors: i) generation time, defined as the time separating two generations of the same population, to a faster generation time corresponds a faster evolutionary rate; ii) selection: certain alleles can confer advantages of any kind to an individual or population in

specific environments, those alleles will be subject of *positive selection*, this means that they will be preferentially represented in future generations, and thus increase their frequency in the population over generations; vice-versa, alleles which confer a disadvantage will be *negatively selected* in the population and possibly disappear over a certain number of generations; iii) genetic drift, a stochastic mechanism which strongly affects allele frequencies, it can counteract the effects of selection.

1.3 Ancient DNA studies

The term ancient DNA (aDNA) defines all the genetic material extracted from degraded biological samples, like bones, teeth, hair, seeds or various tissues, generally older than 100 years (Cabana *et al.* 2013). Differently from the use of modern genetic data to infer hypotheses on the past, aDNA gives the advantage of providing a direct view of the genetic pattern at that given point in time. This allows researchers to formulate new hypotheses regarding genetic affinities between ancient individuals or populations and to test and verify those that have been formulated only using modern populations.

Due to its unique characteristics, the mitochondrial DNA played a crucial role in the development of aDNA studies. First, compared to the nuclear genome, its limited dimension and complexity made it an easier target for sequencing and screening at a reasonable cost. It is believed that its circular structure makes it less accessible to exonuclease enzymes and so its decay rate over time in *post-mortem* specimen is reduced (Schwarz *et al.* 2009; Allentoft *et al.* 2012). Ancient DNA was first used in 1984, sequencing fragments of mitochondrial DNA extracted from the remains of the Quagga, an extinct member of the *Equus* genus (Higuchi *et al.* 1984). Its use was rapidly extended on many different species, including humans in 1985 (Pääbo *et al.* 1985), although this early human work suffered from contamination. The first studies on complete mitochondrial genomes arrived a few decades later, with the publication of ten Siberian mammoth genomes by Gilbert and colleagues (Gilbert *et al.* 2007) and the first ancient human from Europe by Ermini *et al.* (Ermini *et al.* 2008). In the last decade, with the innovative introduction of NGS techniques, understanding of DNA structure, damage pattern, contamination and source material for the extraction of DNA, it has become possible to sequence not only the entire mtDNA but even the nuclear genome. Rasmussen and colleagues in 2010 sequenced the first complete

ancient human genome belonging to a 4000 years old paleo-Eskimo from Greenland, extracted from a tuft of hair (Rasmussen *et al.* 2010).

Typically, aDNA is characterised by a variable degree of degradation which starts right after the death of the organism by the action of endogenous nucleases, not regulated anymore by the cell repair mechanism, or by external organisms (like bacteria or fungi) (Hofreiter *et al.* 2001). Other examples of *post-mortem* damages are fragmentation and deamination (Pääbo 1989). With the innovative use of NGS it was possible to better understand the causes of this fragmentation. Experiments using known adapters, ligating at both ends of each DNA fragment, helped the identification, during the alignment process, of the bases right next to them. The study of Briggs *et al.* revealed the presence of an over-representation of purines next to the 5' ends, this is caused by a chemical reaction called depurination, where the bond between a sugar and a purine (adenine A or guanine G) is broken (Briggs *et al*. 2007). The presence of a newly formed apurinic site causes the activation of a class of enzymes which cleaves the bond between the sugar and the phosphate resulting in single-strand breaks (Pääbo and Wilson 1991) (Figure 1A). Deamination, instead, is a process where the amine group of a nucleotide is cleaved by hydrolysis causing an apparent base transition (Gilbert *et al*. 2007; Brotherton *et al*. 2007). Studies showed that the rate of deamination increases towards the ends of the sequence and typically involves citosine, thus producing uracil (Figure 1B). For this reason, a treatment with UDG (uracil-D-glycosylase) is usually recommended since it excises deaminated cytosines and therefore reduces the amount of damage-induced nucleotide sustitutions (Pääbo 1989; Hofreiter *et al*. 2001). In aDNA studies, it is common practice to not consider nucleotide transitions during the SNPs calling due to the high risk of false substitutions caused by DNA damage (Orlando *et al.* 2015; Kivisild *et al.* 2017). However, this approach leads to a severe loss of data.

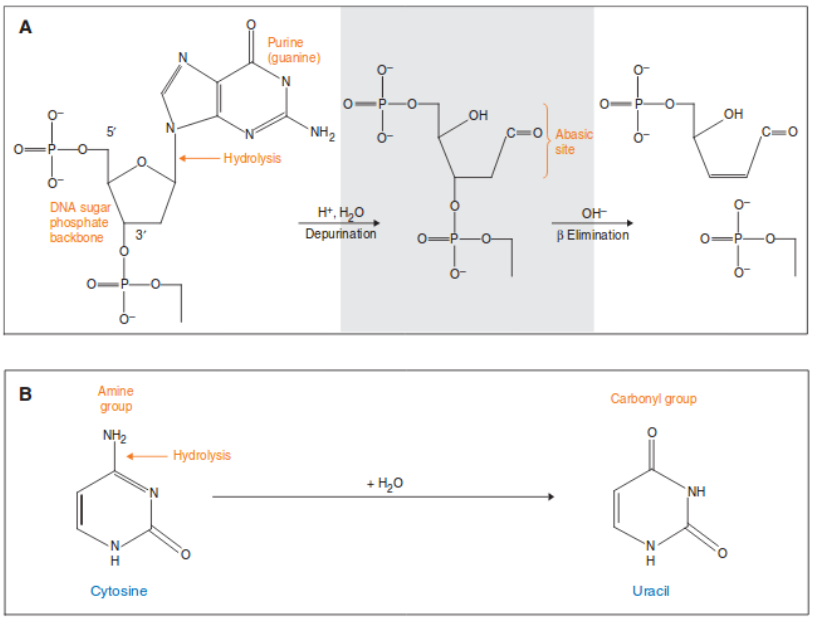


Figure 1. (A) Fragmentation of DNA by depurination. Hydrolysis of the bond between a sugar and a purine (adenine or guanine) resulting in an abasic site. The DNA strand is cleaved through β elimination. (B) Deamination process of cytosine into uracil (Dabney et al. 2013).

The rate of decomposition of a body is deeply influenced by the environment where the organism dies (Lindahl 1993). Cold, dry and low-radiation conditions favour DNA preservation by slowing the processes described above, so specimen retrived from high latitudes areas, especially those found in permafrost or buried deep inside caves and protected from the action of external agents, like water, hold the highest percentage of endogenous DNA. The term "endogenous DNA" defines all the genetic material belonging specifically to that organism, in ancient biological remains the percentage of endogenous DNA can vary significantly. The DNA extracted from ancient remains is a mixture of endogenous and exogenous DNA, with the latter mainly deriving from plants, fungi, bacteria and other organisms colonising the remains *post-mortem*. The exogenous DNA, source of contamination, can be distinguished in two categories: I) modern DNA contamination caused by the lack of precautions in handling the samples, relatively easy to identify and isolate due to the absence of the aDNA features described above; II) aDNA contamination, very hard to identify especially when the source of contamination belongs to the same species of the target. MtDNA is often used to determine the degree of contamination by identifying how many mtDNA lineages are present after a preliminary sequencing screening of the sample.

Typically the best target to retrieve the highest amount of endogenous DNA in an ancient human remain is the petrous bone (Figure 2) (a part of the temporal bone) due to its elevated density and its position protected inside the skull by external degrading agents (Pinhasi *et al.* 2015; Hansen *et al.* 2017). However, this is not true for mitochondrial DNA, as the petrous bone has a reduced cellular activity and so the mitochondria are not extremely abundant (they are involved in the production of energy). As reported in a study in 2015, teeth are the best target for an overall DNA retrieval (Higgins *et al.* 2015).

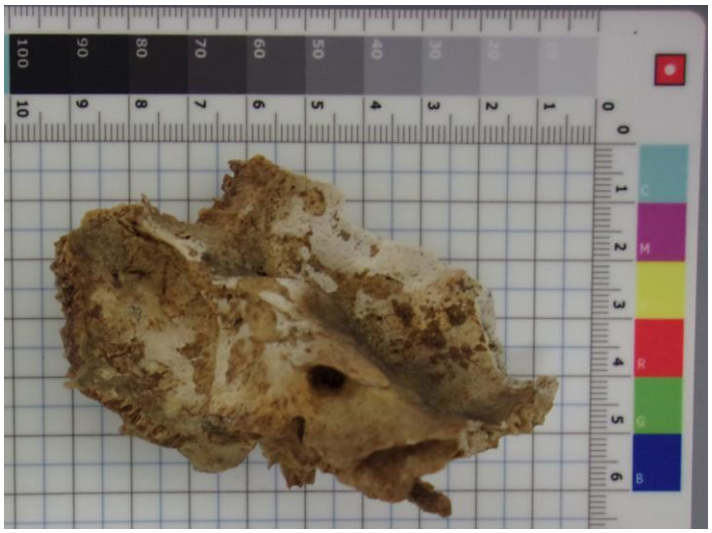


Figure 2. Human petrous bone.

1.4 Molecular Phylogenetics

Molecular phylogenetics aims to investigate the evolutionary history of a species by identifying relationships between DNA sequences or molecules. The tenet of phylogenetics is genetic variability, without it no explicative information can be extrapolated. One of the best procedures to illustrate evolutionary relationships is through a phylogenetic tree. Resembling the shape of a tree with a body and various branches, phylogenetic trees can be used to explain variability inside a specific population or among species, more or less related.

Mitochondrial DNA has been for many years the only tool to study the genetic history of species and the preferred candidate for phylogenetic studies; because of its main characteristics: i) high evolutionary rate, even in a rather short timeframe it can accumulate mutations and differentiate a numerous haplotypes, defined as a combination of alleles

present on different loci that are transmitted as a single unit; ii) small size, making it an easy target to analyse (although this has become less relevant with the advent of NGS technologies) and iii) the lack of recombination, that allows mitochondrial DNA sequence variation to arise uniquely from the sequential accumulation of new mutations along the maternal line of descent. During the course of time, this process has given rise to monophyletic units or clades. Thus a clade is a group of haplotypes that share a specific set of variants inherited from a common ancestor; a named clade is also called haplogroup. A paragroup, instead, refers to all the lineages within a haplogroup that are not defined by additional mutations. Paragroups are generally marked by a "*" following the haplogroup's name (e.g. HV*).

With regards to our species, the process of differentiation occurred mainly when modern humans were colonizing different regions and continents, so haplogroups and sub-haplogroups tend to be restricted to specific geographic areas and populations. Using a molecular clock, based on the time when the split between the genera *Homo* and *Pan* occurred, and the mutation rate of the DNA molecule, it is possible to estimate the age of a node or the moment in time when two haplogroups diverged, the so called coalescent age (Soares *et al.* 2010). Therefore, when the interpretation of genetic data allows the analysis of the diversity and geographical distribution of the lineages, precise information about the demographic history can be drawn, especially about the phenomena of expansion, migration and dispersal of lineages, and by proxy of populations carrying those lineages. This approach is commonly referred to as "phylogeographic approach" (Avise 2000; Templeton *et al*. 1995). Phylogeography is the phylogenetic analysis of geographically distinct genetic data to prove hypotheses regarding possible relationships among populations, species distributions, and the mechanisms of speciation. A phylogeographic study should therefore be interdisciplinary: data from molecular biology, population genetics and phylogenetic analyses have to be combined with data coming from climatology, ecology, geography, demography, history, archaeology, anthropology, and linguistics in the case of humans. In the last few years, phylogeographic studies concerning mammalian species have proliferated, and the main reason for this rapid growth is the increasing level of molecular and phylogenetic resolution made possible by the development of NGS methods mentioned above.

1.5 Mitochondria

The mitochondria are double-membraned intracellular organelles with an inner membrane enclosing the matrix and an outer membrane delimiting the perimembrane space (Figure 3). Every eukaryotic cell contains from one to several hundred of mitochondria, each with a variable number of mitochondrial DNA (mtDNA) molecules.

Mitochondria are involved in several cellular functions. They play a part in intracellular signalling and apoptosis and in the metabolism of aminoacids, lipids, cholesterol, steroids and nucleotides. The most important function is related to cellular energy metabolism.

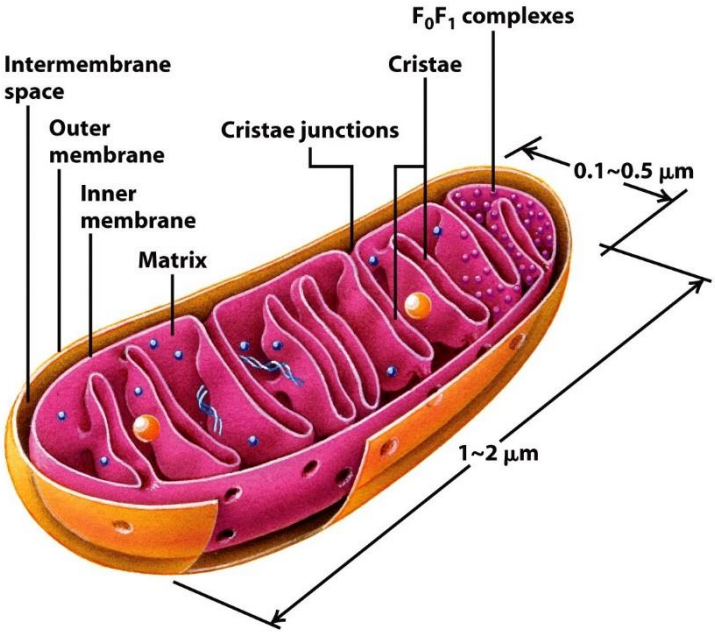


Figure 3. Structure of a mitochondrion. (Molecular Cell Biology, Sixth Edition, 2008 W.H. Freeman and Company)

1.5.1 Mitochondrial genome organization

The first human mtDNA complete sequence was published by Anderson *et al* in 1981 thus becoming the so-called Cambridge reference sequence (CRS) (Anderson *et al.* 1981). In 1999 Andrews and colleagues, re-sequenced the original mtDNA used by Anderson's group revealing a number of errors, leading to publish a new revised reference sequence (rCRS) (Andrews *et al*. 1999). In 2012, a study by Behar and colleagues proposed the use of an alternative reference for the human mitochondrial genome, the Reconstructed Sapiens Reference Sequence (RSRS). The rationale behind this choice was to solve practical problems during the classification of mitochondrial lineages using rCRS. Since rCRS belongs

to a European lineage, H2a2a1, during the haplogroup classification, and haplotype analyses, mutations are called, to use a tree analogy, versus a leaf within a tree, instead of its root. Because of this topology, all the mutations present between rCRS and the analysed sequence must be taken into account. RSRS, instead, is placed at the putative root of the human mitochondrial phylogeny, thus providing a clear identification of ancestral and derived states of the alleles analysed in the sequences (Behar *et al*. 2012). Admittedly, from an evolutionary perspective, the use of RSRS should be preferable to rCRS, however, "reference sequence" and "ancestral sequence" are two distinct concepts with very different functions. The replacement of rCRS with RSRS would imply that all haplotypes called versus rCRS, since its establishment as reference sequence, should have to be re-called, and the related discussions amended. This replacement would open the gates to confusion and misunderstanding that would affect not only phylogenetic studies, but extend to other disciplines such as medical genetics and forensics (Bandelt *et al.* 2013). Therefore, despite the introduction of this new reference sequence, rCRS remains the most used reference sequence for studies involving human mitochondrial DNA.

The human mitochondrial genome is 16,569 bp (base pairs) long, it is compact, lacks introns and is almost entirely coding, with the exceptions of some non-coding regions interspersed in the molecule (Figure 4). Perhaps the most notable of such sequences is the control region (CTR, also called displacement loop or D-loop), which is involved in the regulation of transcription and replication of the molecule. In human mtDNA the control region is 1,121 bp long and it starts from np (nucleotide position) 16,024 and it ends at np 576. This region, in comparison with the rest of the genome, is highly variable. Until few years ago, mtDNA population analyses were based almost exclusively on this region of the genome, useful for screening entire populations. Now, with advanced technologies, these studies include full mtDNA genomic sequences (and nuclear genomic data) to look into the complete variability of a given population.

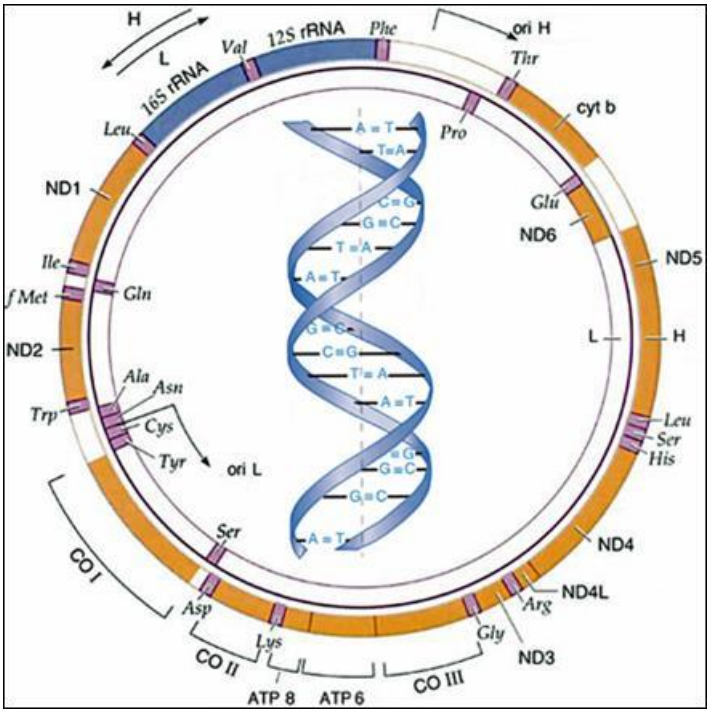


Figure 4. Map of the human mitochondrial genome. (Ross OA.)

The mtDNA encodes for 37 genes: 13 for subunits of the enzymes involved in the process of oxidative phosphorylation, two for ribosomal RNAs (12S and 16S rRNA) and 22 for transfer RNAs (tRNAs), which are needed for intra-mitochondrial protein synthesis (Table 1) (Boore 1999).

Table 1. Human mtDNA genes and products.

Gene	Product
COI, COII, COIII	Cytochrome oxidase subunit I, II and III
Cytb	Cytochrome b
ND1-6, 4L	NADH dehydrogenase subunit 1 to 6 and 4L
ATP6, ATP8	ATP synthase subunit 6 and 8
lrRNA	Large ribosomal RNA subunit
srRNA	Small ribosomal RNA subunit
tRNAs	18 amino acids-specific transfer RNAs
L(CUN) and L(UUR)	Two leucine tRNAs
S(AGN) and S(UCN)	Two serine tRNAs

1.5.2 Replication mechanism

The replication of mitochondrial DNA is a process that occurs independently from the cell cycle. This genome is continuously replicated even in non-dividing tissues (Bogenhagen and Clayton 1977; Birky 2001).

The mechanism that regulates the mitochondrial replication is still under study and debated. Currently there are two models: the "strand-asymmetric model" proposed by Clayton in 1982, postulates that mtDNA molecules replicate unidirectionally from two distinct origins. According to this model, the replication starts at the origin of the heavy strand O_H with the synthesis of a primary transcript that continues until the origin of the light strand replication O_L is exposed. At this point, the light strand is synthesized in the opposite direction (Clayton 1982). The alternative model, proposed more recently in 2003, supports a "rolling circle" replication mechanism, affirming that the replication of mtDNA begins in several origins in a region between the D-loop and the ND4 gene. The replication then proceeds in both directions completing the replication cycle. The light strand completes the cycle with the ligation of Okazaki fragments (Bowmaker *et al*. 2003; Lightowlers 2012).

1.5.3 Mitochondrial DNA inheritance

Mammalian mitochondrial DNA is inherited along the maternal line and therefore it does not undergo recombination (Giles *et al.* 1980; Macaulay *et al.* 1999; Sutovsky *et al.* 2000). It was originally thought that paternal mitochondria do not enter the oocyte, since they are localized in the spermatozoan tail and paternal mitochondria appear to be destroyed by ubiquitination in a mechanism still partially unclear (Sutovsky *et al.* 2000). Under natural conditions maternal inheritance is strongly controlled, however rare cases of paternal transimission in humans have been recorded and studied (Awadalla *et al.* 1999; Schwartz *et al.* 2002; Luo *et al.* 2018).

1.5.4 Heteroplasmy

As stated before, mammalian cells contain multiple copies of mtDNA. During the replication process, these molecules can be target of mutations leading to the scenario that not all the copies are identical. The presence of multiple mtDNA types within an individual is called heteroplasmy. Originally heteroplasmy was believed to be associated to diseases, aging and cancer (Monnat *et al.* 1985), however, recent studies have revealed that low level mtDNA heteroplasmy appears even in healthy individuals (Irwin *et al*. 2009; He *et al.* 2010; Payne *et al.* 2013). Heteroplasmic sites are typically individual-specific and eliminated by genetic drift, however if they appear in the maternal germline they could go through a genetic bottleneck and be passed to the offspring (Hauswirth *et al.* 1982). Even though the process is still not

completely known, the reduction in the number of mitochondrial genomes occurs during the embryogenesis of the female germ line, removing possible deleterious mutations from the population (Cao *et al.* 2007; Cree *et al.* 2008; Samuels *et al.* 2010).

1.5.5 MtDNA phylogeny

The nomenclature system to define all the different human mitochondrial haplogroups was established in 1993 when Torroni and colleagues identified the four main Native American haplogroups A, B, C and D (Torroni *et al.* 1993). The nomenclature follows an alphanumeric logic. Considering a hypothetical haplogroup A, its two sub-branches will be named A1 and A2. If the clade A1 contains three sub-clades nested within, they will be called A1a, A1b and A1c, and so on (Richards *et al.* 1998).

The oldest branches, all sub-haplogroups of L, are geographically limited to the African continent (Figure 5). The human variation outside Africa falls under the mitochondrial macro-haplogroup L3 dated around 70 kya, then diverging into M and N both dated between 50 and 70 kya (Soares *et al.* 2009; Behar *et al*. 2012). M is limited to the South and East Asia and Oceania today. Haplogroup N, instead, is widely spread going from Asia, to the American continents and Europe. Haplogroup U, nested within R, the main clade of N, is the mtDNA haplogroup with the oldest coalescence age in Europe. However, U is not indigenous of Europe and it is currently found in modern southern Asian populations (Achilli *et al.* 2005; Soares *et al.* 2010). The spread or the extinction of some haplogroups in certain areas of the world, or in specific populations, is often related to historical events, like haplogroup L3 with the "Out of Africa" model, or the drastic increase in frequency of haplogroup HV in Europe after the Neolithic revolution (see section 1.8.4 for further details). Currently, haplogroup H, a sub-clade of HV, is the most common haplogroup in modern Europeans (Achilli *et al*. 2004).

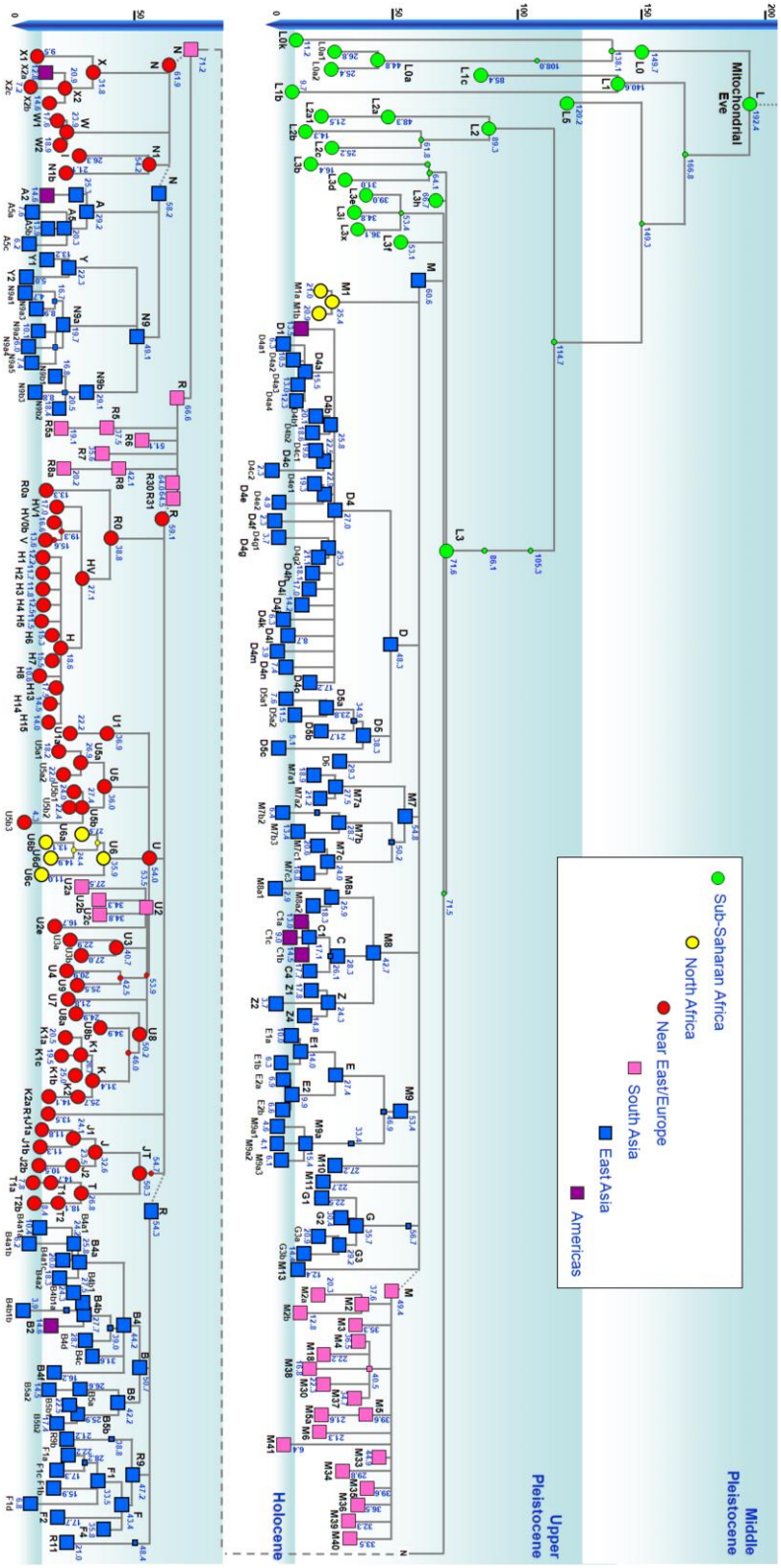


Figure 5. Complete human mitochondrial DNA phylogeny with present-day haplogroup distribution across the world and coalescence ages (in kya) (Soares et al. 2009).

1.6 Y chromosome

The Y chromosome is the male specific uniparental marker as it is passed from father to son. It consists of two main regions: the MSY (male-specific region), which covers 95% of the entire length of the chromosome, and two pseudo-autosomal regions at both ends of the chromosome (Skaletsky *et al*. 2003). The recombination between the X and Y chromosomes is limited to the pseudo-autosomal regions, while it is suppressed in the MSY, where all the sex-determining genes are located (Graves 2006).

Consistent with the mitochondrial DNA phylogeny, the root of the Y chromosome tree shows an extreme haplogroup variability mainly found in African populations, represented by haplogroups A, B and DE (Cruciani *et al*. 2011; Mendez *et al*. 2013; Poznik *et al*. 2013; Wei *et al*. 2013; Haber *et al.* 2019). The genetic variability of human populations in the rest of the continents is limited to three main haplogroups, C, D and F (Hallast *et al*. 2014; Scozzari *et al*. 2014; Karmin *et al*. 2015) (Figure 6). Specific sub-clades of C have been found in hunter-gatherers and Neolithic farmers from Europe, possibly suggesting that this haplogroup was more wide spread in Eurasia in pre-historic times (Scozzari *et al.* 2012; Olalde *et al.* 2014; Mathieson *et al.* 2015). The biggest branch within F is represented by haplogroup K. The distribution of haplogroup K is extremely widespread covering most of the world from West Eurasia to the American continents. Similarly to the pattern shown in the mitochondrial phylogeny, the distribution and frequency of certain Y chromosome haplogroups is deeply related to specific events in history. A remarkable example is the strong replacement of all the paternal lineages in Europe with the Yamnaya population migration from the Steppe around 4.5 kya (see section 1.8.5 for further details). Prior to that, European hunter-gatherers belonged mostly to haplogroup I. Haplogroup I derives from the node IJ, brother of haplogroup K and its distribution is limited to the European continent (Rootsi *et al*. 2004). During the Neolithic, European farmers were a mix of G2a, H2 and partly I2a (inherited by the local hunter-gatherers). With the arrival of the Yamnaya there is a clear shift in the haplogroup frequencies, with the rise of a particular R1b clade, called M-269. R1b derives from R, a branch within K. However, different R1b sublineages were already present in Europe before the arrival of the Yamnaya, during the Late Neolithic/Early Bronze Age, such as R1b-V88, found in two individuals from Italy

(Epigravettian) and Iberia (Early Neolithic) (Haak *et al.* 2015; Fu *et al.* 2016). Today R1b-M269 is the most common European Y chromosome haplogroup.

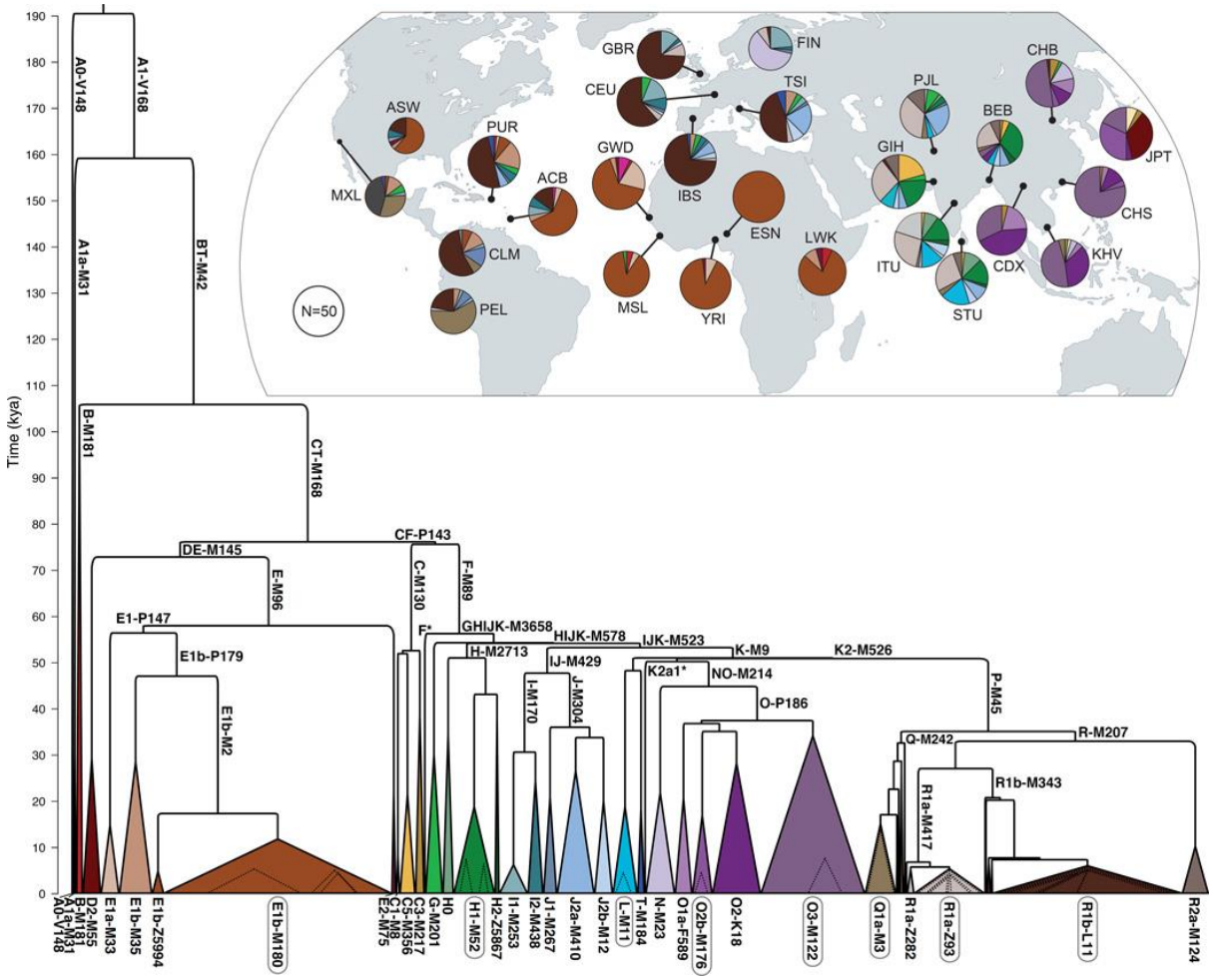


Figure 6. Complete human Y-chromosome phylogeny with estimated ages. The dimensions of the triangles are proportional to the sample size. The map in the upper part shows the geographic origin and Y chromosome haplogroup composition of the 26 populations analysed (Poznik et al. 2016).

1.7 Human nuclear genome

The human genome has been completely sequenced, for the first time in 2004 by the International Human Genome Sequencing Consortium. From there, it was possible to better understand its structure and organization, we now know that it contains over 6 billion base pairs (bp), distributed among the 46 chromosomes packed inside the cell nucleus. The genome includes both (a surprising larger fraction of) non-coding DNA and protein-coding DNA genes. The sizes of the human chromosomes are very heterogeneous, spanning from almost 249 million bp in the biggest, chromosome 1, to 46 million bp for chromosome 21

(Ludwig 2002). The information contained inside the chromosome can be analysed by genotyping polymorphic sites spread over the whole genome.

The first genotypic study, in 1987, was based on a very limited number of markers and individuals (Bowcock *et al.* 1991). In 2002 a project called HapMap aimed to develop a haplotype map of the human genome. The catalog included SNPs involved in the development of diseases and various phenotypes. Using eleven different populations across the world, it was possible to compare the genetic variation and allele frequency in human populations (The International HapMap Consortium 2003). In the last decade additional international collaborations, such as 1000 Genomes Project and the Simons Genome Diversity Project, helped to increase the knowledge regarding human genetic variation, allowing in-depth studies between populations, within the same population and at individual level. On average, more than 99% of the genome is identical across all humans. Moreover, the private polymorphisms which characterise different populations are limited to a very small portion of the variation and can be distinguished in common variants, which evolved early during the history of humankind and so are shared among different populations, and more rare variants, usually restricted to isolated populations (Figure 7).

Several approaches, such as ADMIXTURE analysis and Principal Component Analysis (PCA), have been developed to analyse and estimate the ancestry of individuals or populations. ADMIXTURE is a widely used software that applies a maximum likelihood approach to estimate different ancestral components of individuals, from genotyped datasets (Zhou *et al.* 2009). It is based on the same statistical model used by the previous software STRUCTURE, however ADMIXTURE has a faster calculating algorithm and allows for larger dataset to be analysed. PCA, instead, first used by Menozzi and colleagues in 1978 to study European population structure (Menozzi *et al.* 1978), is a tool to analyse multiple genetic variants by assigning them to fewer major components, which can then be graphically displayed in a multi-dimensional plot, with the distance between the points reflecting the relatedness among individuals (Novembre *et al.* 2008).

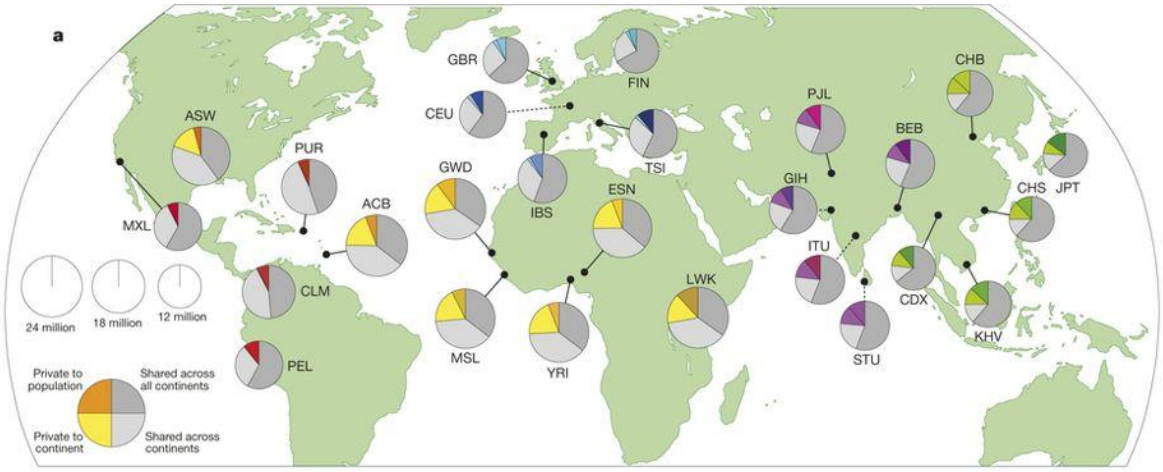


Figure 7. Genetic variants across different human populations. The shades of the colours in the charts represents how common the variants are, from very dark colour being private to a population, to dark grey being shared across all continents (The 1000 Genomes Project Consortium 2015).

1.8 *Homo sapiens* origin and dispersal

1.8.1 Origin of modern humans and Out of Africa

Homo sapiens originated in Africa, possibly as early as 315 kya, date associated to the oldest anatomically modern human found until now, located in Morocco (Hublin *et al.* 2017; Richter *et al.* 2017). Anatomically modern humans (AMH) generally differ from other members of the genus *Homo*, such as Neanderthals, for their lighther body build, structure of the skull, weaker jaw structure and smaller teeth (Hublin *et al.* 2017; Stringer *et al*. 2017). Several theories exist regarding the routes taken by human populations migrating out of the African continent and the number of migrating events which took place. According to some of them, based on palaeoclimatic studies, the Out of Africa was the result of a series of multiple migrations. This scenario suggests that between 125 and 40 kya, several windows of favourable climatic conditions occurred, transforming the land between northeast Africa and the Arabian Peninsula into a green corridor, and allowing groups of humans to cross it (Larrasoaña *et al.* 2013; Parton *et al.* 2015; Timmermann and Friedrich 2016).

It is possible to estimate a window of time during which the firsts modern humans left the African continent using the coalescence ages of the mitochondrial haplogroup L3, clade under which all the modern non-African human populations fall, and its main sub-clades. The estimated age of L3, around 70 kya (79 kya if the 95% upper bound is considered), represents the upper boundary, while the ages of haplogroups M and N, between 50 and 65 kya, represent the lower boundary (Soares *et al*. 2012; Mellars *et al.* 2013). According to this

scenario, the successful Out of Africa migration that allowed the colonization of the world took place in the range between 70 and 50 kya. However, a recent study by Pagani and colleagues pointed at a small percentage of nuclear genome, around 2%, in modern-day Papuans which could be the marker of an earlier and largely extinct out of Africa migration (Pagani *et al.* 2016).

1.8.2 Peopling of Europe

It is not certain when AMH first arrived in Europe but thanks to archaeological discoveries, it was possible to point their arrival around 45 kya. This date matches the discovery, made in 2003 by Trinkaus and colleagues, of the oldest European modern human remain in Romania: Oase Man. Dated to around 37-42 kya (Fu *et al.* 2015), he was named after the cave where he was found (Trinkaus *et al.* 2003). The oldest human remains from the entire Eurasia, instead, is represented by Ust'-Ishim, from Siberia dated around 45 kya (Fu *et al*. 2014). Both individuals belong to the same Y chromosome haplogroup K. K is a sub-haplogroup of F, "father" of most lineages spread all over Eurasia, Oceania and Americas (Bergstrom *et al*. 2016). Their very early dating and affiliation with the root of the haplogroup, whose coalescence age is around 50 kya (Poznik *et al.* 2016), suggest that Oase Man and Ust'-Ishim belonged to the early waves of modern humans spreading over the Eurasian continent. The same scenario is reflected in the maternal line, with both individuals belonging to the macro-haplogroup N. Due to the highly fragmented DNA an accurate classification within N was not possible for Oase Man, however, current data suggests that it probably diverged earlier in the phylogeny and should be classified as pre-N; while Ust'-Ishim has been further classified as belonging to mitochondrial haplogroup R. Both haplogroups represent early offshoots of the human mitochondrial tree in Eurasia (Mellars *et al.* 2013).

Although the Aurignacian complex is generally considered to be the first well documented autochtonous cultural complex associated to AMH from Europe, recent discoveries point at the so-called "transitional complexes" (TA) as the earliest evidence of AMH in Europe (Hublin 2015). TA is chronologically placed between the Initial Upper Palaeolithic and the rise of the Aurignacian and defines a group of lithic industries that includes Szeletian, Lincombian-Ranisian-Jerzmanowician (LRJ), Châtelperronian and Uluzzian (Hublin *et al.* 2015). The Aurignacian culture, dated later than 40 kya, was named after a small town in

the south of France, Aurignac, where the firsts archaeological remains were found (Mellars 2006; Hoffecker *et al.* 2009). Some features of this culture were the use of bone tools, blades crafted using flints and the production of some of the earliest human-shaped figurines and cave paintings, such as the ones at Chauvet cave in southern France (Soffer *et al.* 1993; Chauvet *et al.* 1996).

Around 33 kya, another culture arose in Europe, the Gravettian. Similarly to the Aurignacian, it was named after the archaeological site in southern France where it was first discovered. The Gravettian has been identified all over Europe from Ukraine to Portugal (Marks *et al.* 1994). The Gravettians were skilled hunter-gatherers specialized in large game hunting, such as mammoths, constantly moving to gather food (Kipfer 2000). They produced a variety of ornaments and figurines, and implemented one of the first burial rite systems based on offerings for the deceased (Trinkaus 2006). Between 17 and 11 kya, the Magdalenian culture was widely spread across Western Europe and represented the climax of the Upper Palaeolithic cultural development. Archaeological findings have retrieved abundant examples of spear points and flint tools, and impressive cave art, some of the finest European Upper Palaeolithic cave paintings are associated with this culture. However, it seems that Magdalenian people were not only outstanding artists but also skilled hunters, in particular of reindeers and bisons notably abundant towards the end of the Palaeolithic (Schwendler 2012).

Close to the Last Glacial Maximum (LGM), a period during which ice sheets covered most of North America, northern Europe and Asia, humans were forced to retreat towards lower latitudes areas, such as the Mediterranean region, these zones are called glacial refugia. The harsh climatic conditions had a deep impact on the environment, changes in sea level, humidity and temperature and strongly affected floral and faunal populations (Birks and Ammann 2000; Brooks and Birks 2001; Clark *et al.* 2009). Among the known refugia in western Eurasia are the Franco-Cantabrian one, between the Iberian Peninsula and France, the Alps, delimiting the Italian peninsula, the Balkans, Ukraine and the Near Eastern one (Ashcroft 2010; Olson *et al.* 2012). The populations who took refuge in the favourable climatic areas went through a genetic bottleneck, followed by expansion when they later repopulated the entire continent (Wells *et al*. 2001). This caused a decrease in the genetic variation and in the number of individuals, confirmed by the high values of *runs of*

homozygosity (ROH) in hunter-gatherers individuals across Europe, compared to Neolithic populations (Jones *et al.* 2015; Cassidy *et al.* 2016).

During the LGM in Europe, different cultures start to arise in the area of the known refugia, such as the Solutrean culture between Iberia and France and the Epigravettian in the Italian peninsula, Balkans and Ukraine. The Solutrean culture was characterised by the production of a particular kind of bifaced flint point, used for hunting and as a tool, and remarkable cave paintings (Hugh 1911). In the eastern part of Europe, instead, the Epigravettian replaced the previous Gravettian culture. Covering a wide area of Eurasia, from Italy to western Russia, this culture showed a degree of variability in the artefacts and tools that reflected the distribution across diverse habitat, such as mollusk remains used probably as decorations found in sites in Italy and the Balkans or ivory artifacts from mammoths in the tundra in Ukraine. Overall, it was characterised by an intensive production of bladelets or backed points (Kozlowski 1999; Olenkovskiy 2010). The Epigravettian lasted until 11.5 kya, a a period of time marked by a warmer temperature, an increase of forested environments, changes in the communities and the start of the Mesolithic period.

1.8.3 Mesolithic in Europe

The Mesolithic (literally, Middle Stone Age) was a period whose length varies geographically but started with the rapid warming after the end of the Younger Dryas, at 11.5 kya. During this time, Europe was inhabited exclusively by hunter-gatherer populations. They were specialised in hunting game, such as aurochs, deer and wild boars, fishing and gathering fruits and nuts (Zilhao 2000). They were socially organised in small and highly mobile groups, with seasonal occupation of sites along rivers and shores (Stanko 2007), which helped on creating a high degree of cultural homogeneity (Whittle and Cummings 2007). Among the identified Mesolithic sites, probably the most famous is represented by Lepenski Vir, a site located in the area of modern Serbia associated with the Iron Gates cultures. Described as the “oldest city of Europe”, Lepenski Vir included a large settlement and several satellites villages around it, probably inhabited almost permanently (Srejovic 1972).

The Y chromosome haplogroup I, seems to represent the signature of European Mesolithic populations (Lazaridis *et al*. 2014; Fu *et al*. 2016; Lipson *et al.* 2017). In some cases, the presence of Y chromosome haplogroup I in Neolithic/Bronze Age individuals suggests a

measure of genetic continuity from the Mesolithic, although the source of assimilation needs to be identified in each case (Haak *et al*. 2015; Mathieson *et al*. 2015; Olalde *et al.* 2019). It is possible to differentiate European Mesolithic populations on the basis of genome-wide data into three genetically, and geographically, distinct clusters, known as: western hunter-gatherers (WHG), eastern hunter-gatherers (EHG) and Scandinavian hunter-gatherers (SHG). As shown in the Principal Component Analysis (PCA) in Figure 8, the HGs cluster outside modern European genetic diversity, with a cline from West to East (Lazaridis *et al.* 2016; Mathieson *et al.* 2018). The Caucasus hunter-gatherers (CHG) form a separate cluster from the rest of European HG, being genetically close to modern Caucasus and Iranian populations (Figure 8). In support of the hypothesis of mtDNA haplogroup U being carried by the first modern human settlers of Europe and being the oldest mitochondrial lineage (Richards *et al.* 1998; Richards *et al.* 2000; Soares *et al.* 2010), almost all pre-Neolithic individuals analysed so far belong to haplogroups U (more precisely U2, U4, U5 and U8) and K (haplogroup K is a sub-clade of U8) (Bramanti *et al.* 2009; Krause *et al.* 2010; Hervella *et al.* 2012; Sanchez-Quinto *et al.* 2012; Bollongino *et al.* 2013; Der Sarkissian *et al.* 2013; Fu *et al.* 2013; Lazaridis *et al.* 2014; Olalde *et al.* 2014; Skoglund *et al.* 2014; de la Rua *et al.* 2015; Haak *et al.* 2015; Fu *et al.* 2016; Mathieson *et al.* 2018; Feldman *et al.* 2019).

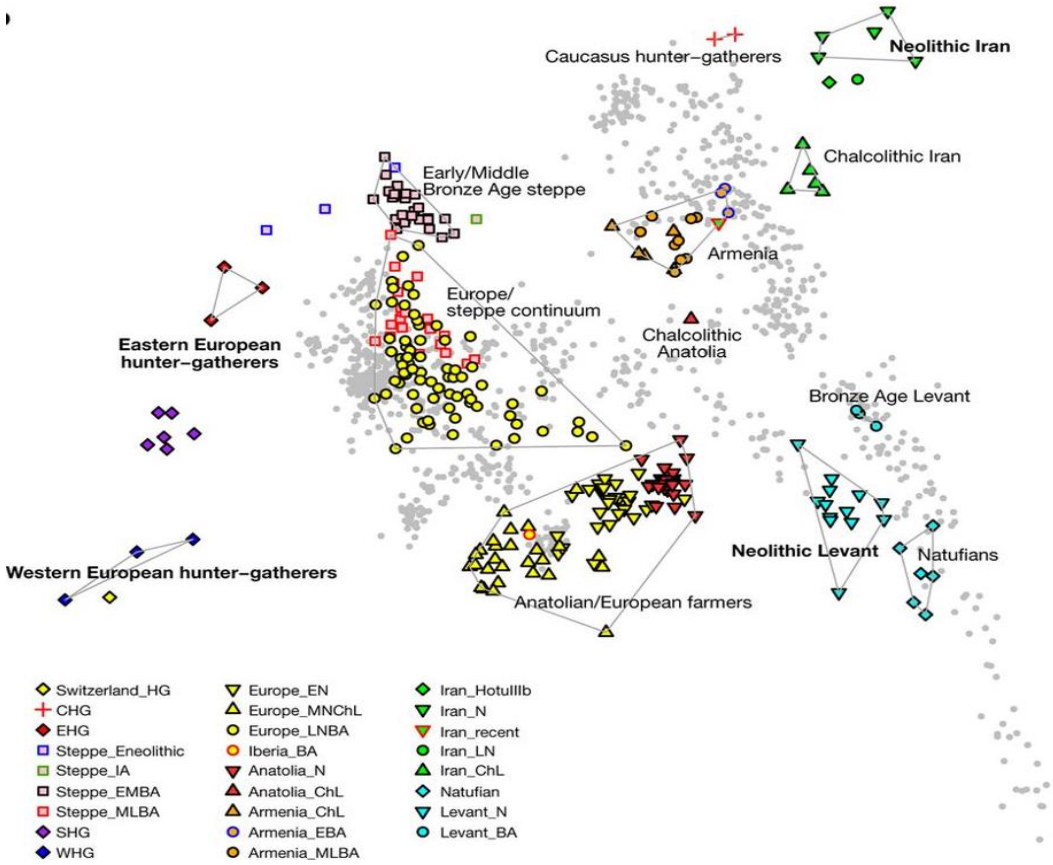


Figure 8. PCA plot of modern Eurasians (grey dots in the background) with ancient individuals spanning from the Mesolithic to the Iron Age. Symbol shapes and colours represent geographical coordinates and culture association (see legend in the lower left corner) (Lazaridis et al. 2016).

1.8.4 Neolithic revolution

The Neolithic transition started around 10,000 years ago in the Near East, more specifically in the western part of the Fertile Crescent (Whittle 1996; Price 2000), a region that spans across modern-day Israel, Palestine, Jordan, Lebanon and western Syria, into southeast Turkey and, along the Tigris and Euphrates rivers, into Iraq and the western part of Iran (Figure 9). Around 10,000 years ago this area offered a unique combination of factors: favourable climatic conditions, seasonally flooded expanses of lowland, and a natural abundance of plants and animals susceptible to domestication (Sherratt 2007; Zeder 2008; Weiss and Zohary 2011). The first population to adopt this new Neolithic lifestyle were the successors to the Natufians from the Levant (Kuijt and Goring-Morris 2002), which had already started to shift towards a progressively more sedentary way of life. The term "Neolithic revolution" was first coined by V. Gordon Childe in 1923 to describe the transition from a highly mobile hunting and gathering lifestyle to a more sedentary one, dedicated to the domestication of plants, like wheat and barley, and animals, such as goats, pigs and

cattles (Harlan and Zohary 1966; Köhler-Rollefson *et al.* 1988). With the Neolithic humans started developing more refined polished tools and producing large volumes of pottery.

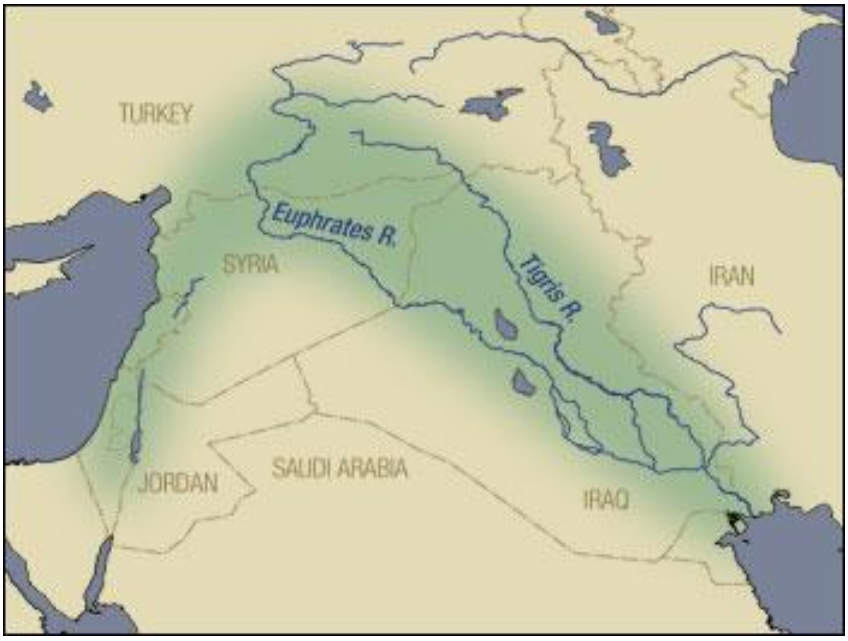


Figure 9. Map of the Fertile Crescent (Simmon R, 2003).

The change from a hunting/foraging lifestyle to one based on domesticated animals and crops had a large impact on the diet, primarily focused on cereals and, later on, on animal-derived products (Richards 2003). The reconstruction of the Neolithic farmers' dietary patterns has been mainly achieved using carbon and nitrogen stable isotope analyses (Figure 10). The isotopes consumed in the diet are directly incorporated into the organism's tissues. The ratio of carbon isotopes $^{12}C/^{13}C$ ($\delta^{13}C$) gives a view of the kind of plants consumed by the individual; for example, low values of $\delta^{13}C$ reflects a diet based on C_3 plants such as wheat, oats or rice, whereas higher values of $\delta^{13}C$ suggest a diet based on C_4 plants like sugarcane, maize and millet (Tauber 1981; Richards 2002). The ratio of nitrogen isotopes $^{14}N/^{15}N$ ($\delta^{15}N$), instead, reflects the source of animal protein in the diet and trophic level position (Richards *et al*. 2000). Low levels of $\delta^{15}N$ correspond to a diet based on terrestrial animals; higher levels of $\delta^{15}N$ to marine animals.

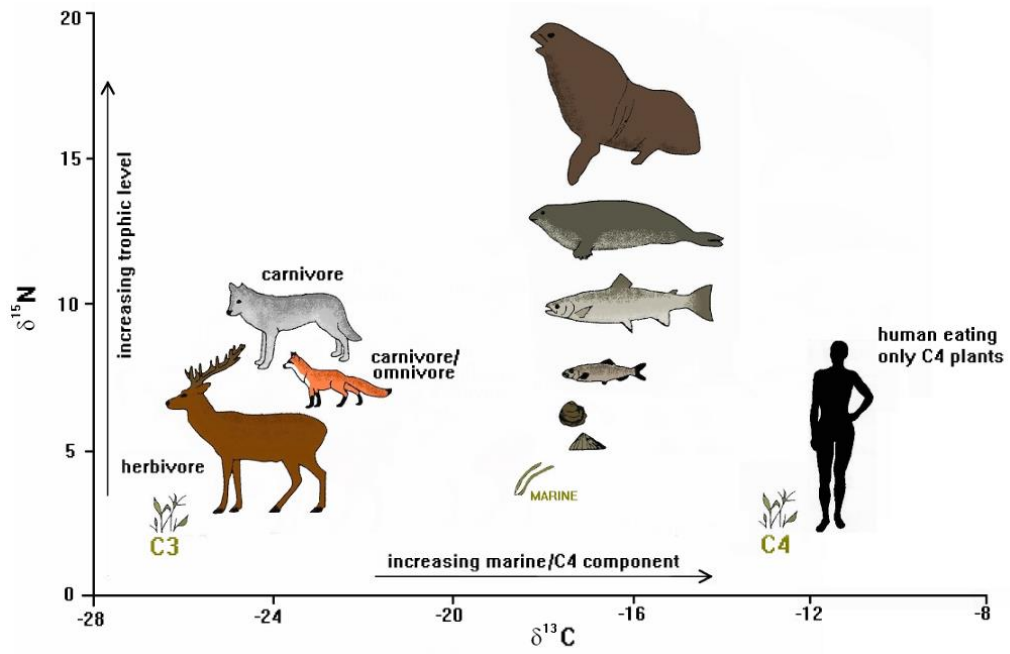


Figure 10. Stable isotopes plot used for dietary reconstruction. On the x axis the ratio of carbon isotopes $^{12}C/^{13}C$ ($\delta^{13}C$) reflects the intake of C_3 or C_4 plants. On the y axis, instead, the ratio of nitrogen isotopes $^{14}N/^{15}N$ ($\delta^{15}N$) reflects the source of animal protein in the diet, with a terrestrial versus marine based diet (Schulting 1998, modified by Dr. S Svyatko, www.chrono.qub.ac.uk/IRMS/).

An important effect of the adoption of farming as exclusive, and progressively more specialised and efficient, source of sustenance was the abundance of food available. More food led to larger families, as more people could be fed with almost no costraints from the availability of natural resources. However, growing population numbers led to a series of detrimental changes: sustaining settlements became harder, people needed more resources and highly dense populations were easy target of conflict, natural disasters and disease, related to both living in close contact with animals and nutritional disorders (Larsen 1995).

Different theories and models have been formulated regarding the spread of agriculture in Europe. From the geographical point of view, early farmers followed mainly two routes: the continental route and the Mediterranean route (Figure 11). The continental or Danubian route is supported by archaeological data found across central Europe that links it to the *Linearbandkeramik* (LBK) culture. This culture arose in the area of modern Serbia and Hungary and, following the course of the Danube, spread rapidly across central Europe (Baldia 2006; Brandt *et al*. 2013). The LBK culture was characterised by new features in ceramics technology, with fine band-decorated potteries. Their settlements were usually

located along the course of rivers or streams of water and were composed of several timber longhouses. In some cases, the settlements were surrounded by fortifications.

The Mediterranean route, instead, is characterised by the rise of another Neolithic culture, the Cardial Ware (Barnett 2000; Guldogan 2010). Archaeological evidences suggest that this culture spread along the northern coastes of the Mediterranean Sea from the Balkans, crossing the Italian Peninsula and southern France, as far as the Iberian Peninsula. A distinctive feature of the Cardial Ware culture was the production of pottery decorated with shell impressions (Price 2000; Whittle and Cummings 2007; Rowley-Conwy 2011; Özdoğan 2011).

However, genetic evidence has suggested that these two routes were not completely distinct and cultural and/or demographic interactions likely took place. A recent genetic study on the neolithization of Southeast Europe (Lipson *et al.* 2017; Mathieson *et al.* 2018) underlined the strong genetic affinities between Cardial Ware and LBK farmers.

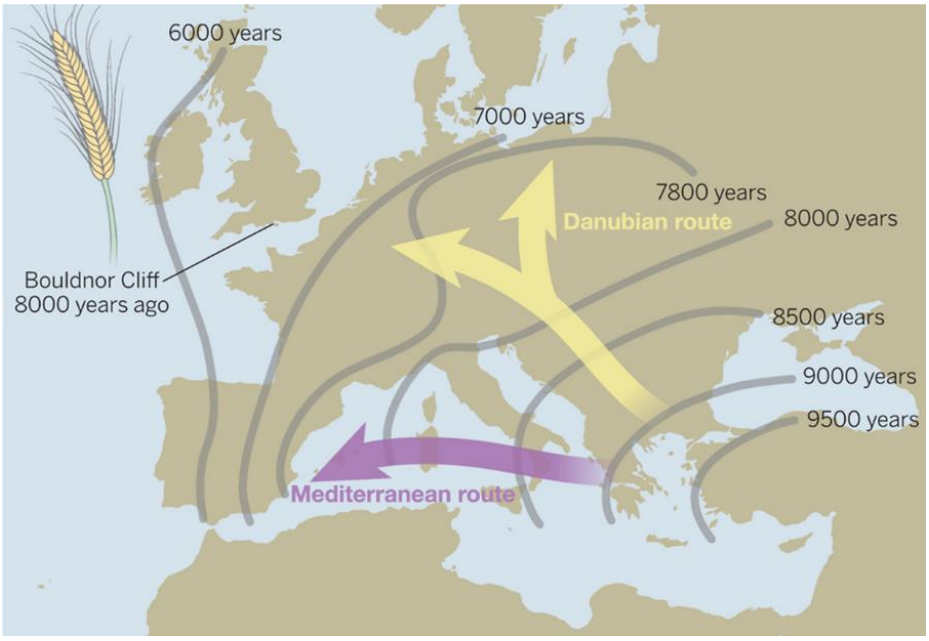


Figure 11. Expansion of farmers during the Neolithic transition (Smith et al. 2015).

Regarding the modes of dispersal, two of the most famous theories are “demic expansion” (Childe 1925; Ammerman and Cavalli-Sforza 1984), which postulates that the diffusion of agriculture was not only due to spread of ideas but also to physical movement of people carrying those ideas, and hence of genes; and “cultural diffusion” (Whittle 1996), which instead states that the social and technological changes observed in Europe during the

Neolithic were mainly due to the diffusion of ideas, with limited gene flow. However, a single model cannot explain the process across all of Europe, most likely this expansion was not homogeneous, and the speed and the mode were based on several factors, including climatic conditions and autochthonous cultures. In 1984, Zvelebil and Rowley-Conwy proposed the "agricultural frontier" model (Zvelebil and Rowley-Conwy 1984). According to this model, the transition from hunting and gathering involved pioneer or "leapfrog" colonisation but happened heterogeneously across Europe, with multiple frontiers, geographically and chronologically distinct. The Balkan Peninsula was the first European area to experience the Neolithic revolution since it was the closest geographically to the Near East. During that time, the more western and northern areas of the continent were still populated by hunter-gatherers and they kept the Mesolithic lifestyle for an additional thousand years or more until the agricultural frontier reached them (Figure 11). Even after the farming technology spread all over the continent, some hunter-gatherer groups persisted in central and northern Europe, possibly including the Limburg and La Hoguette cultures in France and Belgium and the Pitted Ware culture in southern Scandinavia (Noe-Nygaard 1989; Gronenborn 1999; Jeunesse 2000).

Following the agricultural frontier model, the adaptation to the new lifestyle can be divided into three phases: availability, substitution and consolidation (Zvelebil 1986). In the first phase, availability, Mesolithic hunter-gatherers and Neolithic farmers locally co-existed in the same area, sharing information and establishing contacts, but still mantaining economic independence. The second phase, substitution, starts with the adoption, by the hunter-gatherers, of some farming techniques. The last phase, consolidation, marks the full replacement of the mobile hunting lifestyle with the farming one, based on agriculture and animal domestication (Zvelebil 1998).

Genetic data provides evidence on the degree of the admixture between hunter-gatherers and farmers. European hunter-gatherer ancestry appears to be higher in Neolithic farmers across Europe compared to early Anatolian farmers (Figure 12), suggesting that with increasing distance from the Levant, a stronger hunter-gatherer influence and admixture occurred. The same pattern is displayed in the PCA plot in Figure 8, where European farmers occupy an intermediate position between Natufians, the ancestral population leading to farming, and hunter-gatherer populations (Gamba *et al.* 2014; Haak *et al.* 2015; Gunther *et*

al. 2015; Lazaridis *et al.* 2016). Analyses on mtDNA from Neolithic individuals across Europe showed a drastic change in haplogroup frequencies compared to the Mesolithic. The most common mtDNA haplogroups among Neolithic farmers are N1a, T2, K, J, HV, W and X. However, typical Mesolithic haplogroups, such as U5, was still present during the Neolithic (Gamba *et al*. 2012; Hervella *et al*. 2012; Sanchez-Quinto *et al*. 2012; Brandt *et al.* 2013; Hofmanová *et al*. 2016). Despite the preponderant presence of Y chromosome haplogroup I in Neolithic individuals, it is possible to observe new lineages characterising European famers, like G and H. These lineages seem to be associated with the movement of the farmers out of Anatolia since G reaches its frequency peak among Anatolian farmers, and H has been found both in Anatolian farmers and European Neolithic individuals (Mathieson *et al*. 2015; Broushaki *et al*. 2016; Lazaridis *et al*. 2016; Lipson *et al.* 2017; Mathieson *et al.* 2018).

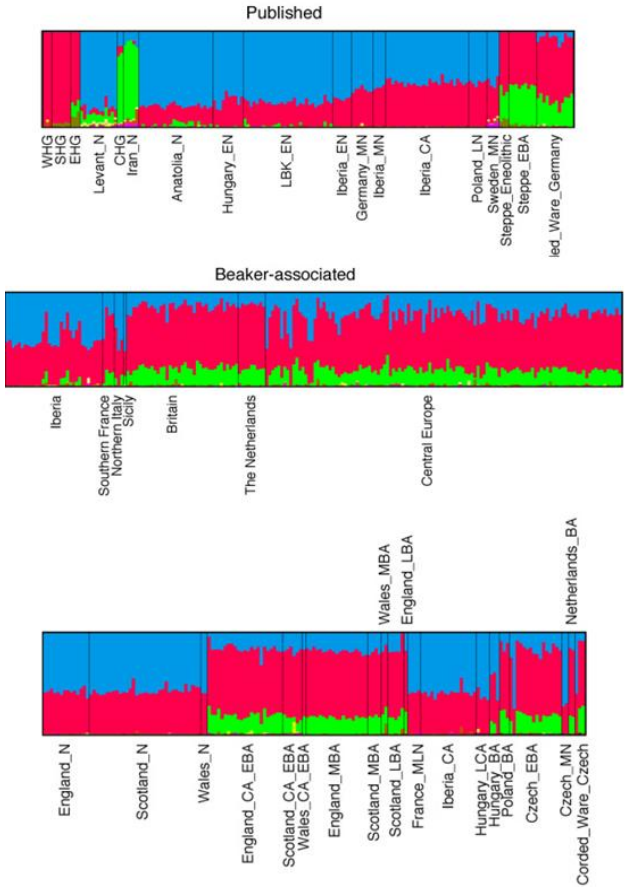


Figure 12. ADMIXTURE analysis (K=8) of ancient West Eurasian populations, from the Mesolithic to the Bronze Age (Olalde et al. 2018).

1.8.5 The dawn of the Metal Ages

Another event that drastically changed the lifestyles and cultures of modern human populations was the ability to extract and work different metals and the development of metal tools. This period is usually defined as Metal Age and can be divided chronologically, and according to the type of metal discovered and used, into three periods: Copper Age or Chalcolithic, started around 7 kya, Bronze Age appeared around 4 kya in the Balkans, and around one thousand year later along the Atlantic coasts, and Iron Age ~2 kya.

The first traces of copper in the production of tools and ornaments in Europe have been found in an archaeological site in Serbia, dated 7 kya (Bower 2010). However, the main spread of copper happened after 5 kya, with the Remedello culture in the Italian peninsula, the Bell Beaker funerary complex in Western Europe and Yamnaya in the East (Sklenář 1983; Cunliffe and Koch 2010; Beckerman 2015).

The Bell Beaker complex was named after the distinctive bell shape of their pottery. Its exact origin is still not clear, although the earliest date for a Bell Beaker vessel has been obtained so far from the Leceia site in the Estremadura region of Portugal (Cardoso *et al.* 2014); Bell Beakers covered most of western Europe, maintaining its characteristic pottery shape, but with sometimes profound variation of the decorative pattern (Czebreszuk 2004). A recent study on Bell Beaker individuals revealed that the geographical heterogeneity was also reflected in the genetic patterning of the individuals (Olalde *et al*. 2018). Bell Beakers from Iberia seemed genetically closer to Neolithic/Chalcolithic individuals from the same area, whereas Beakers from Central Europe carried a novel Pontic-Caspian Steppe-related genetic component (Figure 12), possibly as a result of the contact with the Corded Ware culture in the area of modern northern Germany, along the river Elbe (Czebreszuk 2004). The Corded Ware was a very mobile culture which based its economy on pastoralism. Its name derives from the distinctive cord-impressed pottery that has been found in graves related to the culture. Other important characteristics of this culture were the common use of single-burials and the production of battle axes (Neustupny 1969). Expanding eastward, the Corded Ware culture eventually made contact with a Bronze Age culture from the Pontic Steppe, the Yamnaya. Some hypotheses link the cultural association between the Corded Ware and the Yamnaya with the introduction of Indo-European languages into Europe (Cunliffe 2008; Anthony 2010).

The Yamnaya culture arose in the area of the Pontic-Caspian Steppe, probably around the Volga River. Archaeological records suggest that they were a highly mobile population, specialised in animal husbandry and possibly also horse riding (Anthony 2007). Two different studies, in 2015, indicated that the Yamnaya were genetically the result of an admixture between the EHG and CHG (Haak *et al*. 2015; Jones *et al*. 2015). Starting from the Late Neolithic, this Steppe-related component associated with the Yamnaya began to appear in many populations across Europe, spreading from east to west. Analyses of the Y chromosome showed an almost complete replacement in many regions of the European Neolithic haplogroups G2 and I2 with R1a and R1b, previously associated with EHG. Haplogroups R1a and R1b, which were uncommon in Europe before the Late Neolithic, drastically increase thier frequency and distribution around the continent with the possibly male-driven migrations of the Yamnaya from the Steppe (Allentoft *et al*. 2015; Haak *et al*. 2015; Mathieson *et al*. 2015).

Studies of modern Europeans show that they are largely the result of admixture between three different ancient populations, which ancient DNA has identified as Mesolithic hunter-gatherers, early Neolithic farmers and Late Neolithic/Bronze Age Steppe populations (Haak *et al.* 2015). Among present-day Europeans, northeastern Europeans tend to have the highest genetic affinity with the hunter-gatherers, while southern Europeans (especially Sardinian and Basques) tend to be genetically closer to European Neolithic farmers (Batini *et al*. 2015). However, many historical events described in the previous chapters are still not completely clear. An interdisciplinary approach is essential in order to shed light on the mechanisms which shaped the history of the European continent and its inhabitants.

1.9 Belgium: a western European case study

The area of Europe presently occupied by the Kingdom of Belgium was inhabited by hominins before the arrival of AMH in Europe, as suggested by the finding of several *Homo neanderthalensis* remains. The first specimen, named Engis 2, was found in 1829 in a cave in the western part of the country by the paleontologist Philippe-Charles Schmerling, and represents the first find of a Neanderthal fossil. At the time however, the fossil was identified as *Homo sapiens* and only in 1936 (Wood 2011), almost seventy years after the discovery of Neanderthal fossils in the Neander Valley in Germany, was it finally assigned to

the Neanderthal species. A second group of Neanderthal remains was found in the late 19th century in the Spy Cave, along the Meuse River and has been dated around 38 kya (Semal *et al.* 2009; Pirson *et al.* 2012). The last cave system holding Neanderthal remains in Belgium is Goyet, placed in the same area of the Spy Cave (Toussaint 2006). A recent study hypothesized that, according to the marks on the bones, cannibalism was practiced by the Neanderthals in the Goyet caves (Rougier *et al.* 2016). When the firsts AMHs arrived in the area around 35 kya, they occupied the same cave systems previously, or contemporaneously, used by Neanderthals, as suggested by the remains found in the Spy and Goyet caves. In 2013, Fu and colleagues analysed the genetic composition of two AMH from Goyet caves, dated around 35 kya, associated with the Aurignacian culture (Fu *et al.* 2013). Results showed that the two Goyet Upper Palaeolithic individuals belonged to a deep branch of the first founder European populations; however, they were genetically closer to Late Glacial individuals from Iberia than to other Upper Palaeolithics from Eurasia (Fu *et al.* 2013).

At the end of the Mesolithic, possibly before the arrival of the first Neolithic populations in northern Europe, hunter-gatherer communities may have adopted certain elements typical of the Neolithic, such as pottery production. Remarkable examples are Limburg and La Hoguette pottery styles (Gronenborn 2003). Both have been found in northern France, Belgium, and the Netherlands along the Rhone river valleys (Jeunesse 2002; Crombé *et al.* 2005; Hauzeur 2009). The place of origin and timing of the spread of these two different pottery styles, and their association with Mesolithic or Neolithic cultures, have been the focus of intense investigation, and as a result four main hypotheses have been formulated: a) the pottery reflects early Mesolithic adoption of Neolithic culture through direct contact with LBK groups arriving from the east (Crombé *et al.* 2005); b) local hunter-gatherers had contact with migrating groups from southern Europe, reflected in similarities with the Cardial Ware regarding the pottery style and composition (Jeunesse 2002); c) they were a product of non-LBK Neolithic nomadic herders following the Rhine valley before the establishment of the LBK (Zimmermann 2002); d) they were a local Mesolithic invention without any Neolithic influence (Gronenborn 2003).

With the Neolithic revolution, the LBK culture rapidly occupied a vast area of central Europe, mostly centered on the loess rich areas of the Netherlands, Belgium and Luxembourg,

creating settlements along the course of major rivers like the Danube, the Rhine and the Meuse. Particularly in Belgium, most of the archaeological sites found so far in the southern part of the country are present along rivers; the area is characterised by dense forest which possibly did not favour the development of farming techniques. During the Late Neolithic, the Michelsberg culture was present in the area of France, Belgium and Netherlands (Kreuz *et al.* 2014). Several excavations in the area, mainly France, gave more insight on the lifestyle of Michelsberg people. Probably they used to live, permanently or semi-permanently in small houses placed in rows and used lithic tools to prepare flour from cereals (Stika 1996; Kreuz *et al.* 2014). However, the current lack of ancient genetic data from Belgium prevents a clear view of the population dynamics during the Neolithic transition.

During the Late Bronze Age, a new culture arose in western Europe, the Hallstatt culture (Mariën 1999; Leblois 2010). This culture was characterised by a fine and intensive production of metal tools and weapons. These objects have been commonly found in graves, such as in burials in Austria, or in river beds, like the Meuse (Fontijn 2003). Their economy was mainly based on the extraction of salt and iron from caves and their trading routes followed the course of several main European rivers.

1.10 Haplogroup HV

Mitochondrial haplogroup HV is the major sub-clade of R0. It is defined by a transition (T to C) at the nucleotide position 14766 of the mtDNA genome (Torroni *et al.* 2006). Several HV subclades are characterized by mutations in the control region of the mtDNA. This region has high variability, so these mutations are often recurrent in the phylogeny in unrelated clades. A remarkable example is represented by the subclade HV+16311 (Figure 13), previously named HV3, which is defined by a transition at the nucleotide position 16,311. The same mutation appears 21 times across the entire haplogroup HV in PhyloTree (van Oven and Kayser 2009). To have a reliable classification, especially inside haplogroup HV, an analysis involving the complete mitochondrial genome is therefore essential. HV includes twelve sub-haplogroups, among those there are two major subclades: H and HV0 (Figure 13). Haplogroup HV (excluding haplogroups H and HV0) is uncommon in Europe, with frequencies between 0% and 4%, with a cline of increasing frequencies from West to East of

Europe (Karachanak *et al.* 2011). Outside Europe, instead, HV reaches a frequency of ~18% in Iranian populations (unpublished data).

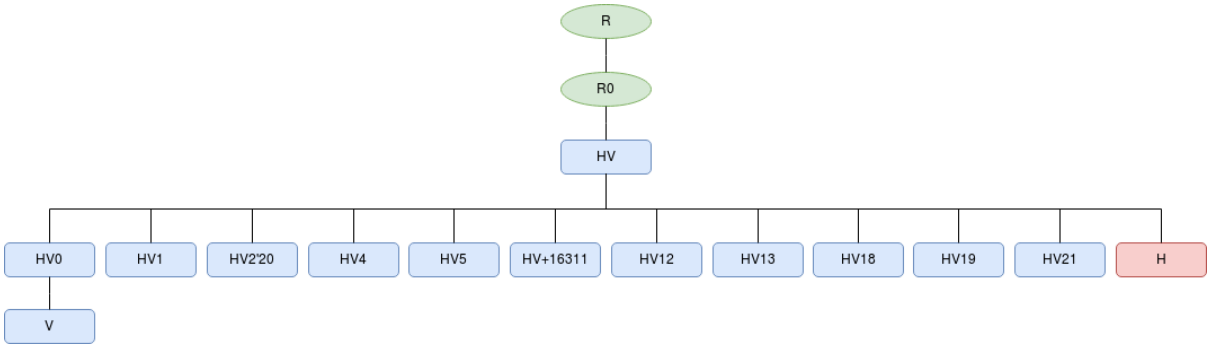


Figure 13. Schematic representation of mitochondrial haplogroup HV phylogeny, following the last version of PhyloTree (build 17) (van Oven and Kayser 2009).

The analysis of HV variation among modern populations traces the origin of haplogroup HV to the Near Eastern Upper Palaeolithic (Pereira *et al*. 2005). Most of its subclades show Late Glacial or postglacial coalescence dates, which was, at the time, interpreted by some scholars as later expansions, especially from the European glacial refugia (Torroni *et al.* 1998; Richards *et al*. 2000; Torroni *et al.* 2001; Achilli *et al*. 2004; Pereira *et al.* 2005). Additional demographic reconstructions based on modern data showed a major event of expansion of haplogroup HV in pre-Neolithic times, around 10 kya (De Fanti *et al.* 2015). However, aDNA evidence does not support this scenario. Haplogroup HV has not yet been found in ancient Europeans before the Early Neolithic, except for a single H13 individual belonging to the Iron Gates culture at the Lepenski Vir site in Serbia, dated 7785-7581 BP (Haak *et al.* 2015; Cassidy *et al.* 2016; Lazaridis *et al.* 2016; Lipson *et al.* 2017; Margaryan *et al.* 2017; Mathieson *et al.* 2018; Olalde *et al.* 2018). The major clade H shows strong demographic growth during the Neolithic (Fu *et al.* 2012; Brotherton *et al.* 2013). Given this evidence, and that the distribution and frequency in modern and ancient data of haplogroup HV and its subclade H suggest an origin in the Near East, its spread across the European continent may be related to the Neolithic revolution and migration of early farmers from Anatolia.

1.11 Aims of the project

This work plans to expand the knowledge of the pre-history of Europe and investigate population movements across the continent. To this end, two projects were carried out in

parallel, the origin and expansion of mitochondrial haplogroup HV and the process of the Neolithic in a frontier region of Western Europe.

Considering that Belgium is characterised by particular pottery cultures, such as Limburg and La Hoguette, which origins are still partially unknown, these cultures were investigated by assessing the genetic composition of ancient Belgian individuals. Therefore, Mesolithic and Neolithic human remains from three different sites were collected and both whole genome and uniparental markers analyses were performed. During this collection it was also possible to sample 166 modern Belgians that were incorporated in the mitochondrial analysis and served as modern term of comparison.

As mentioned in the previous chapter, mitochondrial haplogroup HV (including H and HV0) is the most common haplogroup amongst modern Europeans and a potential marker of Neolithic transition in Europe. However, its origin and dispersal are still poorly understood. Investigating the history of HV should help shed light on population movements across Europe from the Mesolithic until the Metal Ages. In order to do so, its distribution and frequency were analysed and new data was added to the complete published mitogenomes available in literature, amounting to a total of more than a thousand mitogenomes.

2. MATERIALS & METHODS

This project combines different experimental approaches so, to reflect this, the Materials and Methods section has been divided into two main parts: the first focuses on the ancient DNA analysis of the 38 ancient human remains from Belgium, while the second part describes the mitochondrial analysis of 166 modern Belgians and the analysis of modern published and unpublished Europeans belonging to mitochondrial haplogroup HV.

2.1 Ancient data analysis

2.1.1 Ancient sample description

To address the main question of this study, a total of 38 ancient human remains from three different sites in Belgium were collected in collaboration with the University of Liege and Bournemouth University. Fourteen samples were excavated from Trou Al'Wesse, a limestone cave in the Modave region, thirteen from Abri Sandron, an archaeological site in the Huccorgne region, and eleven from the Grotte du Mont Falise, a cave in the province of Huy (Table 2 & Figure 14). The three sites are distributed in a 10 Km radius one from the other, and are located in proximity of small tributaries of the Meuse River. As discussed in the Introduction, the skeletal elements that yield the highest amount of endogenous DNA from ancient remains are the petrous bone and teeth. For this reason, the rationale behind the sample collection was to retrieve these optimal skeletal elements. However, for many cases this was not possible due to the lack of material, so other elements were chosen for the collection.

Dr Rebecca Miller, archaeologist and collaborator at the University of Liege, sadly passed away during the progression of the project.

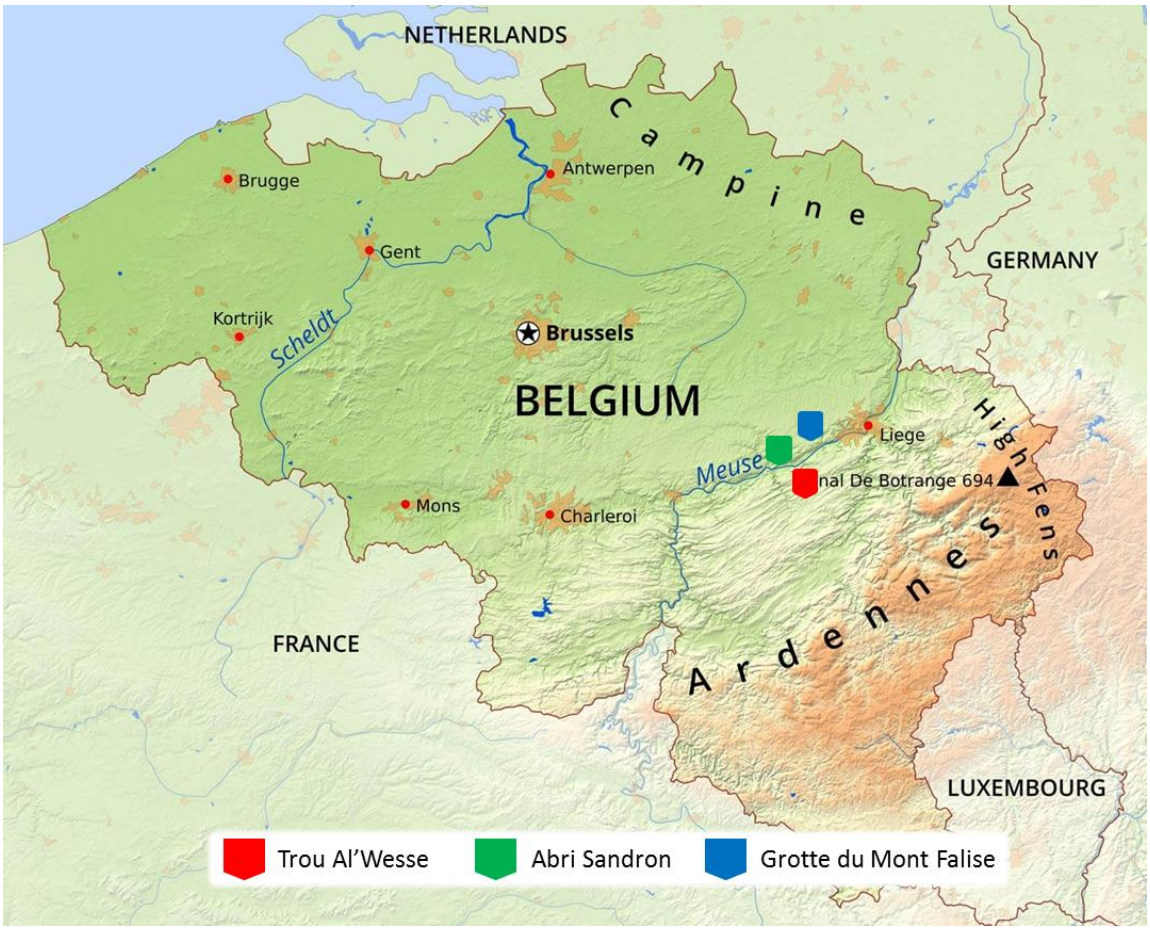


Figure 14. Physical map of Belgium containing the marked locations of the three archaeological sites.

Table 2. List of ancient human remains from the three archaeological sites in Belgium. After the collection, each sample was internally renamed with a unique code (Sample ID).

Original ID	Sample ID	Era	Site	Skeletal element[a]
1 / 3123	AF001	Neolithic	Trou Al'Wesse	molar
TAW 2010 / H13.266	AF002	Mesolithic	Trou Al'Wesse	LRM1
TAW 2010 / H13.267	AF003	Mesolithic	Trou Al'Wesse	LRM2
1 / x4	AF004	Neolithic	Trou Al'Wesse	petrous bone
3 / 3116 x2	AF005	Neolithic	Trou Al'Wesse	femur
2 / 3119	AF017	Neolithic	Trou Al'Wesse	ulna
2 / 3120	AF018	Neolithic	Trou Al'Wesse	humerus
2 / 3263	AF019	Neolithic	Trou Al'Wesse	cubitus
3 / 3267	AF020	Neolithic	Trou Al'Wesse	femur
3 / 3116	AF021	Neolithic	Trou Al'Wesse	femur
3 / 3116 x	AF022	Neolithic	Trou Al'Wesse	femur
4 / 3118	AF023	Neolithic	Trou Al'Wesse	long bone
4 / 3262	AF024	Neolithic	Trou Al'Wesse	tibia
4 / 3269	AF025	Neolithic	Trou Al'Wesse	tibia
213 / 3259	AF006	Neolithic	Grotte du Mont Falise	LLM2
213 / 3274	AF007	Neolithic	Grotte du Mont Falise	molar
213 / 15.019	AF008	Neolithic	Grotte du Mont Falise	femur
213 / 3250	AF009	Neolithic	Grotte du Mont Falise	tooth
213 / 3250	AF013	Neolithic	Grotte du Mont Falise	tooth
213 / 3252	AF014	Neolithic	Grotte du Mont Falise	molar
212 / 1x.021	AF026	Neolithic	Grotte du Mont Falise	femur
212 / x3	AF027	Neolithic	Grotte du Mont Falise	long bone
212 / 3245	AF028	Neolithic	Grotte du Mont Falise	long bone
213 / 10.006	AF029	Neolithic	Grotte du Mont Falise	humerus
213 / 3251	AF030	Neolithic	Grotte du Mont Falise	molar
97	AF010	Neolithic	Abri Sandron	LRM3
98	AF011	Neolithic	Abri Sandron	LLM2
89 / 6763	AF012	Neolithic	Abri Sandron	petrous bone
79 / 6167 / 85.03	AF015	Neolithic	Abri Sandron	petrous bone
96	AF016	Neolithic	Abri Sandron	LLM1
X1	AF031	Neolithic	Abri Sandron	molar
88 / 6165	AF032	Neolithic	Abri Sandron	molar
90	AF033	Neolithic	Abri Sandron	LLM3
91	AF034	Neolithic	Abri Sandron	LM3
93	AF035	Neolithic	Abri Sandron	LRM1
94	AF036	Neolithic	Abri Sandron	tooth
95	AF037	Neolithic	Abri Sandron	LLM2
99	AF038	Neolithic	Abri Sandron	LRM3

[a]LRM: lower right molar; LLM: lower left molar; LM: lower molar.

2.1.1.1 Trou Al'Wesse

The archaeological site of Trou Al'Wesse (Wasp Cave) is composed of two sections, a limestone cave, part of a karstic system connected to the Condroz Plateau, and the terrace below the cliff (Figure 15 and 16), overseeing the Hoyoux River, a tributary of the Meuse. The opening of the cave is currently located 50 meters from the river and 8 meters above the current water surface. The cave extends almost horizontally for 35 meters inside the cliff; at the back, the ceiling opens up through a 9 meters long chimney that connects the cave to the above plateau, the chimney is partially filled with detritus. Human remains and other archaeological findings were excavated both from the terrace and from the chimney at the back of the cave, possibly used as a burial site. The first human occupation of the site has been attributed to hunter-gatherer groups, which, as suggested by chronostratigraphic studies (Table 3), lasted until the Late Neolithic period (Derclaye *et al.* 1999; Miller *et al.* 2012; Miller unpublished). Radiocarbon dates obtained from the chimney burials were younger (~5,200 cal BP) than the dates obtained from the terrace (spanning between 10,500 and 5,600 cal BP) (Table 4), possibly reflecting a change in the function of the site, with the terrace used for herding activity and the burial at the back of the cave for funerary purpose (Miller unpublished) The dated stratigraphy of portion of the terrace excavate so far, includes the Early Mesolithic layers: 4b-alpha, 4b-beta and 4b-gamma; Late Mesolithic later: 4b-delta, Early Neolithic layers: 5a and 4b-LaH; and the Middle Neolithic layer 4a. Early Mesolithic layers, 4b-alpha, beta and gamma were probably the result of seasonal occupation of the site due to the retrieval of burnt hazelnut and wild boar remains (Miller unpublished). Ceramic sherds found in layer 4b-LaH display features that differentiate them from the LBK type and, instead, are probably associated with the La Hoguette culture (Figure 16) (Miller unpublished). AC and ACOF layers have not been dated due to lack of collagen in the samples associated with them, but they have been classified as Mesolithic due to their deposition prior to the Late Mesolithic layer 4b-delta (Figure 16). Out of the fourteen samples collected, only two were found in the terrace section of the site, in the ACOF layer, and the approximate position is known (Figure 16); the remaining twelve samples were found in the chimney burial, across all four layers (Table 4).

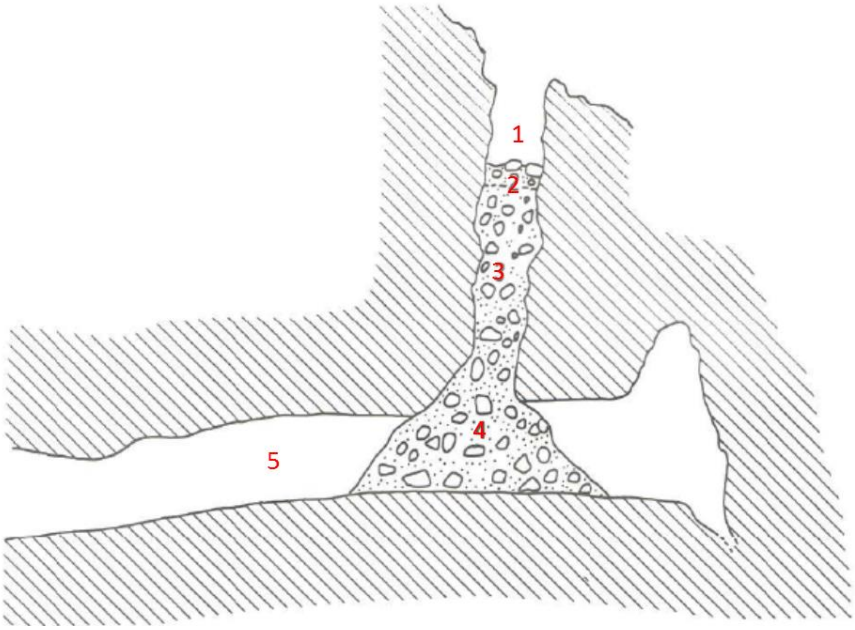


Figure 15. Stratigraphic representation of the chimney burial located at the back of the cave in Trou Al'Wesse. (1) 2 m high empty layer; (2) 0.5 m high layer; (3) 5.5 m high layer; (4) 2 m high scree cone inside the cave; (5) Limestone cave (modified from Masy 1993).

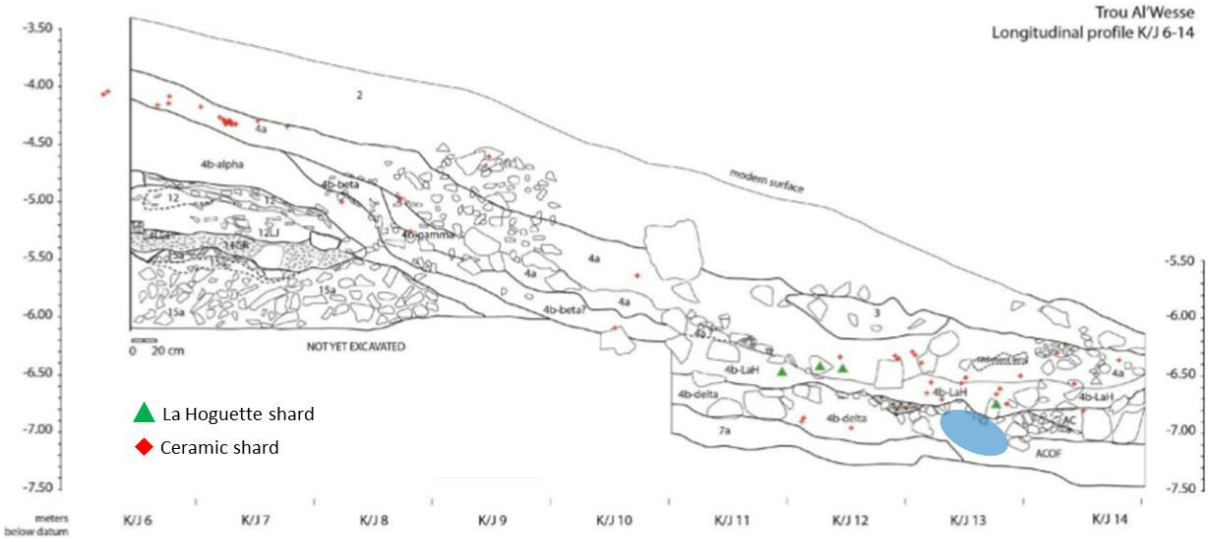


Figure 16. Stratigraphic representation of the terrace of Trou Al'Wesse with marked position of the ceramic and La Hoguette shards discovered. The blue ellipse marks the proximal position of the remains of individuals AF002 and AF003 (modified from Miller unpublished).

Table 3. Radiocarbon dates from Trou Al'Wesse (modified from Miller unpublished).

Lab number	Section of the site	Layer	Uncal BP	Cal BP	Material	Period
Beta-209871	Terrace	4b-alpha	9,000±40	10,250-10,130	Hazelnut (*Corylus*)	Early Mesolithic
Beta-224152	Terrace	4b-beta	9,240±40	10,550-10,540	Hazelnut (*Corylus*)	Early Mesolithic
Beta-224153	Terrace	4b-gamma	9,130±40	10,390-10,320	Hazelnut (*Corylus*)	Early Mesolithic
Beta-224150	Terrace	4b-delta	6,790±40	7,680-7,580	*Bos* astragalus	Late Mesolithic
Beta-251056	Terrace	4b-delta	6,890±40	7,830-7,670	*Bos* astragalus	Late Mesolithic
Lv-1752	Terrace	5a	5,950±70	6,980-6,630	Unidentified bone	Early Neolithic
OxA-26535	Terrace	4b-LaH	6,850±40	7,786-7,608	Hazel charcoal	Early Neolithic
Beta-251057	Terrace	4b-LaH	6,910±40	7,800-7,660	Burned bone	Early Neolithic
Beta-224151	Terrace	4a	4,810±40	5,600-5,560	*Canis lupus familiaris* lower molar	Middle Neolithic
OxA-7633	Terrace	4a	5,045±45	5,910-5,660	Unidentified bone	Middle Neolithic
Beta-319270	Chimney burial		4450±30	5280-5160	*Homo sapiens* 5th left	Late Neolithic
Beta-319269	Chimney burial		4560±30	5320-5280	*Homo sapiens* 5th left metatarsal	Late Neolithic

Table 4. List of samples from Trou Al'Wesse classified according to the section of the site and the relative layer where they were found.

Sample ID	Era	Section of the site	Layer
AF001	Neolithic	Chimney burial	1
AF004	Neolithic	Chimney burial	1
AF017	Neolithic	Chimney burial	2
AF018	Neolithic	Chimney burial	2
AF019	Neolithic	Chimney burial	2
AF005	Neolithic	Chimney burial	3
AF020	Neolithic	Chimney burial	3
AF021	Neolithic	Chimney burial	3
AF022	Neolithic	Chimney burial	3
AF023	Neolithic	Chimney burial	4
AF024	Neolithic	Chimney burial	4
AF025	Neolithic	Chimney burial	4
AF002	Mesolithic	Terrace	ACOF
AF003	Mesolithic	Terrace	ACOF

2.1.1.2 Abri Sandron

Abri Sandron is a sepulchral site first discovered by Julien Fraipont at the end of the 19th century AD (Fraipont 1898). The cave is located near Huccorgne, in the Liege province; it is 12 meters deep, opening into a large terrace close to the river Mahaigne, a tributary of the Meuse River. In the deepest part of the cave, several Pleistocene mega-faunal remains, human bones and pottery fragments have been found. Interestingly, among the human remains found at the site, particular care was reserved for the skulls, which were all stored inside a funerary urn. The rest of the remains, such as long bones, vertebrae, metatarsals and metacarpals were placed around the urn (Fraipont 1898). Radiocarbon dating on three human metacarpals recovered from the site suggested Late Neolithic occupation (4235±45 uncal BP / 4765-4619 cal BP; 4183±38 uncal BP / 4769-4607 cal BP; 4280±40 uncal BP / 4966-4815 cal BP) (Toussaint 2002).

2.1.1.3 Grotte du Mont Falise

The third site, Grotte du Mont Falise, is a natural cave located near Antheit (Liege province). It opens up below a limestone cliff facing the Meuse River. Stratigraphic studies on the site revealed the presence of five different layers spanning from the Middle Ages to the Upper Palaeolithic, containing human remains, pottery fragments and tools (Fraipont 1897; Fraipont 1898; Haeck 1964). Radiocarbon dating on two bones recovered from the site suggested Late Neolithic occupation (4195±40 uncal BP / 4768-4611 cal BP; 4265±40 uncal BP / 4892-4807 cal BP) (Toussaint 2003).

2.1.2 Ancient sample preparation and processing

The ancient human remains analysed in this study were collected from the University of Liege, Belgium, and shipped to the Ancient DNA Facility at the University of Huddersfield, where preparation and processing were carried out in specialised clean rooms. Full-body suits, gloves, hairnets and face masks were used for the entire time spent inside the Facility. Surfaces and tools were repeatedly treated with LookOut DNA Erase (SIGMA Life Sciences) and exposed to UV light. Prior to processing, all the samples were photographed, to keep a record of the original state, and then exposed to UV light for 30 minutes on each side (total

of 1 hour) to sterilise the surface. Each sample was then stored in a sterile plastic bag and transferred to the drilling room for processing.

2.1.2.1 Sample processing

The surface of each bone was cleaned inside a fume-hood, using a SWAM-Blaster compressed air abrasive system, with 29 µm aluminium oxide powder. The processing step from the bone to the powder varied significantly with the skeletal element analysed, and different protocols were adopted accordingly.

- In the case of post-cranial bones (femur, tibia, humerus, cubitus) the densest part was targeted, which was usually the head of the bone. Using a round-tipped drill bit for the hobby drill several holes were drilled in the bone, and the powder was collected inside sterile tubes.

- For petrous bones, a section of the dense part was excised using a diamond-tip saw; the fragment was then placed inside a grinding jar (25 ml) made of zirconium oxide. Each jar contained a grinding ball (15 mm diameter) of the same material. The jars were clamped to the designated jar holders of the Mixer Mill (Retsch MM400). After setting the required frequency of oscillations (30 Hz/s) and time (30 s) for the grinding process, the machine milled the bone section, producing very fine powder.

- To process teeth, the diamond-tip saw was used to cut the root from the crown. The root was then pulverised using the Mixer Mill, following the same procedure as described for the petrous bone.

The bone/tooth powder was transferred to 2 ml O-ring tubes, and the weight was measured. Optimal measurements were around 0.15 g. If the powder exceeded that amount, the whole was split and stored in spare tubes.

After obtaining the bone powder, the samples were transferred to the extraction room.

2.1.2.2 DNA extraction

The extractions performed followed the procotol reported in Yang *et al*. (1998); with some modifications as in MacHugh *et al*. (2000).

The first step was to prepare the extraction buffer as described in Table 5.

Table 5. DNA Extraction buffer mix.

Reagents	Volume X 1 (µl)
Tris-HCl (1 M)	20
SDS (20%)	17
EDTA (0.5 M)	950
Expose the mix to UV light for 15 minutes.	
Proteinase K (~20 mg/ml)	13
Total	**1000**

1 ml of Extraction buffer was added to each sample and the tubes were placed on a rotating wheel at 37°C for 24 hours. The rotation is essential to ensure that the solution is constantly mixed, so as to avoid formation of a pellet. The tubes were then centrifuged at 13,000 rpm for 10 minutes, the supernatant was removed and the process was repeated using a freshly made extraction buffer (as above). After a second centrifuge at 13,000 rpm for 15 minutes, the supernatant of each sample was transferred to 6 ml 30kDA columns. 3 ml of 10 mM Tris-HCl were added to each column, and the columns were centrifuged at 2,500 rpm for 30 minutes and the filtered liquid discarded. This step was repeated a second time, but, at the end, the filtered liquid, which contained DNA, was collected and transferred to sterile 2 ml O-ring tubes and stored in the fridge.

To purify the extracted DNA, the QIAQuick MinElute Purification Kit was used following the standard protocol (Qiagen), but with two PE wash steps, and elution in 100 µl of EBT buffer (the supplied EB Elution buffer, with 0.05% Tween added).

2.1.3 Library Preparation

The library preparation followed the protocol described in Meyer *et al*. (2010), with modifications as in Gamba *et al*. (2014) and Cassidy *et al*. (2016).

2.1.3.1 USER enzyme

Treatment with USER (Uracil-Specific Excision Reagent) enzyme is recommended for ancient DNA since it cleaves DNA molecules at uracil residue location removing the damage caused by deamination, which usually located at both ends of the fragments. In order to validate the genuineness of the ancient DNA, one library extract per sample did not go through this step and thus maintained the damage at the end of the fragments (See section 2.1.5.6 for further details).

16.5 µl of DNA extract for each sample were mixed with 5 µl of USER enzyme and incubated at 37°C for 3 hours.

2.1.3.2 Blunt-end repair

To repair the ends of the DNA fragments generated during the USER enzyme digestion, each 21.5 µl DNA extract was treated with 48.5 µl of blunt-end repair enzyme mix (NEBNext) (Table 6) and incubated first at 25°C for 15 minutes and then at 12°C for 5 minutes.

Table 6. Blunt-end repair mix preparation.

Reagents	Volume X 1 (µl)
End prep enzyme	3.5
End repair reaction buffer (10x)	7.0
ddH_20	38.0
Total	**48.5**

2.1.3.3 Sample clean-up

The DNA extracts were purified using the QIAQuick MinElute Purification Kit (Qiagen). 350 µl of Binding buffer (PB) were added to each sample and then centrifuged at 13,000 rpm, and the filtered liquid discarded. 700 µl of Wash buffer (PE) were added to each column and again centrifuged at 13,000 rpm for 1 minute. The columns were placed into fresh tubes and centrifuged for a further minute to allow drying. The columns were then transferred into 1.5 ml tubes and the solution eluted in 22 µl EBT buffer (Table 7), incubated for 1 minute at room temperature and then centrifuged at 13,000 rpm for 1 minute. The filtered liquid was transferred into 0.2 ml PCR strip tubes.

Table 7. EBT (Qiagen EB buffer + Tween) buffer mix.

Reagents	Volume X 1 (µl)
Qiagen EB buffer	59.97
Tween	0.03
Total	**60.0**

2.1.3.4 Adapter ligation

To proceed with the library preparation, the adapter sequences had to be ligated to the DNA molecules. 20 µl of the DNA from the earlier step was treated with 20 µl of oligonucleotide T4 ligase mix (Table 8). The mixed solutions were then incubated at 22°C for 30 minutes.

Table 8. T4 ligase mix preparation.

Reagents	Volume X1 (µl)
ddH_2O	10
T4 DNA ligase buffer (10X)	4
PEG-4000 (50%)	4
Adapter mix (100 µM)[a]	1
T4 DNA ligase (5 U/µl)	1
Total	**20**

[a]custom-made by Sigma-Aldrich (P5+P7)

2.1.3.5 Sample Clean-up

Another purification step was performed as before (2.1.3.3), but with 200 µl of Binding buffer (PB).

2.1.3.6 Adapter Fill-in

For the adapter polymerisation, 20 µl of each sample was treated with 38.2 µl of *Bst* polymerase mix (Table 9) and incubated at 37°C for 30 minutes, then at 80°C for 20 minutes.

Table 9. Bst polymerase mix preparation.

Reagents	Volume X 1 (µl)
ddH_2O	13.5
Thermopol reaction buffer (10X)	4
dNTPs (10 mM)	1
Bst polymerase, large fragments (8 U/µl)	20
Total	**38.5**

2.1.3.7 Amplification

The amplification mix was prepared by combining 42 µl of the amplification reaction mix with 2 µl of the appropriate indexing oligo to each sample (Table 10). This was added to 6 µl of the sample from the above step.

Table 10. Amplification reaction mix preparation.

Reagents	Volume X 1 (µl)
Accuprime Pfx Supermix	41
Primer IS4 (10 uM)	1
Total	**42**
+	
Indexing oligo	2

This was the final step performed in the Ancient DNA Facility, as the final PCR (Table 11) was run in the modern DNA laboratory.

Table 11. PCR protocol for ancient whole genome amplification.

Step	Temperature	Time	Cycles
Initial denaturation	95°C	5 minutes	1
Denaturation	95°C	15 seconds	12
Annealing	60°C	30 seconds	
Extension	68°C	30 seconds	
Final extension	68°C	5 minutes	1
Hold	4°C	forever	1

2.1.3.8 Sample purification

To complete the library preparation, another purification step was performed using the QIAQuick MinElute Purification Kit (Qiagen). This was undertaken as before, but with 125 µl of Binding buffer (PB) and elution in 25 µl of EBT buffer. The eluted purified library was stored in 2 ml O-ring tubes in the fridge.

2.1.4 Library size distribution and pooling

The concentration of each library for each sample was measured using the Qubit 3.0 Fluorometer (ThermoFisher Scientific).

A small volume was diluted 5 times using ddH_2O and loaded into a High Sensitivity DNA chip from Agilent Technologies and run on a BioAnalyzer. This was to ensure that successful library preparation had occurred, and that fragments of the correct length were present. Each BioAnalyzer chip could run 11 samples, with a volume of 1 µl each. The protocol provided by Agilent Technologies was used as described.

Before setting up the DNA chip, a sterile syringe was placed in the correct position into the priming station. The gel-dye mixture was prepared by adding 15 µl of High Sensitivity DNA dye concentrate (blue cap) into the High Sensitivity DNA gel matrix vial (red cap). The mix was transferred into the spin filter and centrifuged at 6,000 rpm for 10 minutes at room temperature. An unused High Sensitivity DNA chip was placed on the priming station and 9 µl of gel-dye mix were pipetted into the correct well (marked with ***G***). The priming station was closed to allow the syringe to apply pressure on the chip for 60 seconds. 9 µl of gel-dye

mix were pipetted in the other three wells marked with *G*. 5 µl of High Sensitivity DNA marker (green cap) were added to the remaining twelve wells of the chip and 1 µl of High Sensitivity DNA ladder (yellow cap) into the well marked with a ladder symbol. In each of the remaining wells (eleven in total), 1 µl of library sample was added. The chip was mixed using a vortex at 2,400 rpm for 60 seconds and subsequently analysed using the BioAnalyzer.

2.1.5 Bioinformatic data processing

Pair-end shotgun sequencing, which produces partially overlapping sequences from both ends of each fragment, was performed at Macrogen, Rep. of Korea, on an Illumina HiSeq4000. Illumina HiSeq is based on a clonal amplification method, which generates clusters of DNA templates at high densities on a glass slide (flow cell). Templates are then sequenced using a sequencing-by-synthesis approach, whereby reversible terminator nucleotides for the four bases, each labelled with a different fluorescent dye, are used. The output data was returned in the format of two .fq.gz files per sample. The screening of the samples consisted of sequencing the 38 individuals divided across four lanes, with one USER-treated library per sample. After the initial screening, the six samples with the highest endogenous DNA were chosen in order to have data from both time periods (Mesolithic and Neolithic) and from all three archaeological sites. Further sequencing was performed with two samples per lane, and three libraries per sample, for a total of three lanes. One library of each sample was not USER treated, so as to assess DNA degradation.

The remaining 32 samples were sent for SNP capture sequencing to the Department of Genetics, Harvard Medical School, Boston, to the laboratory led by Professor David Reich. SNP capture method consists of sequencing target regions or specific positions using customised DNA chip arrays complementary to the genome.

2.1.5.1 Trimming and merging

Adapters were trimmed and the two Illumina sequencing reads merged using leeHom (Renaud *et al*. 2014). One advantage of merging sequences from the same sample is to increase the sequence depth and to reduce the chances of mis-calling sequencing errors during the later SNP calling steps. The trimming process is essential to avoid the presence of adaptor sequences at both ends of each read during mapping and assembly. LeeHom is an algorithm based on a Bayesian maximum *a posteriori* probabilistic approach. It has been

chosen for its speed, capability of processing large volumes of data, and accuracy handling short sequences, as is the case for ancient DNA. The command used in the trimming and merging process and its relative options are described in Appendix 1.

2.1.5.2 Mapping

After obtaining the merged fastq file, mapping against the reference genome was carried out for each merged library using BWA (Burrows-Wheeler Aligner) (Li and Durbin 2009). BWA is widely used software to map sequences against a reference genome. It consists of multiple algorithms specific for different sequence lengths. Originally BWA was developed to exclusively align modern data but, with modifications first introduced by Schubert and colleagues in 2012, it is now possible to overcome this limit and work with highly damaged and fragmented reads (Schubert *et al*. 2012). In the case of ancient DNA, where the reads are tipically shorter than 100 bp, the algorithm suggested and widely used is BWA-backtrack (Li and Durbin 2009).

The reads were mapped to the human genome reference hg19 (Human genome build version 19). To obtain more coverage on the mitochondrial genome, the sequences were separately mapped exclusively against rCRS (mitochondrial genome reference). The alignments were generated in the .sam format as single-end reads (bwa samse). The commands used in the mapping process and their relative options are described in Appendix 2.

2.1.5.3 Mapping quality filtering

After obtaining the alignments for each library, a mapping quality filtering was applied using the software Samtools. Samtools contains several options but, in this pipeline, it has been used to filter and sort the reads. A mapping quality of 20 was set as threshold. In order to proceed with the following steps each .bam file had to be sorted. The command used in the mapping quality filtering and its relative options are described in Appendix 3.

2.1.5.4 Marking duplicate reads

In order to remove PCR duplicates, which could contain random errors introduced during the library amplification or sequencing, the “marking duplicates” step was performed. The marked duplicates were retained inside the file but “flagged”, and were not considered for the SNP calling process. The result of this filtering step was a significant drop in the number

of reads, as only the reads with the highest quality and reliability were kept, mitigating the effects of errors. The tool used was Picard, software developed in Java to manipulate sequencing data (Picard Toolkit, Broad Institute). The command used in the mark duplicates step is described in Appendix 4.

2.1.5.5 Read length filtering

In the command described in Appendix 5, samtools was used to scan the information inside the file and export only the reads that had a length above 30, set as a threshold for good quality read legth in aDNA as per Schubert *et al*. (2012).

2.1.5.6 Assessing damage

The software mapdamage 2.0 was used to assess the presence of deamination at the end of the reads, which is a diagnostic feature of aDNA (Jonsson *et al.* 2013). Both USER-treated and not USER-treated libraries were assessed for damage. The resulting reports show a very distinctive pattern: in non USER-treated libraries (Figure 17a), the mapdamage report shows an increased frequency of C->T transitions at the 5' end and G->A at 3' end of the reads, this is caused by the deamination process. In USER-treated libraries (Figure 17b), the report shows that the effect of the deamination has been heavily reduced by the treatment. See Appendix 6 for the command used to generate the mapdamage report.

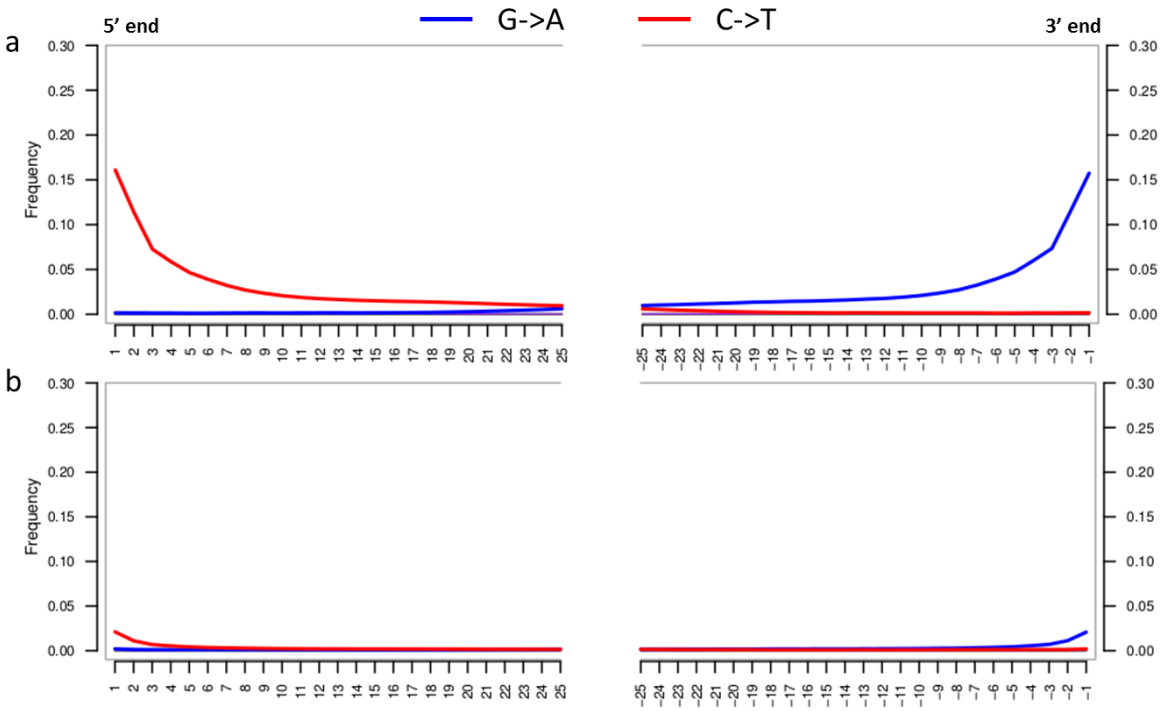


Figure 17. Mapdamage report of the USER-treated (b) and untreated (a) library of individual AF010. The x axis shows the nucleotide position along the first 25 bases of the read; the y axis indicates the frequency of the nucleotide base at a certain position of the read.

2.1.5.7 Softclipping

In order to call genuine SNPs, and avoid consideration of false calls caused by deamination, the .bam files were softclipped. Softclipping is a procedure where the last few bases at the ends of each read are modified into "N", so they are not considered during SNP calling. The number of bases softclipped changed according to the mapdamage report, and untreated libraries had a higher number of bases softclipped compared to USER-treated ones.

To softclip the reads, the software bamUtil-master was used (Jun *et al.* 2015). The command used with its relative options is described in Appendix 7.

2.1.5.8 Qualimap report

The Java platform Qualimap 2.2.1 (Okonechnikov et al. 2015) was used to assess the quality of each library after applying all the filtering processes described above. Qualimap is an application that examines the alignment file and provides an overall report on the content and quality of it, including number of total, mapped and duplicated reads, GC content, mean coverage and mapping quality. The command used to run Qualimap with its relative options is described in Appendix 8.

2.1.5.9 Merging bam files

Once all quality checks were completed, libraries from the same sample were merged together in one .bam file by using the function MergeSamFiles from the software Picard (Picard Toolkit, Broad Institute). The command used to merge the bam files with its relative options is described in Appendix 9.

2.1.5.10 SNP calling

The SNPs for each sample were called using GATK (Genome Analysis Toolkit) HaplotypeCaller (McKenna *et al.* 2010). The reference SNP arrays used in this study were the Affimetrix Human Origins array with 600K SNPs (Lazaridis *et al.* 2014) and the 1240K array with 1.2M SNPs (https://reich.hms.harvard.edu; David Reich Lab). The Affimetrix Human Origin dataset was used to perform the PCA, IBD and ADMIXTURE analyses. The 1240K dataset, instead, was used to perform IBD and D-statistics. Around 500K SNPs are shared between the two lists. The command used to call the SNPs for each genome is described in Appendix 10.

2.1.6 Sex assignment and uniparental lineage classification

Sex assignments were estimated using the method described by Skoglund *et al*. (2013), calculating the number of reads aligned to the Y chromosome as a fraction of the total number of aligned reads to both X and Y chromosomes (Skoglund *et al*. 2013). The Y chromosome haplogroup classification for the male individuals was performed using the software Y-leaf (Ralf *et al.* 2018). Due to the low coverage of the data, in many cases it was possible to define only the root of the haplogroup without a more precise classification.

The alignments (bam files) were manually inspected using the Integrative Genomic Viewer (IGV). The classification followed the latest version of the human mitochondrial phylogeny, Phylotree build 17 (van Oven and Kayser 2009). See Results section for both Y chromosome and mtDNA haplogroup classification.

2.1.7 Principal Component Analysis (PCA)

A bidimensional plot was generated using high quality and high coverage genotypic data of modern human individuals; subsequently, the ancient data was projected into the newly created reference plot. The command is described in Appendix 11.

The output, represented by a table, was imported in Rstudio and graphically plotted, using the plot package of R (See Results section).

2.1.8 ADMIXTURE analysis

In the ADMIXTURE analysis the number of ancestral populations is represented by the variable *K*. Gradually increasing *K*, it is possible to observe a more complex genetic structure and shared ancestries of the populations analysed. A cross-validation (CV) procedure is integrated into the software that allows the error estimation of the predictive model (Alexander *et al*. 2009). The analysis was repeated ten times for each *K*, from *K* = 2 to *K* = 15, and the *K* with the lowest CV error was chosen. The command used to run the ADMIXTURE analysis is described in Appendix 12. The resulting table was imported in Rstudio and the containing data plotted using the barplot package of R (See Results section).

2.1.9 Identity By Descent (IBD)

In order to assess the presence of duplicated samples, or first/second/third degree relationships among the individuals analysed, Identity By Descent (IBD) analysis was performed. IBD is a method to infer pairwise relatedness using DNA sequences. The software PLINK was used (Chang *et al.* 2014). Two individuals were defined as "identical by descent" when they shared an identical portion of DNA inherited by a recent common ancestor; the larger the scale of identical segments, the closer the relationship between the individuals was. See Appendix 13 for a description of the command used.

2.1.10 *f*-statistics

f-statistics were calculated via ADMIXTOOLS (Patterson *et al*. 2012) to infer admixture between populations using allele frequency correlations. *f*-statistic test calculates the amount of genetic shared drift between the populations analysed, three in the case of f3-statistics or four in the case of D-statistics.. When one of the populations is replaced by a known outgroup population (Yoruba in this case), the test calculates the degree of admixture between the target population X and two given populations A and B in the case of D-statistics, or between the target population and a given reference population in the case of f3-statistics. In the format *D*(Yoruba, X, pop A, pop B), a positive value of the test indicates admixture between the target population X and the population B. Instead, when a negative value is obtained, there is sign of admixture between the target population X and

population A. Values close to zero indicate that populations A and B are genetically closer to each other than the target X (Patterson *et al.* 2012). Instead, in the format f3(Yoruba, pop A, pop B), the values of f3 indicate the amount of shared drift between pop A and pop B.

The command used to run *f*-statistic test is described in Appendix 14. The output file of the *f*-statistics test was represented by a data table containing all the different combinations of populations used in the analysis to estimate admixture, the degree of admixture, the standard error, the Z-score and the number of SNPs used in each combination. The results were plotted in R using the admixturegraph package (https://github.com/mailund/admixture_graph).

2.1.11 Dietary stable isotopic analysis

The collagen for carbon and nitrogen isotopic analysis was extracted from tooth root or bone powder of all the individuals.

The dietary stable isotopic analysis of carbon and nitrogen was performed by Dr. Peter Ditchfield at the Research Laboratory for Archaeology at the University of Oxford. The isotopic data of only 19 of the 38 individuals is presented in the Results section, as the remaining samples have yet to be analysed.

2.1.12 Phenotypic analysis

Alignments were manually inspected using the Integrative Genomic Viewer (IGV). The phenotype prediction for skin pigmentation, eye and hair colour was performed using the HiriSplexS platform (Walsh *et al.* 2014; Walsh *et al.* 2017; Chaitanya *et al.* 2018). The prediction was based on 41 SNPs across the genome. Due to the low coverage of the samples, it was possible to retrieve data for only six individuals, most of whom had most of the 41 SNPs covered.

The lactase persistance prediction was based on two SNPs under selection in modern European populations, rs4988235 on the *LCTa* gene and rs182549 on the *LCTb* gene (Bersaglieri *et al.* 2004). Both genes regulate the function of the lactase enzyme during childhood and adulthood. It was possible to retrieve data for at least one of the two SNPs in 14 individuals; only three individuals had both SNPs covered.

2.1.13 Network analysis

Phylogenetic networks were drawn using the Network software (*www.fluxus-engineering.com/sharenet.htm*). The reduced median and median-joining algorithms were applied, as described in Bandelt *et al.* (1995, 1999). The application of these algorithms allowed building a network with the most likely links between the individuals.

The output file was graphically edited using Network Publisher (*www.fluxus-engineering.com/sharenet.htm*).

2.2 Modern data analysis

In order to verify a possible genetic continuity thorough time in Belgium, 166 modern samples were collected during a two day sampling trip at the University of Liege. Donors were selected by having at least two generations of Belgian or Luxembourgish ancestry. The sampling followed ethical approval, where ancestry information was collected anonymously and personal data could not be correlated with the genetic data. Also there were no medical or commercial consequences of taking part in the study. And participation was voluntary and without risk.

The dataset used, instead, for the study of mitochondrial haplogroup HV (see 1.11 for further details) included: 175 unpublished complete sequences produced in our laboratory at the University of Huddersfield, 213 unpublished complete sequences produced at the Institute of Genomics, University of Tartu, thank to collaboration with the research group led by Dr Mait Metspalu and 1101 published complete sequences.

2.2.1 DNA extraction of 166 modern Belgians

The 166 modern Belgian DNA samples (code LIE) were extracted from cotton buccal swab, using PureLink® Genomic DNA Kits (Invitrogen), following the protocol here described.

The buccal swabs were placed in sterile 1.5 ml tubes with 400 µl of PBS (Phosphate-Buffered Saline), 20 µl of Proteinase K and 420 µl of PureLink Genomic Lysis/Binding buffer. The samples were incubated at 55°C for 10 minutes. 200 µl of 96-100% ethanol were added to each tube. The lysates were transferred to PureLink Spin Columns and centrifuged at 10,000

g for 1 minute at room temperature. The flow-through was discarded. 500 µl of Wash buffer 1 was added to each column and centrifuged again at 10,000 g for 1 minute. The "wash" step was repeated a second time using 500 µl of Wash buffer 2, and the columns were centrifuged at maximum speed for 3 minutes. The spin columns were placed inside 1.5 ml tubes. 100 µl of Elution buffer were added to each column and then the samples were centrifuged at maximum speed for 1 minute. The eluted solutions containing DNA were stored at -20°C.

2.2.2 DNA amplification by long range PCR

Different sets of primers were tested in order to minimise the number of PCR fragments, while at the same time covering the entire mitochondrial genome. Two partially overlapping fragments were chosen, using a long-range PCR protocol. GoTaq® Long PCR Master Mix (Promega) was used. The mix contains all the reagents required, including a hot-start *Taq polymerase*, buffer, dNTPs and $MgCl_2$. One of the main features of this system is the hot-start recombinant polymerase, which has an activity that is inhibited at lower temperatures, giving time to set-up reactions at room temperature. Activity is restored after the initial denaturation step, which is carried out at high temperature (94°C).

All PCR reactions were performed using the oligonucleotide primers in Table 12.

A new protocol was established, which halved the volumes of the GoTaq Long PCR Master Mix required, without compromising the result of the experiment. This allowed doubling of the reactions that could be undertaken with each kit (Table 13).

Table 14 describes the PCR reaction protocol optimised to work with the same efficiency for both fragments.

Table 12. Oligonucleotides used to amplify the entire mitochondrial genome in two overlapping fragments.

Fragment	Oligo name	Sequence (5' -> 3')	Tm (°C)	GC content
1	5871 for	gcttcactcagccattttacct	58.4	45.5 %
	13829 rev	agtcctaggaaagtgacagcga	60.3	50 %
2	13477 for	gcaggaatacctttcctcacag	60.1	50 %
	6345 rev	agatggttaggtctacggaggc	60.9	50 %

Table 13. Long range PCR mixture.

Component	Starting concentration	Volume (µl)
Nuclease-free Water		11
GoTaq Long range PCR MasterMix (Promega)	2X	12.5
Forward Primer	10 pmol/µl	0.5
Reverse Primer	10 pmol/µl	0.5
DNA		1
Total		**25.5**

Table 14. Long range PCR reaction protocol.

Step	Temperature (°C)	Time
1. Initial denaturation	94	2 min
2. Denaturation	94	30 sec
3. Annealing	55	30 sec
4. Extension	65	9 min
Repeat from step 2 to step 4 for 30 times		
Final extension	72	10 min
Final hold	4	Forever

2.2.3 Gel electrophoresis

To assess the quality of the PCR products 3 µl of PCR product were mixed with 2 µl of Blue/Orange loading dye (Promega) and loaded on a 1% agarose gel, stained with Midori Green (NipponGenetics). The electrophoresis run was performed in 0.5X TBE buffer (0.89 M Tris-HCl, 0.89 M boric acid, 20 mM EDTA), and with an electrical voltage of 80 volts for 60 minutes. The amplified products were visualised under UV light and the sizes of the fragments were determined by comparison to DNA ladder of known length (Quick load purple 1 kb DNA ladder, Biolabs).

2.2.4 MtDNA purification

The two PCR products were purified with Wizard® SV Gel and PCR Clean-Up System (Promega), following the standard protocol. This system is optimised to purify PCR products of 100 bp to 10 kb. PCR products are commonly purified to remove excess nucleotides and primers.

2.2.5 Sample preparation and sequencing

The concentration of the purified PCR products was quantified using a Qubit 3.0 Fluorometer (ThermoFisher Scientific). The PCR products were then normalised at a common concentration of 1 ng/µl, and both fragments for each sample were merged together in a single solution. Each pooled sample was then sent to the Earlham Institute, Norwich, England, where library preparation and sequencing were carried out on an Illumina MiSeq. The raw product of the sequencing process consisted of two fastq files for each sample.

2.2.6 Bioinformatic analyses of modern mitochondrial data

2.2.6.1 Mitochondrial haplogroup classification

The fastq files were aligned to the rCRS reference genome using bwa mem (Li and Durbin 2009) and the duplicates removed using samtools (Li *et al.* 2009). The SNP calling process was performed by GATK (Genome Analysis Toolkit) HaplotypeCaller (McKenna *et al.* 2010). The haplotypes were extracted from the vcf files and the haplogroup classification was obtained using Haplogrep (Kloss-Brandstätter *et al*. 2011; Weissensteiner *et al.* 2016).

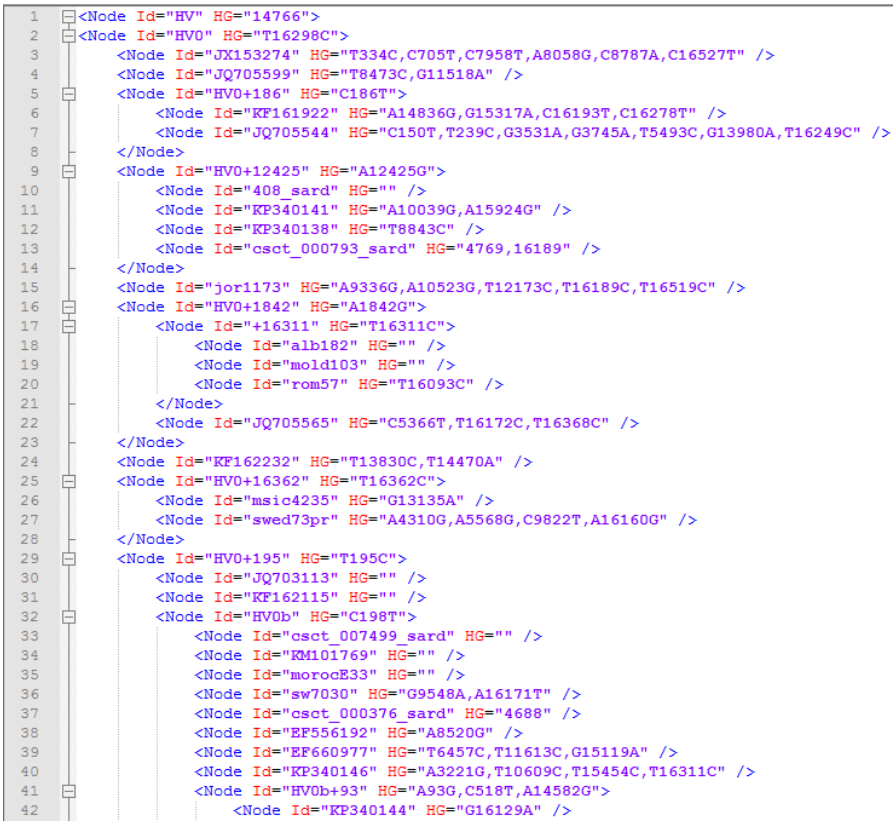
2.2.6.2 Age calculation

Coalescence times were calculated using both ρ statistic, defined as the average distance of the haplotypes of a clade from the respective root haplotype (Saillard *et al.* 2000), and with a maximum likelihood approach, using PAML (Phylogenetic Analysis by Maximum Likelihood) (Yang 1997, 2007). Mutational distances were converted into years using the substitution rate corrected for purifying selection for the human overall mtDNA genome (Soares *et al*. 2009).

2.2.6.3 Tree construction

The phylogenetic tree for haplogroup HV, containing published and unpublished complete modern mitogenomes, was created using mtPhyl (*https://sites.google.com/site/mtphyl/home*). mtPhyl compares the mtDNA sequences with the rCRS reference and creates a maximum parsimony phylogenetic tree using the PhyloTree reference tree, build 17 (van Oven and Kayser 2009). A maximum parsimony tree

minimises the number of mutations across the phylogeny. Each branch was manually reviewed and the final tree was exported in an xml format (Figure 18).



```
<Node Id="HV" HG="14766">
<Node Id="HV0" HG="T16298C">
    <Node Id="JX153274" HG="T334C,C705T,C7958T,A8058G,C8787A,C16527T" />
    <Node Id="JQ705599" HG="T8473C,G11518A" />
    <Node Id="HV0+186" HG="C186T">
        <Node Id="KF161922" HG="A14836G,G15317A,C16193T,C16278T" />
        <Node Id="JQ705544" HG="C150T,T239C,G3531A,G3745A,T5493C,G13980A,T16249C" />
    </Node>
    <Node Id="HV0+12425" HG="A12425G">
        <Node Id="408_sard" HG="" />
        <Node Id="KP340141" HG="A10039G,A15924G" />
        <Node Id="KP340138" HG="T8843C" />
        <Node Id="csct_000793_sard" HG="4769,16189" />
    </Node>
    <Node Id="jor1173" HG="A9336G,A10523G,T12173C,T16189C,T16519C" />
    <Node Id="HV0+1842" HG="A1842G">
        <Node Id="+16311" HG="T16311C">
            <Node Id="alb182" HG="" />
            <Node Id="mold103" HG="" />
            <Node Id="rom57" HG="T16093C" />
        </Node>
        <Node Id="JQ705565" HG="C5366T,T16172C,T16368C" />
    </Node>
    <Node Id="KF162232" HG="T13830C,T14470A" />
    <Node Id="HV0+16362" HG="T16362C">
        <Node Id="msic4235" HG="G13135A" />
        <Node Id="swed73pr" HG="A4310G,A5568G,C9822T,A16160G" />
    </Node>
    <Node Id="HV0+195" HG="T195C">
        <Node Id="JQ703113" HG="" />
        <Node Id="KF162115" HG="" />
        <Node Id="HV0b" HG="C198T">
            <Node Id="csct_007499_sard" HG="" />
            <Node Id="KM101769" HG="" />
            <Node Id="morocE33" HG="" />
            <Node Id="sw7030" HG="G9548A,A16171T" />
            <Node Id="csct_000376_sard" HG="4688" />
            <Node Id="EF556192" HG="A8520G" />
            <Node Id="EF660977" HG="T6457C,T11613C,G15119A" />
            <Node Id="KP340146" HG="A3221G,T10609C,T15454C,T16311C" />
            <Node Id="HV0b+93" HG="A93G,C518T,A14582G">
                <Node Id="KP340144" HG="G16129A" />
```

Figure 18. Extract of the HV xml tree, showing branches nodes, samples ID and defining mtDNA mutations.

2.2.6.4 Founder Analysis (FA)

Founder Analysis (FA) is a phylogeographic approach that uses DNA sequences to identify migration events, estimate arrival times. It quantifies the contribution of each migration from a defined source population (source) into one or many receiving population(s) (sink) (Richards *et al*. 2000) using ρ statistics, which represents the mean number of mutations occurred from the inferred ancestral haplotype of a given clade (Forster *et al*. 1996; Saillard *et al*. 2000). To overcome the possibility of recurrent mutations common in the mitochondrial phylogeny, and back-migration, which can alter the results of the analysis, different criteria of stringency have been developed, called *f*1 and *f*2 (Richards *et al*. 2000). The difference between the two stringency parameters is that *f2* is more stringent, and therefore more accurate, than *f1,* but the downside is that it requires a higher amount of "founder clades" in the sink population. When this requirement is not fulfilled, *f*1 tends to show a more reliable pattern.

The FA software requires two input files that are:

- A phylogenetic tree in the form of an xml file (Figure 18).
- A two-column list, in .txt format, of the dataset that will be used in the calculation. The first column contains the sample ID; the second one contains a code indicating to which group, “source” or “sink”, each sample belongs.

The parameters used to reproduce and estimate the migration events involving mitochondrial haplogroup HV used in this study were as follows:

- The founder haplotype must be represented by at least one (*f*1) or two (*f*2) derived branch(es) in the founder population; this can cause a different distribution of *f*1 and *f*2, where *f*2, which is usually considered more precise, is more affected by sample size, so in many cases both results must be taken in account.
- Since the coalescence age of haplogroup HV is estimated to be ~20 kya, a time frame from modern time until 20 kya, with a 200 year interval for each migration, was set for the analysis (Figure 19).
- A mutation rate of 1 mutation every 2,600 years (Fernandes *et al.* 2015; Gandini *et al.* 2016) was used.

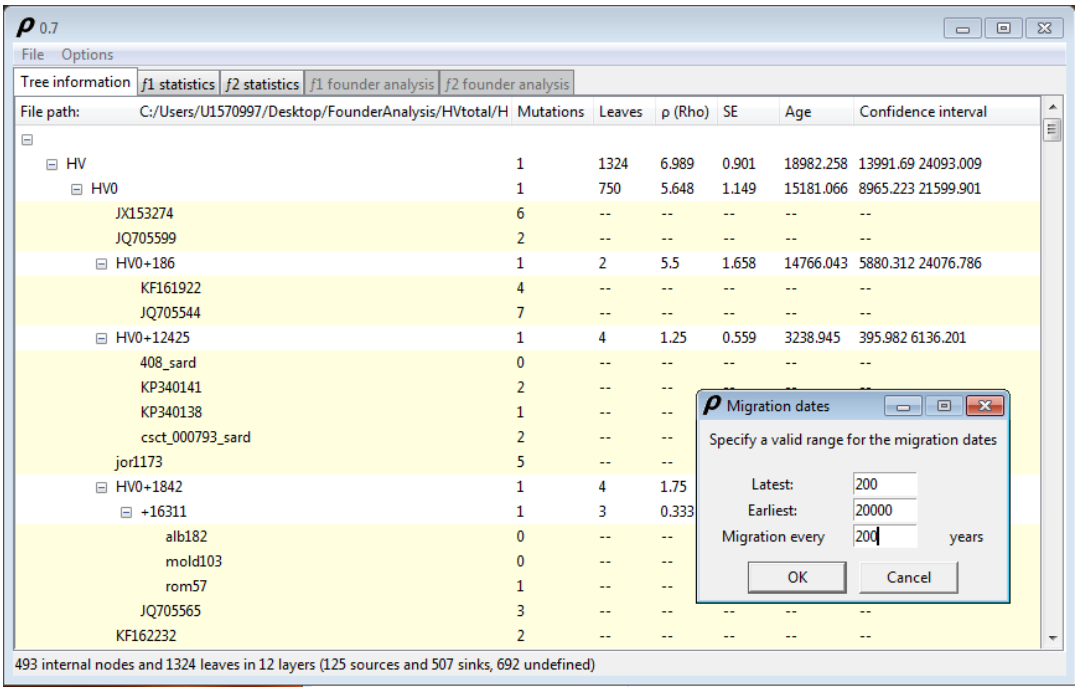


Figure 19. Founder Analysis (FA) software. Example of HV tree information and migration date range.

After loading the two input files in the software and having set the desired parameters described above, it is possible to run the calculation and export the result represented by a graph that shows on the *x*-axis the Migration date (years) and on the *y*-axis the percentage of lineages (0-1) (see Results section).

3.1 Ancient Belgium

After the initial screening, the six individuals chosen for deeper sequencing were: the Mesolithic AF002 and AF003, AF001, AF015, AF025 and AF029. The remaining 32 samples (Table 15) were sent to the Department of Genetics, Harvard Medical School, Boston in the laboratory led by Professor David Reich, for SNP capture. Only five samples did not pass the quality and contamination filtering, giving a total of 33 individuals (Shotgun and SNP capture data) that have been used in the analyses (Table 15). The endogenous DNA content, calculated as the fraction of mapped reads (after applying quality filters) on the total number of raw reads, is present exclusively for the samples sequenced using the Shotgun technique because only in that case the fastq files were available. The mean coverage for the whole genome is present exclusively for the samples sequenced with the Shotgun technique. Since SNP capture data is the product of the sequence of a determined combination of SNPs in the chip array, the coverage would represent the fraction of those SNPs to be covered and not the whole genome. However, the mean coverage for the mtDNA, could be calculated because the SNP array for the capture technique covers the entire molecule (1-16569). The highest value of endogenous DNA content, 55%, is achieved by the samples AF015. The skeletal element of AF015 was a petrous bone, the best candidate to preserve DNA in ancient remains. Surprisingly, AF029, represented by a humerus, yielded the second highest score of endogenous DNA content, 12%. They were followed by AF001 (molar) with 2.7%, AF003 and AF025 both yielding 2% of endogenous DNA, represented by a molar and a tibia, respectively. The lowest score, 1%, was achieved by AF002 whos skeletal element was a loose lower right molar.

The genetic sex assignment was performed using the Skoglund *et al.* 2013 calculation as described in section 2.1.6. 28 individuals out of 33 (~85%) were genetically classified as males; the remaining 5 individuals were genetically classified as females (15%).

*Table 15. Summary of the whole genome sequencing results for the 38 ancient samples from Belgium. Mesolithic samples are in blue. *indicates failes samples.*

Sample ID	Archaeological site[a]	Endogenous DNA content (%)	Mean coverage	MtDNA coverage	MtDNA haplogroup	Y chromosome haplogroup	Sex prediction	SNPs Human Origin SNP list[b]	SNPs 1240k SNP list[c]	Sequencing technique
AF001	TAW	2.7	0.1	90.1	U5a2b4	nd	XY	47,537	87,591	Shotgun
AF002	TAW	1	0.04	21.4	U5b1	I	XY	43,642	102,253	Shotgun
AF003	TAW	2	0.13	24.5	K1a4a1	-	XX	63,437	116,523	Shotgun
AF015	ABS	55	1.57	215.3	K1a4a1	-	XX	381,308	710,979	Shotgun
AF025	TAW	2	0.06	33.8	T1a1n	R	XY	31,727	58,159	Shotgun
AF029	GMF	12	0.36	44.5	U5a2a1	R1b1a1a2a1a1c1a	XY	138,120	257,643	Shotgun
AF004	TAW	-	-	11.6	H4a1a1a	I	XY	11,938	28,558	SNP capture
AF005	TAW	-	-	1.2	-	nd	XY	2,348	5,697	SNP capture
AF006	GMF	-	-	91.2	H3k1	I	XY	21,173	49,719	SNP capture
AF007	GMF	-	-	66.5	K1a4a1	R1b1a1a2a1a2c1a5b1a1a	XY	35,470	83,698	SNP capture
AF008	GMF	-	-	4.8	U5b2b2	I	XY	10,371	24,360	SNP capture
AF009	GMF	-	-	24.3	H	I	XY	10,927	26,562	SNP capture
AF010	ABS	-	-	612.5	U5a2b4	I	XY	264,422	620,487	SNP capture
AF011	ABS	-	-	335.1	J1c3g	R1b1a1a2a1a	XY	320,303	750,584	SNP capture
AF012	ABS	-	-	137	H1q	I	XY	296,280	694,377	SNP capture
AF013	GMF	-	-	40.3	H4a1a1a	I	XY	22,169	52,761	SNP capture
AF014	GMF	-	-	40.7	J1c5	I	XY	151,584	362,482	SNP capture
AF016	ABS	-	-	16.8	H4a1a1a	nd	XY	10,183	24,279	SNP capture
AF017	TAW	-	-	15.5	H1	R1b1a1a2	XY	20,975	49,318	SNP capture
AF018	TAW	-	-	8.5	U5b2b2	J	XY	7,989	18,866	SNP capture
AF019*	TAW	-	-	-	-		-	-	-	SNP capture
AF020	TAW	-	-	22.5	T1a1n	-	XX	16,123	37,736	SNP capture
AF021*	TAW	-	-	-	-		-	-	-	SNP capture
AF022	TAW	-	-	14.8	H	nd	XY	6,959	16,476	SNP capture
AF023	TAW	-	-	16	H1e2	R1b1a1a2	XY	21,173	49,855	SNP capture
AF024	TAW	-	-	7.1	H1c	-	XX	8,781	21,065	SNP capture
AF026	GMF	-	-	19.5	K1a3a	I	XY	20,461	48,526	SNP capture
AF027*	GMF	-	-	-	-		-	-	-	SNP capture
AF028	GMF	-	-	11.3	K1a4a1	-	XX	15,209	36,053	SNP capture
AF030	GMF	-	-	33.5	H4a1a1a1	nd	XY	3,548	8,592	SNP capture
AF031	ABS	-	-	149.8	H3an	I2a1a1a	XY	77,665	185,350	SNP capture
AF032	ABS	-	-	58	H1q	I2a1a1a	XY	140,821	335,549	SNP capture
AF033	ABS	-	-	76.3	U5b2b2	I	XY	87,647	208,742	SNP capture
AF034	ABS	-	-	559.8	H7d	I	XY	179,586	421,681	SNP capture
AF035*	ABS	-	-	-	-		-	-	-	SNP capture
AF036	ABS	-	-	276.7	J1c5	I2a1a1a	XY	226,524	537,659	SNP capture
AF037	ABS	-	-	205.3	U5b2b3	I	XY	54,947	129,120	SNP capture
AF038*	ABS	-	-	-	-		-	-	-	SNP capture

[a]TAW: Trou Al'Wesse; ABS: Abri Sandron; GMF: Grotte du Mont Falise

[b] Lazaridis *et al.* 2014

[c] Reich's Lab.

3.1.1 Identity by descent

The IBD test was used to determine possible degrees of relationship among the individuals and to detect the presence of duplicated samples.

Since the number of SNPs used in the analysis varied drastically according to the SNP dataset used (Table 15), the IBD test was performed twice, one for each SNP dataset (see section 2.1.5.10). When the input file was generated using the Affimetrix Human Origin SNP list (Lazaridis *et al.* 2014) with a total of 600k SNPs, two individuals, AF009 from Grotte du Mont Falise and AF022 from Trou Al'Wesse, appeared as duplicates. Instead, when the input file was generated using the 1240k SNP list (Reich Lab), none of the individuals appeared as duplicates or relatives.

3.1.2 Principal Component Analysis

Despite the low coverage on the nuclear genome, the 33 individuals had enough known SNPs shared with modern and ancient published populations to perform a Principal Component Analysis (PCA).

Firstly, the Belgian individuals were merged with a subset of the Human Origins dataset (Lazaridis *et al.* 2014), which included modern populations from Europe, Caucasus and the Near East (for the list of populations used in the analysis see Appendix 15). From the PCA plot (Figure 20), it is possible to observe the modern genetic diversity of European populations with Basques and Sardinians being very diverse from the rest of Europe and a cline from North to South East Europe going from the left to the right side of the plot. Modern populations from South East Europe are among the closest, in Europe, geographically and genetically, to Near Eastern and Caucasian populations.

The majority of the ancient Belgian individuals cluster outside the modern European variation with the exception of a small group clustering with modern North European populations (Figure 20).

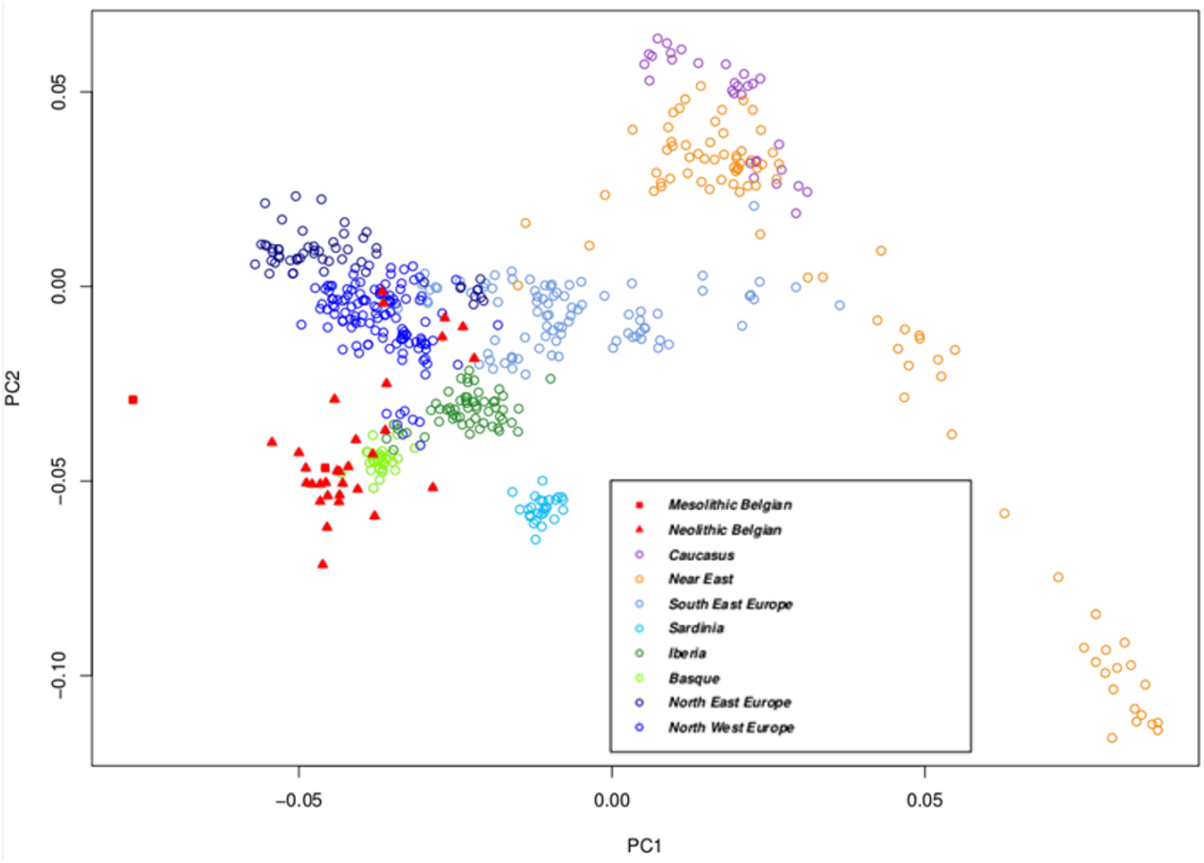


Figure 20. Principal component analysis (PCA) combining the ancient Belgian individuals and a subset of the Human Origins dataset (Lazaridis et al. 2014). This subset includes modern human populations from Europe, Caucasus and the Near East. The plot was created using Rstudio, with the package "plot". See legend for colour code and symbols.

To further investigate genetic affinity among populations, the ancient Belgian individuals were merged with published ancient genomes from Europe, covering a time frame spanning from the Mesolithic to the Bronze Age (Figure 21) (Olalde *et al.* 2014; Skoglund *et al.* 2014; Allentoft *et al.* 2015; Jones *et al.* 2015; Mathieson *et al.* 2015; Cassidy *et al.* 2016; Fu *et al.* 2016; Hofmanova *et al.* 2016; Kilinc *et al.* 2016; Gonzales-Fortes *et al.* 2017; Martiniano *et al.* 2017; Pruefer *et al.* 2017; Saag *et al.* 2017; Mathieson *et al.* 2018; Mittnik *et al.* 2018; Olalde *et al.* 2018; Valdiosera *et al.* 2018).

The ancient Belgian individuals appear to split into three separated clusters (Figure 21). A small group of individuals, renamed now as Neolithic-B, is shifted towards the upper part of the PCA, clustering with European Bronze Age individuals and Bell Beakers. The second group called Neolithic-A, includes the majority of the Neolithic samples, and the Mesolithic AF003, and it occupies an intermediate position between WHGs and Neolithic Europeans. The Neolithic-A Belgians plot separately from the rest of published ancient individuals so far.

The third group is represented by the Mesolithic individual AF002, who clusters with other published WHG, on the left side of the plot.

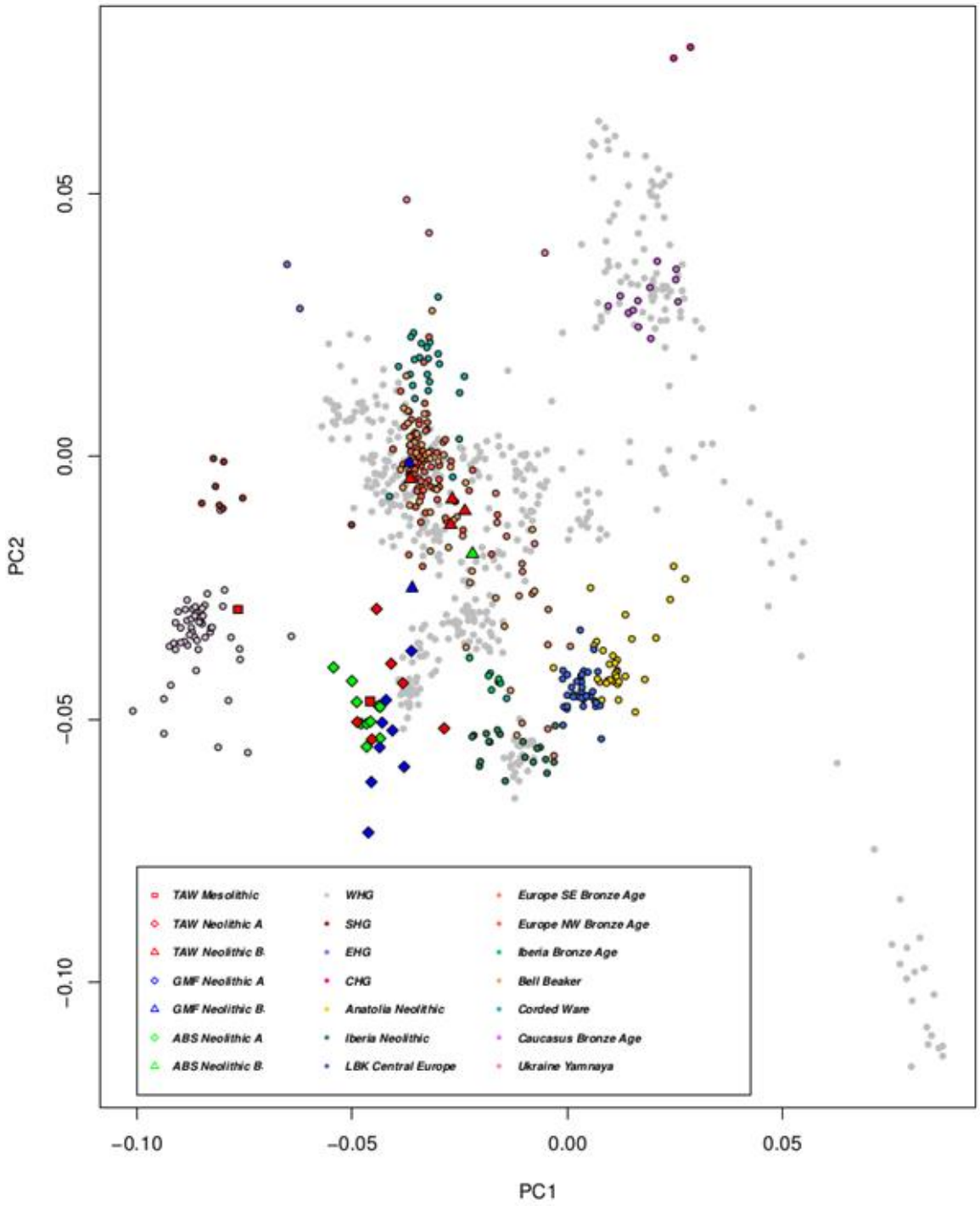


Figure 21. Principal component analysis (PCA) plot combining the ancient Belgian individuals, a subset of the Human Origins dataset (Lazaridis et al. 2014) and published ancient Europeans. The grey dots in the background represent modern European populations. The plot was created using Rstudio, with the package "plot". See legend for colour code and symbols. TAW: Trou Al'Wesse; GMF: Grotte du Mont Falise; ABS: Abri Sandron.

A PCA plot including exclusively the 33 ancient Belgian individuals was made to investigate the genetic diversity within the ancient Belgian population (Figure 22). The two Neolithic

groups, A and B, described previously, include individuals from all the three sites. The position of samples with a number of SNPs below 10,000 could be unprecise, those individuals (five) are marked in red font in Figure 22.

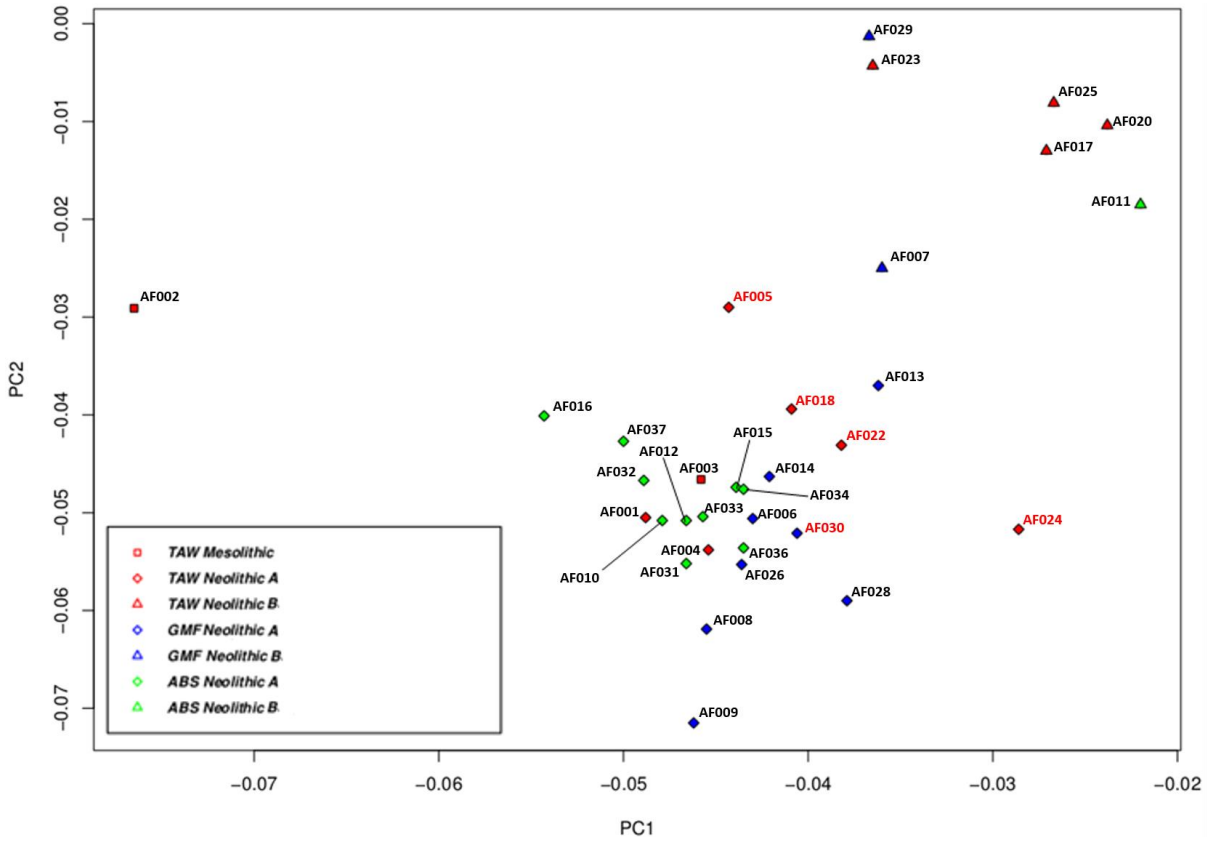


Figure 22. Principal component analysis (PCA) plot of the ancient Belgian individuals with relative IDs. Samples in red font have less than 10,000 SNPs. The plot was created using Rstudio, with the package "plot". See legend for colour code and symbols. TAW: Trou Al'Wesse; GMF: Grotte du Mont Falise; ABS: Abri Sandron.

3.1.3 ADMIXTURE analysis

ADMIXTURE analysis allows a more accurate investigation of the ancestry composition of specific individuals or populations. In Figure 23, is represented the ADMIXTURE plot of Belgian individuals, published ancient and modern individuals. Each bar represents a single individual. From the plot it is possible to identify three main genetic components: a yellow component characteristic of WHG from Europe, a blue component associated with the Neolithic farmers from Anatolia and a red component originary from the Pontic-Caspian populations.

Hunter-gatherer populations from Europe are composed almost entirely of yellow component, except for EHG and CHG characterised by an increasing percentage of CHG component (red). The highest fraction of Neolithic component is observed among Anatolian

farmers (Anatolia_Neolithic), the first farmers. Bronze Age populations (Bulgaria_EBA, Iberia_BA, England_MBA, Ireland_BA, Netherlands_BA, Sweden_BA, Czech_EBA, Latvia_BA, Germany_CW, Ukraine_Yamnaya), instead, are represented by an increasing CHG component, with the lowest values in Early Bronze Age individuals from Europe and the highest values in the Yamnaya from Ukraine. Modern Europeans, here represented by Czech, British, French, Basque, Italian, Sicilian and Sardinian populations, display all the three major genetic components, with southern Europeans typically characterised by a higher Neolithic ancestry and northern Europeans with a higher hunter-gatherer ancestry (Figure 23).

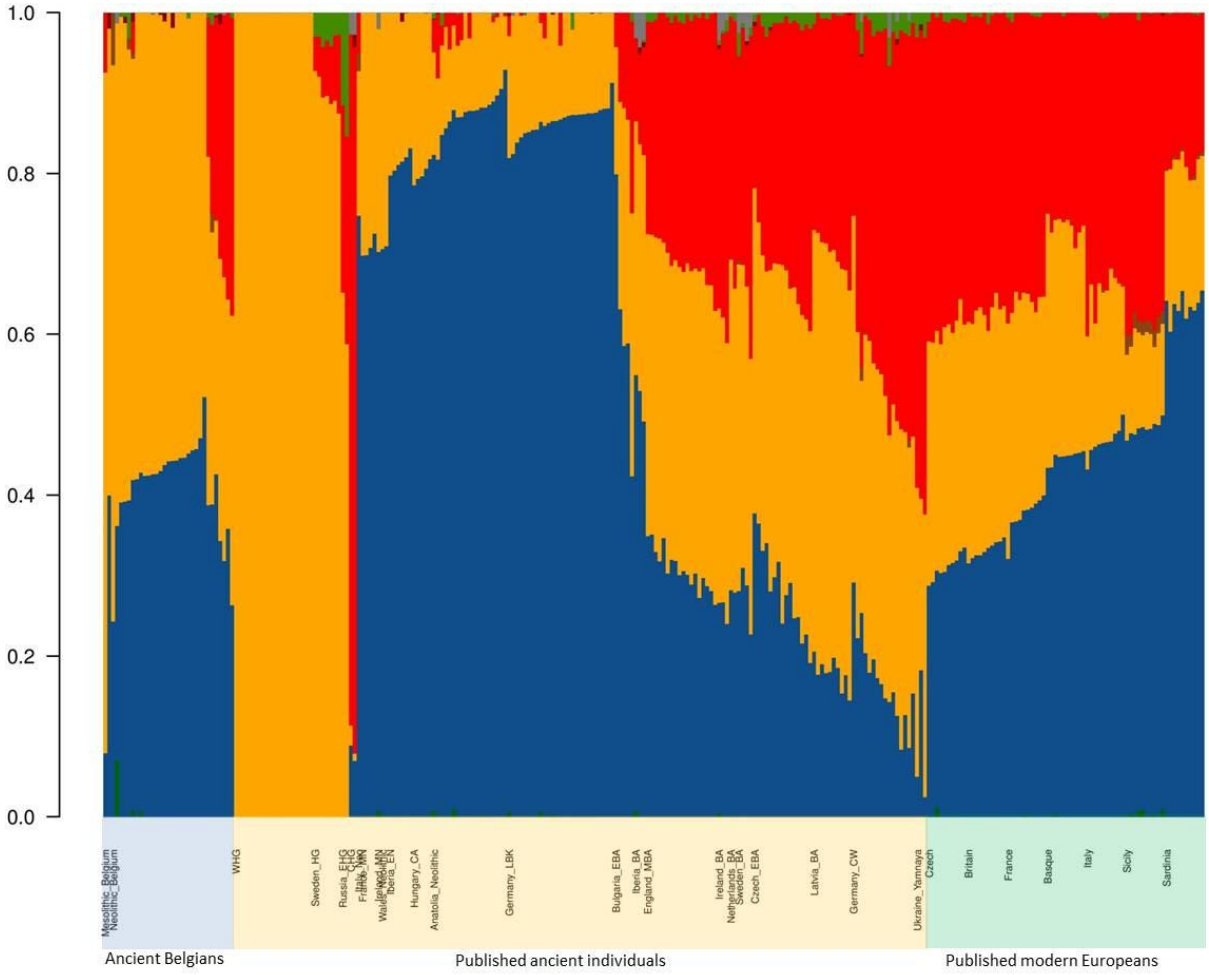


Figure 23. ADMIXTURE plot (k = 8). Each bar corresponds to an individual whose population group is listed below. Major genetic components: Western hunter-gatherers (yellow), Neolithic Levant (blue), Steppe (red). Populations from left to right: Mesolithic Belgium, Neolithic Belgium, WHG, SHG, EHG, CHG, Neolithic Europeans, Bronze Age Europeans and modern Europeans.

When the ancient Belgians are analysed in detail (Figure 24), the ADMIXTURE analysis shows three different groups within the population. The Mesolithic individual AF002 displays the

highest percentage of hunter-gatherer component among the ancient Belgians. The Neolithic-A cluster have two major components: the hunter-gatherer component around 60% and the Neolithic farmer component around 40%. The seven individuals belonging to the Neolithic-B group, display instead an additional component, represented by the CHG component ranging from 20 to 35%. Compared to other Neolithic Europeans shown in Figure 23, the Neolithic-A individuals generally show a higher percentage of hunter-gatherer component, already suggested by their intermediate position in the PCA (Figure 24).

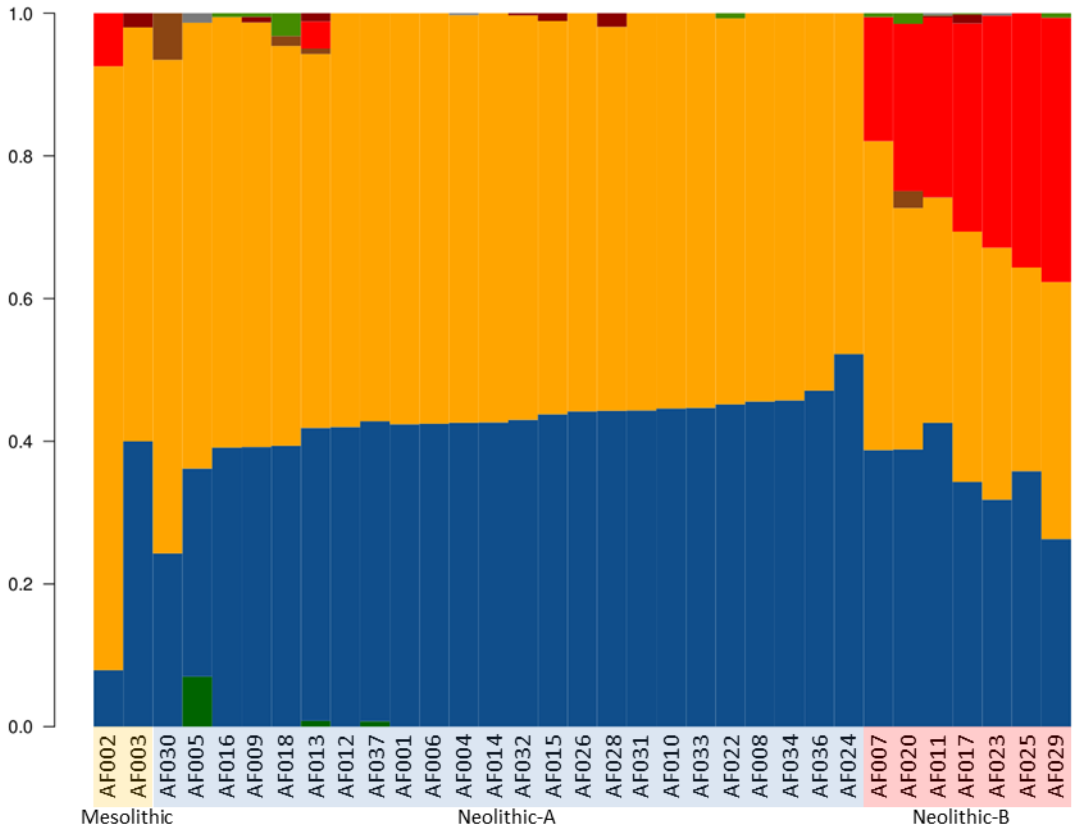


Figure 24. Subset of ADMIXTURE analysis plot, with k=8, showing the genetic composition of the thirty-two ancient Belgian individuals. Each bar corresponds to an individual whos ID is listed below. Major genetic components: Western hunter-gatherers (yellow), Neolithic Levant (blue), Steppe (red).

3.1.4 *f*-statistics

To statistically confirm the the presence of two distinct Belgian clusters (Neolithic-A and Neolithic-B) an outgroup-*f*3 test was performed for each individual in the format of *f*3(Outgroup, test individual AF, target individual AF). As outgroup, the African population Yoruba was used, as this population is genetically distant from the rest of the European populations, both ancient and modern, used in the tests. Here are reported the results for four target individuals: AF001, AF003, AF011 and AF023 (Figure 25). The outgroup-f3 results

for Neolithic-A individuals AF001 and AF003 show a clear distinction between the values scored with other Neolithic-A Belgians and values scored with Neolithic-B individuals, confirming the separation of the two groups. Neolithic-B Belgians AF011 and AF023, instead, appear to be genetically closer to both A and B individuals, suggesting admixture between the two groups. This result could support the hypothesis of genetic continuity from Neolithic-A individuals to Neolithic-B during the course of time.

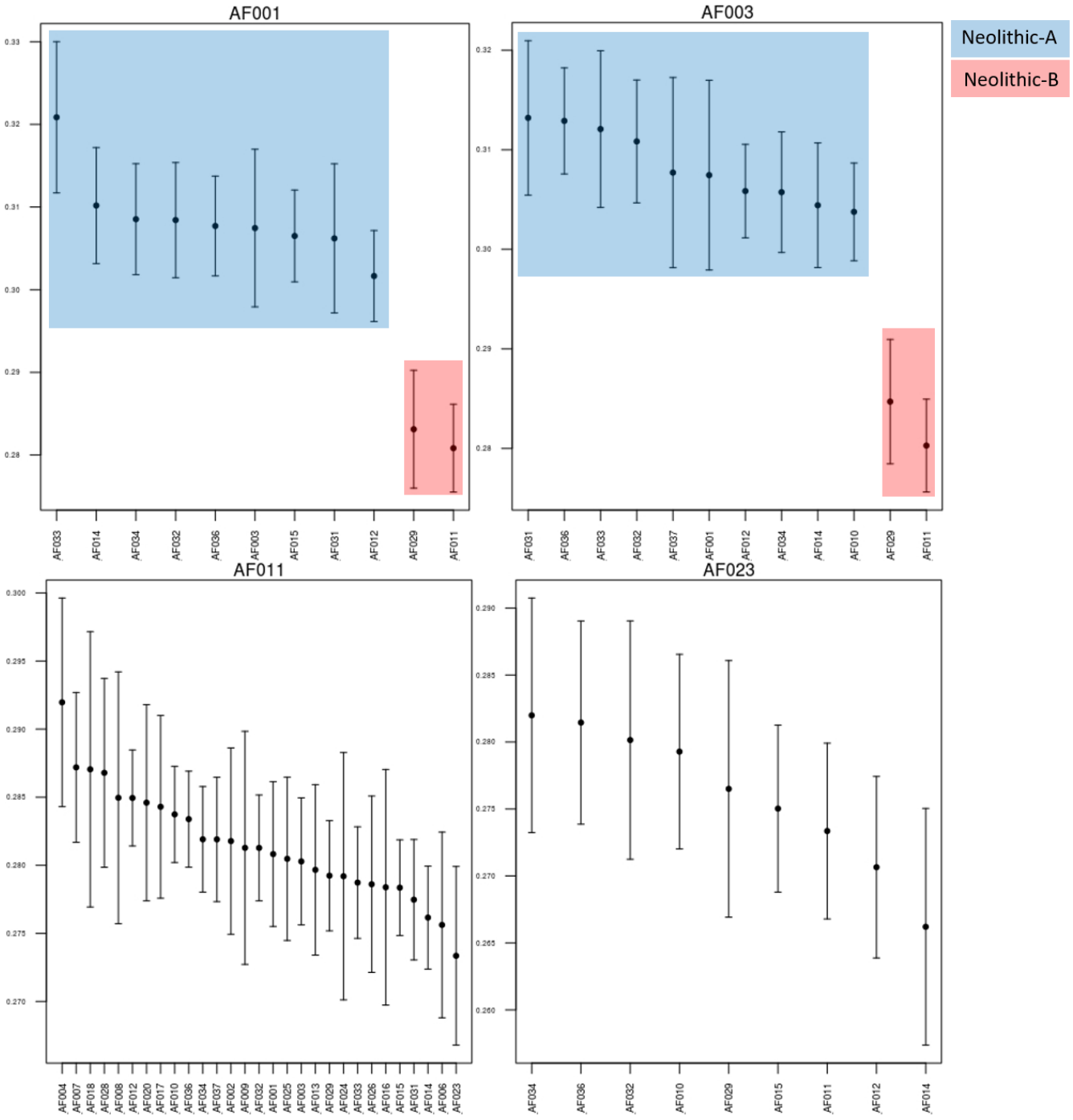


Figure 25. Outgroup-f3 statistics, in the format f3(outgroup, test individual AF, target individual AF) at individual level of two Neolithic-A Belgians (AF001 and AF003) and two Neolithic-B Belgians (AF011 and AF023). Higher values indicate more shared ancestry between individuals.

To formally test which individuals could be grouped in the two different clusters, two additional *D*-statistics tests were performed, in the format of *D*(Outgroup, test population, target individual AF, reference individual AF) (Figures 26 and 27). The two Belgian groups, Neolithic-A and Neolithic-B, in this test, were represented by one "reference Belgian individual" with the highest number of SNPs, AF015 (710,979) for Neolithic-A and AF011

(750,584) for Neolithic-B. Each Belgian individual was compared to several published Mesolithic and Neolithic populations as “test population”. Values on the right of the dashed line indicate admixture between test population and reference individual. Values on the left of the dashed line, instead, indicate admixture between target population and target individual. When, instead, the values overlap the dashed line, the target individual and the reference individual could statistically form a separate clade from the test population.

The D-statistics shows that all the Belgian individuals belonging to the Neolithic-A group, when tested with “AF015 Neolithic-A” as reference, share genetic drift with AF015 and together form a separate clade from the test population (Figure 26). The Neolithic-B Belgians, instead, when tested with “AF015 Neolithic-A” as reference, are genetically distant to both Iron Gates and AF015 (Figure 26).

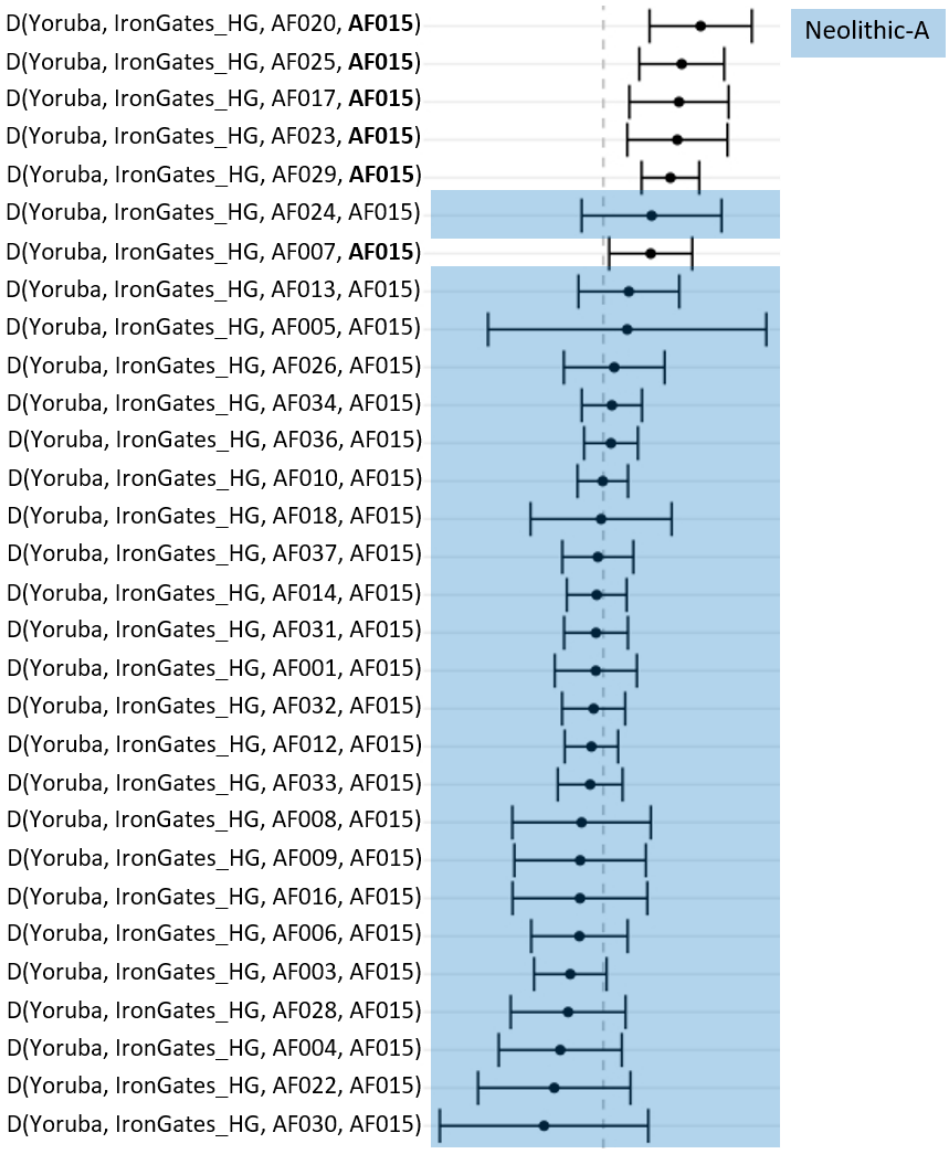


Figure 26. D-Statistics test at individual level. Each row represents a different combination of populations in the format D(outgroup, test population IronGates_HG, target individual AF, reference individual AF015). The values on the right side of the dashed line indicate presence of admixture between the test population and AF015. The values on the left side of the dashed line indicate presence of admixture between the test population and the target individual (AF…). The values that overlap the dashed line indicate that the target individual and the reference individual could statistically form a separate clade from the test population.

When the test was repeated using "AF011 Neolithic-B" as reference, the individuals belonging to the Neolithic-B group form a separate clade from the test population and cluster with AF011 (Figure 27). The Neolithic-A Belgians, instead, form a cluster with the Iron Gates. Only three individuals coulded not be formally grouped: AF007, probably due to its position in the PCA halfway between Neolithic-A and Neolithic-B (Figure 22), and AF005 and AF024 due to the low number of SNPs covered (Table 15).

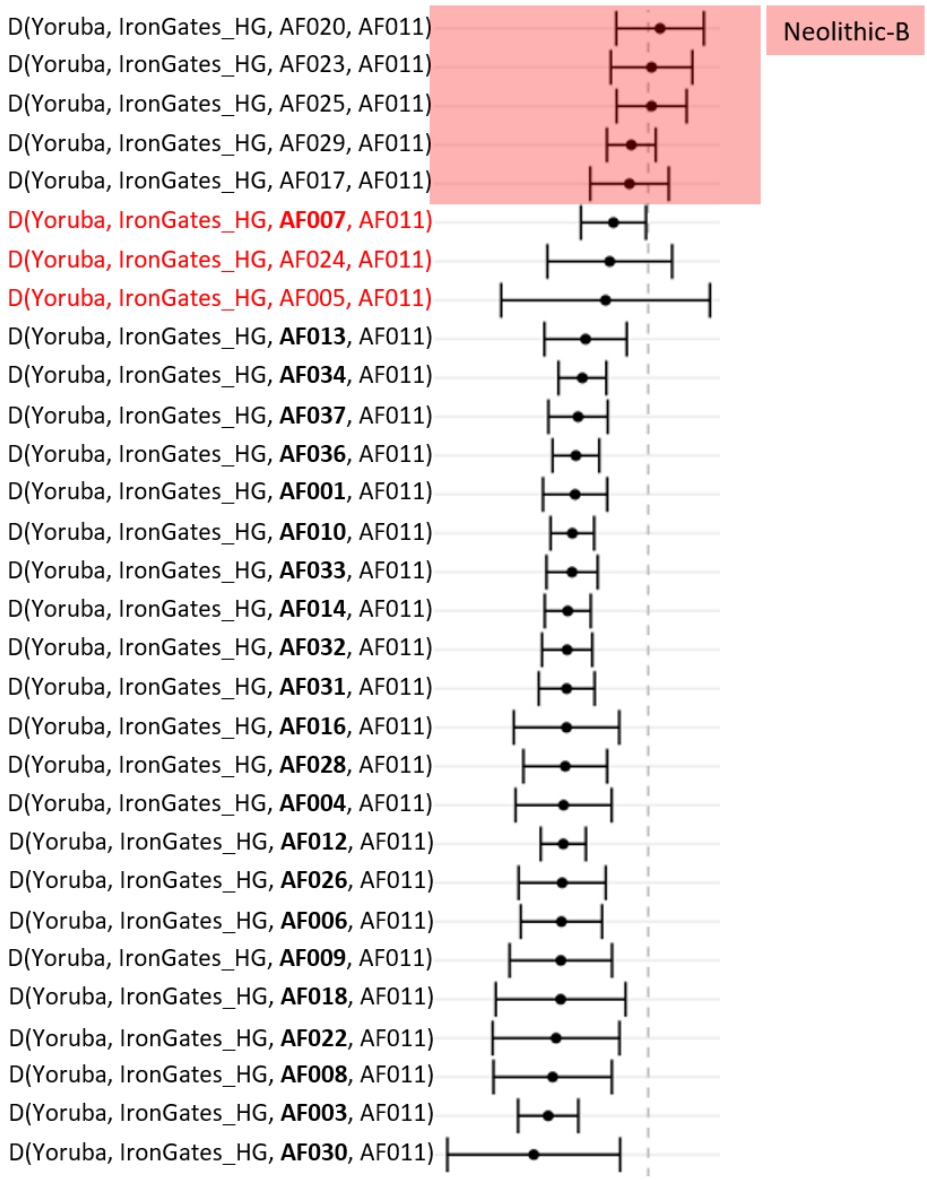


Figure 27. D-Statistics test at individual level. Each row represents a different combination of populations in the format D(outgroup, test population IronGates_HG, target individual AF, reference individual AF011). The values on the right side of the dashed line indicate presence of admixture between the test population and AF011. The values on the left side of the dashed line indicate presence of admixture between the test population and the target individual. The values that overlap the dashed line indicate that the target individual and the reference individual could statistically form a separate clade from the test population.

To further investigate the degree of admixture and to assess the presence of shared ancestry between ancient Belgians and other ancient European populations, a D-statistics test in the form D(outgroup population, target population, population A, population B) was performed (Figures 28 and 29). As outgroup, the African population Yoruba was used.

When testing the Mesolithic individual AF002 (Figure 28), the highest values of admixture are achieved with other European WHGs, such as Loschbour from Luxembourg and Bichon

from Switzerland. Despite being geographically close, the Belgian Upper Palaeolithic individual Goyet does not seem to be genetically close to AF002 (Figure 28). When compared to several Neolithic populations from Europe, AF002 shows highest values with the Belgians Neolithic-A, suggesting presence of genetic continuity in the area. AF002 seem genetically closer to the Neolithic-A group than the Neolithic-B.

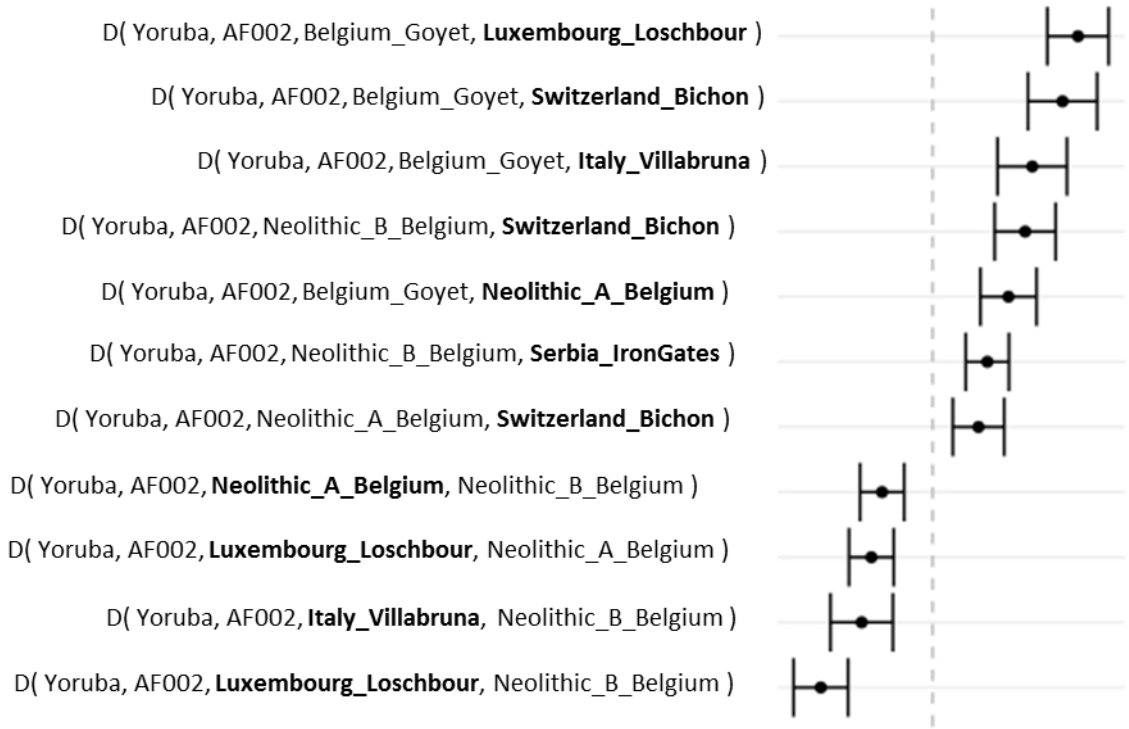


Figure 28. D-Statistics test of AF002 vs European HGs and Belgium Neolithics (A and B). Each row represents a different combination of populations in the format (outgroup population, target population X, population A, population B). The values on the right side of the dashed line indicate presence of admixture between the test population X and B. The values on the left side of the dashed line indicate presence of admixture between X and A.

When testing Neolithic-A Belgian group (Figure 29), the *D*-statistics results indicate that the strongest signal of shared drift is with AF002, due to the shared ancestry, and followed by WHGs. This high level of shared drift with Mesolithic populations confirms the high proportion of WHG ancestry of Neolithic-A Belgians showed in the ADMIXTURE. When the Neolithic-A Belgians are compared to other Neolithic populations, including the Neolithic-B Belgian group, there is a stronger signal with Neolithic populations from France and Britain, than the Neolithic-B Belgian group (Figure 29).

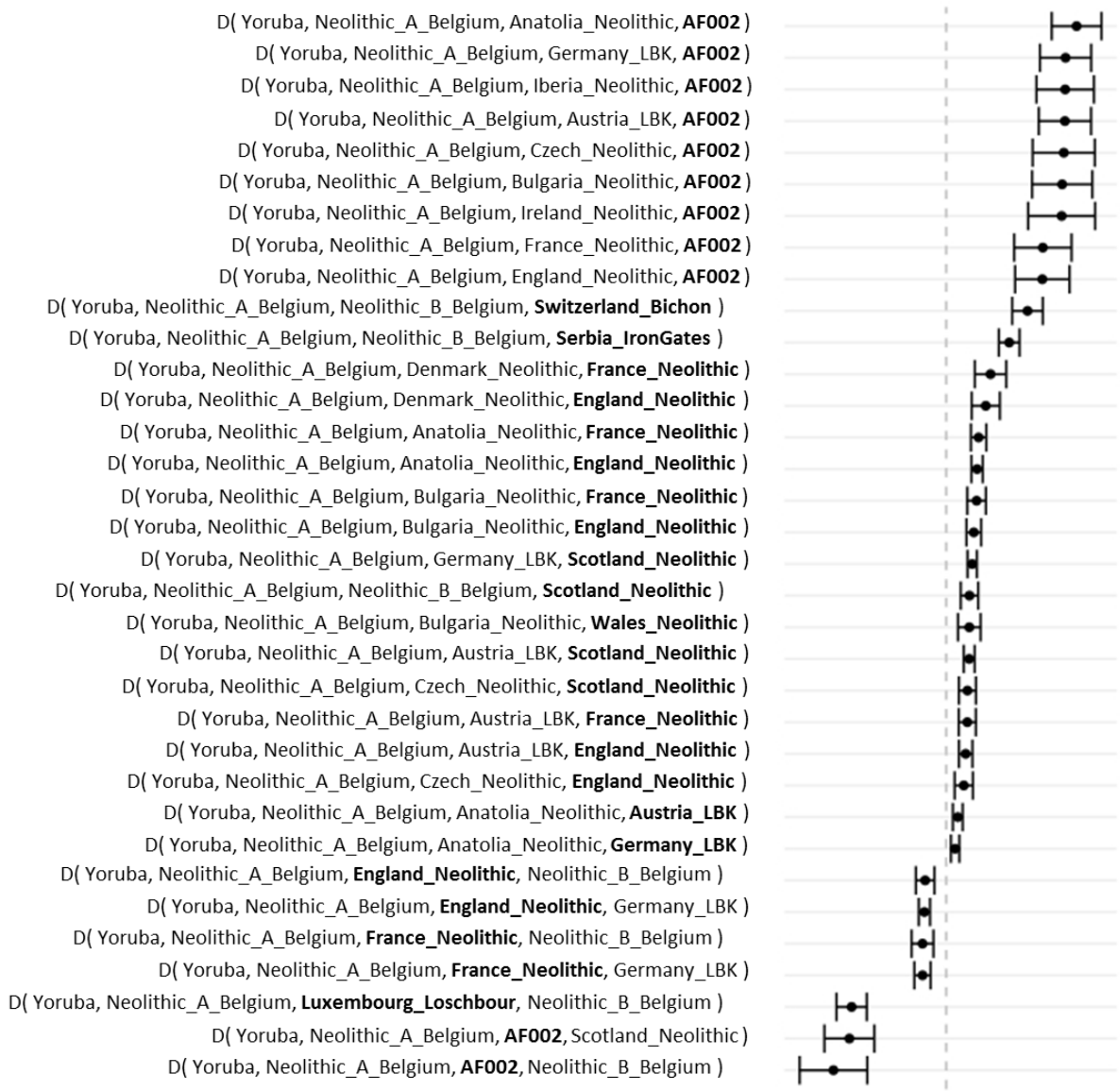


Figure 29. D-Statistics test of Belgian Neolithic-A vs AF002 and other Neolithic and HGs populations from Europe and the Levant. Each row represents a different combination of populations in the format (outgroup, X, pop A, pop B). The values on the right side of the dashed line indicate presence of admixture between the test population X and the pop B. The values on the left side of the dashed line indicate presence of admixture between the test population X and the pop A.

To formally test which published Mesolithic and Neolithic populations had the highest amount of shared drift with the Neolithic-A group, an outgroup-*f*3 in the form of *f*3(Outgroup, test, Neolithic-A) was performed (Figure 30). The results confirmed the D-statistics results, with the Neolithic-A group having the highest shared drift with AF002 and other Mesolithic populations, followed by Middle Neolithic populations from England, France and Iberia. Lower values were achieved with Early Neolithic LBK, Anatolian farmers, Neolithic-B Belgians and Yamnaya population.

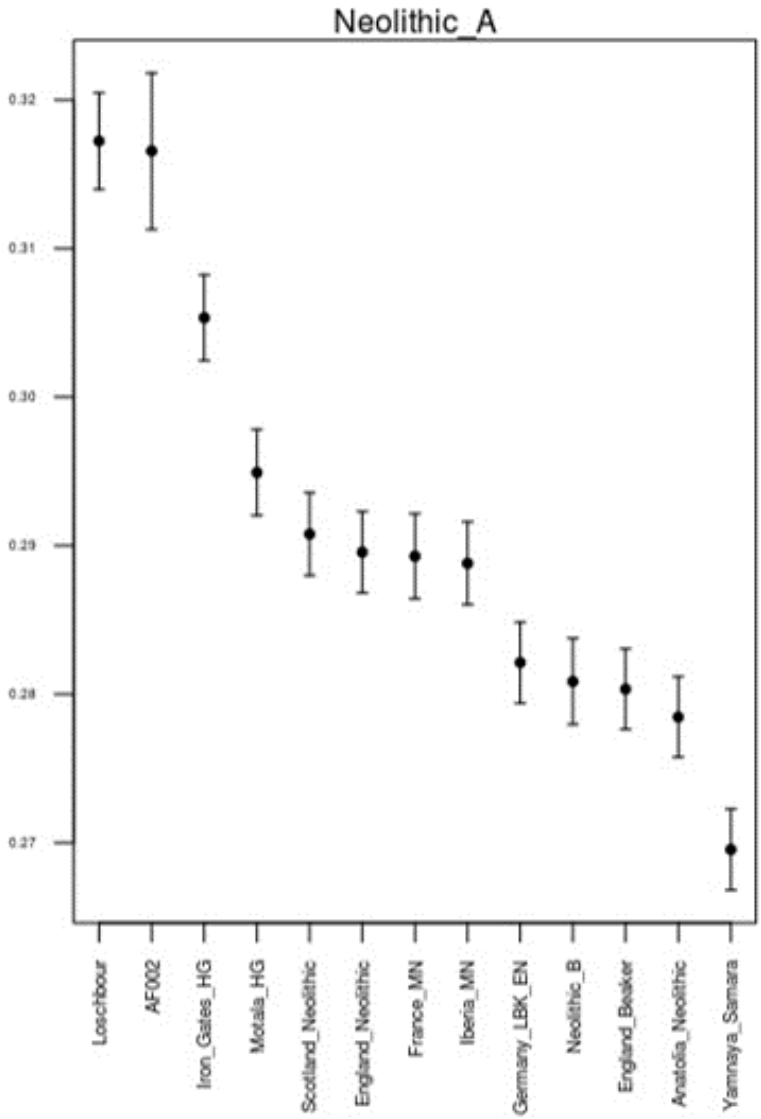


Figure 30. Outgroup-f3 statistics, in the format f3(outgroup, target population, Neolithic-A) at group level of Neolithic-A Belgians. Several published Mesolithic, Neolithic and Bronze Age populations were used as target. Higher values indicate more shared ancestry between groups.

To assess presence of shared ancestry from various hunter-gatherers, Neolithic and Bronze Age populations with the Neolithic Belgians (A and B), a test in the form of D(outgroup, target population, Neolithic-A Belgium, Neolithic-B Belgium) was performed (Figure 31). The result of the D-test showed admixture between Neolithic-B Belgium with the Yamnaya Samara population. Populations who lack CHG component, such as Early LBKs from Germany, Neolithic populations from Western Europe and hunter-gatherers, instead, were genetically closer to the Neolithic-A group (Figure 31). When the Neolithic-B Belgians, instead, were compared with Early Bronze Age populations, such as Corded Ware, and Bell Beakers, they formed a distinct cluster.

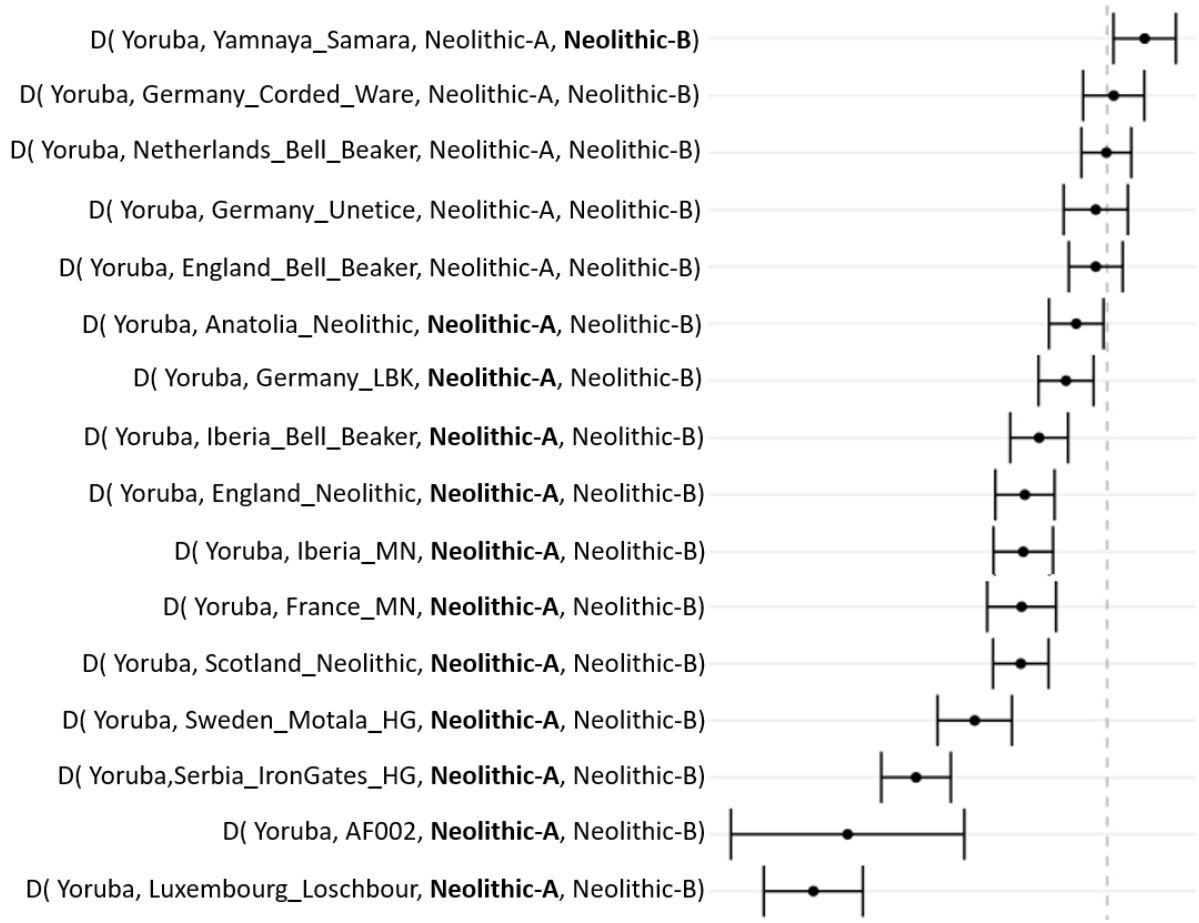


Figure 31. D-Statistics test of various hunter-gatherers, Neolithic and Bronze Age populations vs Belgian Neolithics (A and B). Each row represents a different combination of populations in the format (outgroup, X, pop A, pop B). The values on the right side of the dashed line indicate presence of admixture between the test population X and the pop B. The values on the left side of the dashed line indicate presence of admixture between the test population X and the pop A. Values that overlap the dashed line indicate that the Neolithic A and B groups form a separate cluster from the target X.

3.1.5 Uniparental markers

Mitochondrial genomes were retrieved for 32 individuals out of 33 and assigned to known haplogroups (Table 15, Figure 32, Appendix 16). The mitogenome of AF005 was highly fragmented and so it was impossible to safely assign a mitochondrial haplogroup.

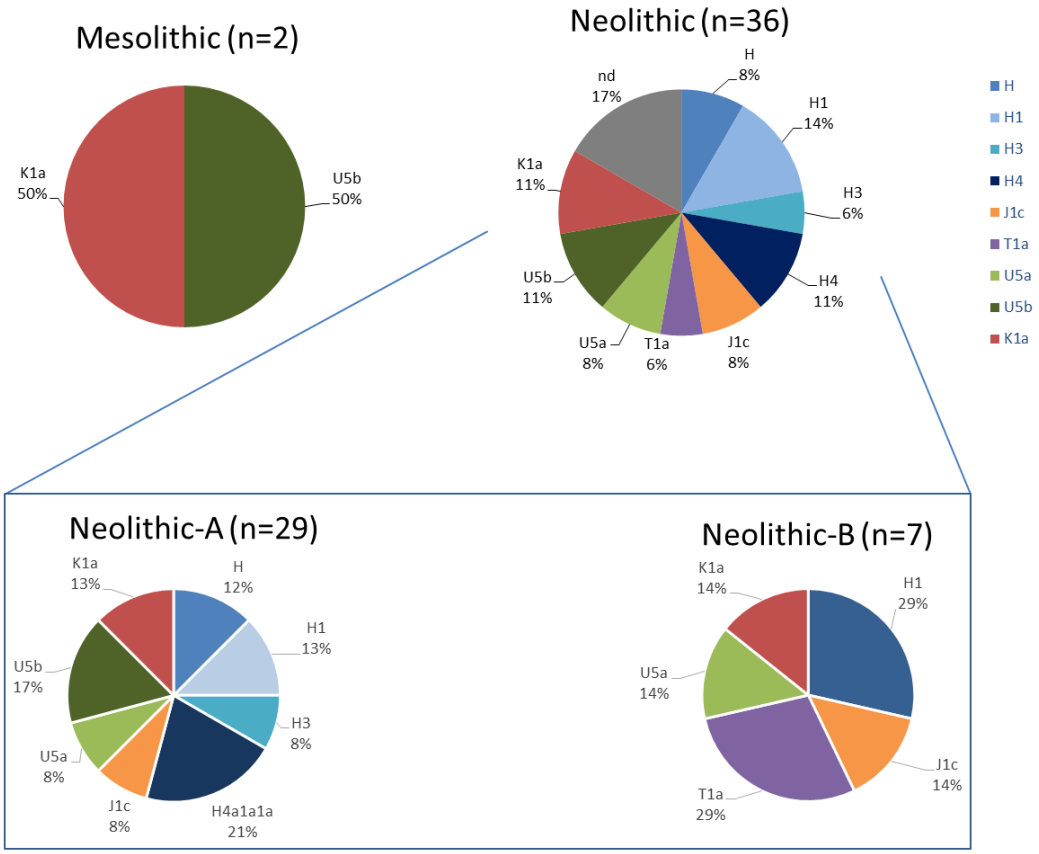


Figure 32. Mitochondrial haplogroup distribution of the 38 ancient Belgian individuals, separated in Mesolithic (n =2) and Neolithic (n = 36). The Neolithic individuals were further separated in Neolithic-A (n=29) and Neolithic-B (n=7). See legend for colour codes.

The two individuals classified as Mesolithic, AF002 and AF003, belong respectively to mitochondrial haplogroups U5b1 and K1a4. The oldest individual that has been found belonging to haplogroup U5b1 is an Upper Palaeolithic individual from France which shares the same haplotype with AF002 (Figure 33). This agrees with the whole genome results for the sample, showing a strong hunter-gatherer component. U5b1 clade additionally harbours two hunter-gatherers from France and Germany, Neolithic individuals from North Europe and Chalcolithic individuals from Spain and Britain. AF003, instead, has the more typically Neolithic haplogroup K1a4; further confirmed by the whole genome results, showing that AF003 clusters with other Neolithic Belgian individuals (Figure 22).

All the Neolithic Belgian individuals carry mitochondrial haplogroups typical of the central and western European Neolithic (*e.g.,* K1a, H1, H3, H4, H7, J1c), as well as U5, typical of European hunter-gatherers (Table 15, Figure 32). Two Neolithic Belgian individuals are placed directly at the root of haplogroup H among other Neolithic, Chalcolithic and Bronze Age individuals from North West and South East Europe (Figure 33). The H4a1a1a clade where four Neolithic Belgian clusters, three of which share the same haplotype, includes Chalcolithic and Bronze Age individuals from the Czech Republic, UK and Germany. Individual AF034 is separated by one transition in the mtDNA control region, at np 152, from a Neolithic individual from the Czech Republic inside the clade H7d. AF006 and AF031 are the only ancient individuals belonging, respectively, to H3k1 and H3an. One of the major branches within H is represented by H1. Five different Neolithic Belgians cluster within this haplogroup. Two of them, AF012 and AF032, belong to H1q and are separated from a Chalcolithic individual from Spain by one transition at np 16188. AF023, instead, shares the same haplotype of a Chalcolithic individual from Portugal. H1c is represented by AF024 along with Neolithic, Chalcolithic and Bronze Age individuals from Northern Europe. Two individuals from Trou Al'Wesse carry mitochondrial haplogroup T1a1n. Individual AF011 belongs to mitochondrial haplogroup J1c3g. This specific haplogroup has been found exclusively in Chalcolithic and Bronze Age individuals (Figure 33). Two individuals from Abri Sandron and Grotte du Mont Falise carry haplogroup J1c5, an entirely Neolithic haplogroup within J, with individuals from Northern Europe (Figure 33). Four Neolithic Belgians cluster inside U5b2, which seems to have a Neolithic North Western European distribution. Two individuals from Trou Al'Wesse and Abri Sandron belonging to haplogroup U5a2b4, share the same haplotype (Figure 33); their cluster harbours from a Neolithic individual from Germany. AF029 is an individual from Grotte du Mont Falise who carries haplogroup U5a2a1. This clade is represented by Chalcolithic, Bronze Age, Anglo Saxon and Iron Age individuals. Five individuals cluster within haplogroup K1a, which is thought to have spread during the Neolithic from the Fertile Crescent (Torroni *et al.* 1996; Richards *et al.* 2000). A Neolithic individual from Turkey shares the same haplotype with AF026 and several other Neolithic and Chalcolithic Northern Europeans (Figure 33).

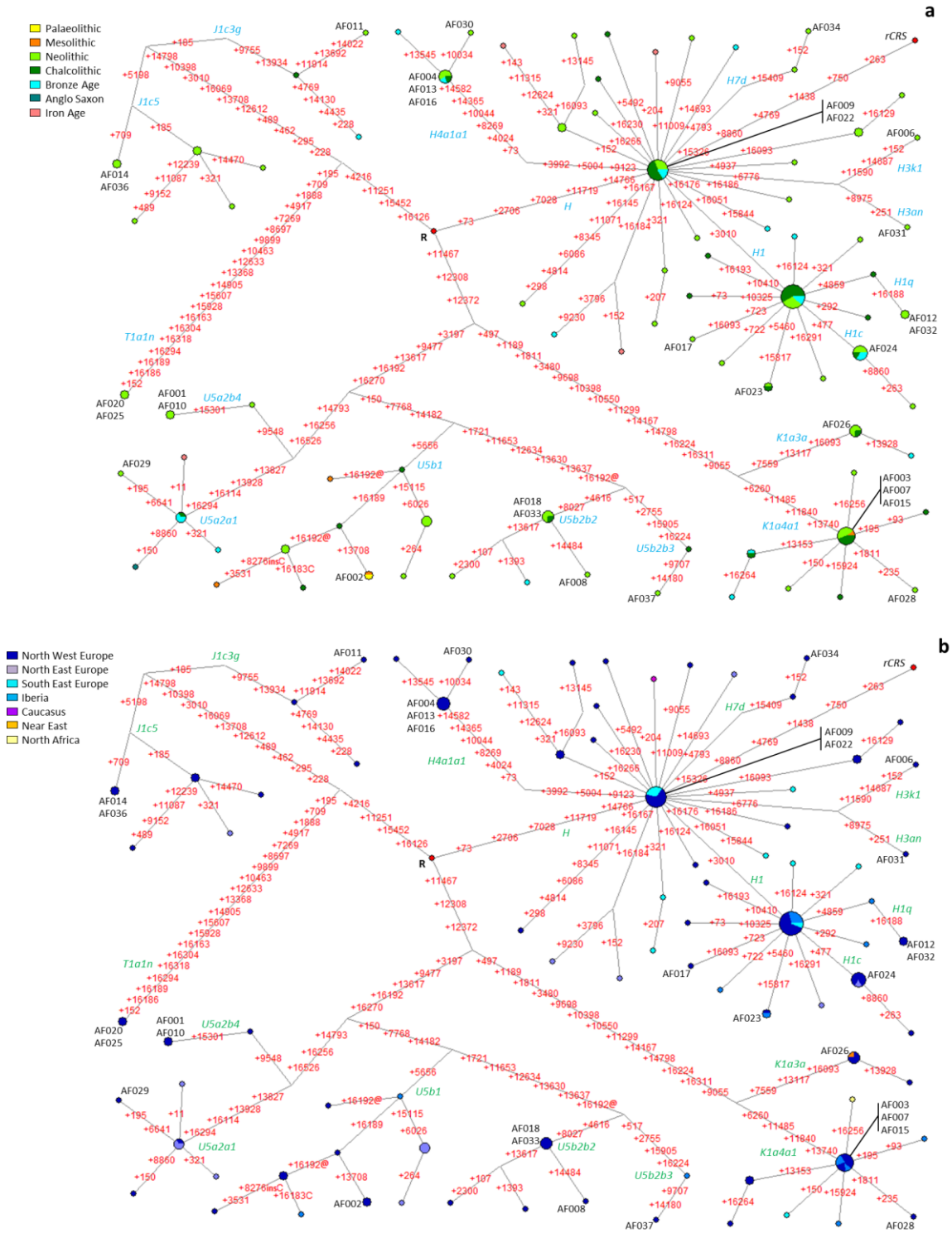


Figure 33. Network analysis of complete ancient mitochondrial genomes organised according to the time period of the samples (a) and geographic location (b). The dataset is composed of the ancient Belgians and other published ancient individuals belonging to the same mitochondrial haplogroup clade. The mutations are highlighted in red font. See legend for colour code.

Several mitochondrial haplogroups are shared across different sites (Figure 34). This is the case for U5a2b4 present in one individual from Trou Al'Wesse and in one individual from

Abri Sandron. Another U5 sub-lineage, U5b2b2, is shared across all three sites, in one individual for each site, with AF008 (Grotte du Mont Falise) carrying a private mutation at np 14484. Haplogroup J1c5 is present in one individual from Grotte du Mont Falise and in one individual from Abri Sandron. Haplogroup K1a4a1 is shared across all sites with one individual from Grotte du Mont Falise carrying a private mutation at np 235. At last, haplogroup H4a1a1a is present in five individuals across all three sites, with one individual from Grotte du Mont Falise further classified as H4a1a1a1.

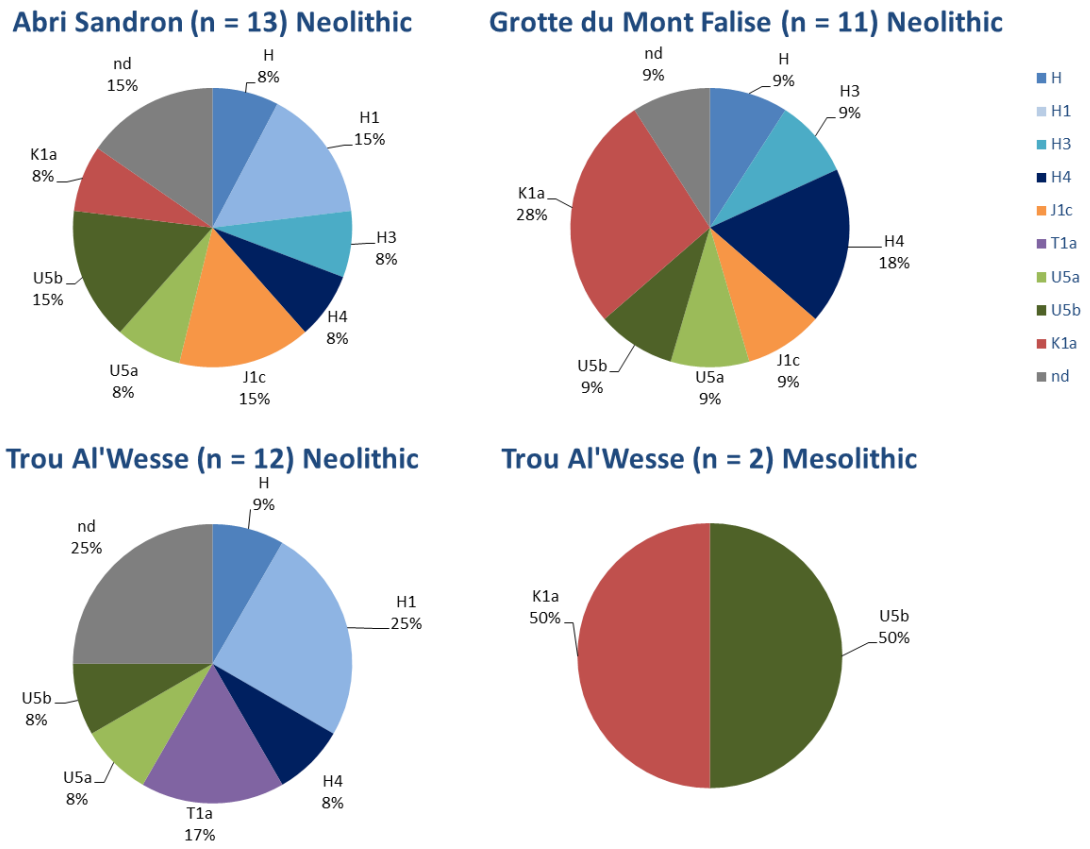


Figure 34. Mitochondrial haplogroup distribution of the 38 ancient Belgian individuals, separated according to the archaeological site of origin. See legend for colour codes.

Due to the limited coverage on the Y chromosome, it was not possible to precisely classify all the male lineages (Appendix 18). Sixteen individuals across the three sites belong to haplogroup I (I2a, when the resolution is sufficient) (Figure 35). One individual, from Trou Al'Wesse, belongs to the less common Y chromosome haplogroup J. Six individuals, instead, belong to haplogroup R, five of which can be further classified as R1b (Figures 35 and 36).

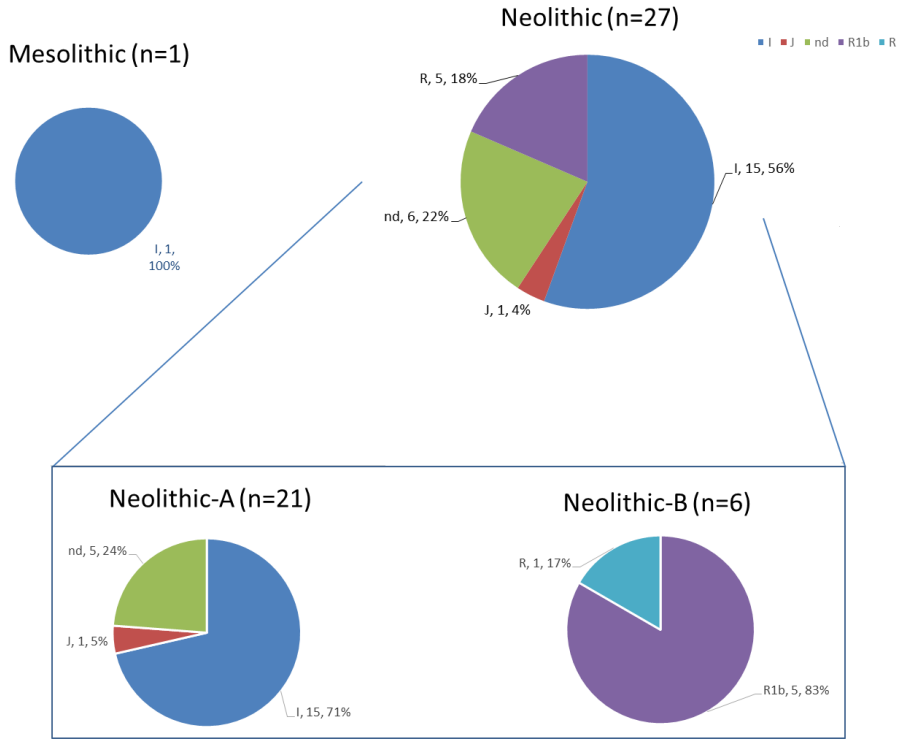


Figure 35. Y chromosome haplogroup distribution of the 28 ancient Belgian males, separated in Mesolithic (n=1) and Neolithic (n=27). The Neolithic males were further separated in Neolithic-A (n=21) and Neolithic-B (n=6). See legend for colour codes.

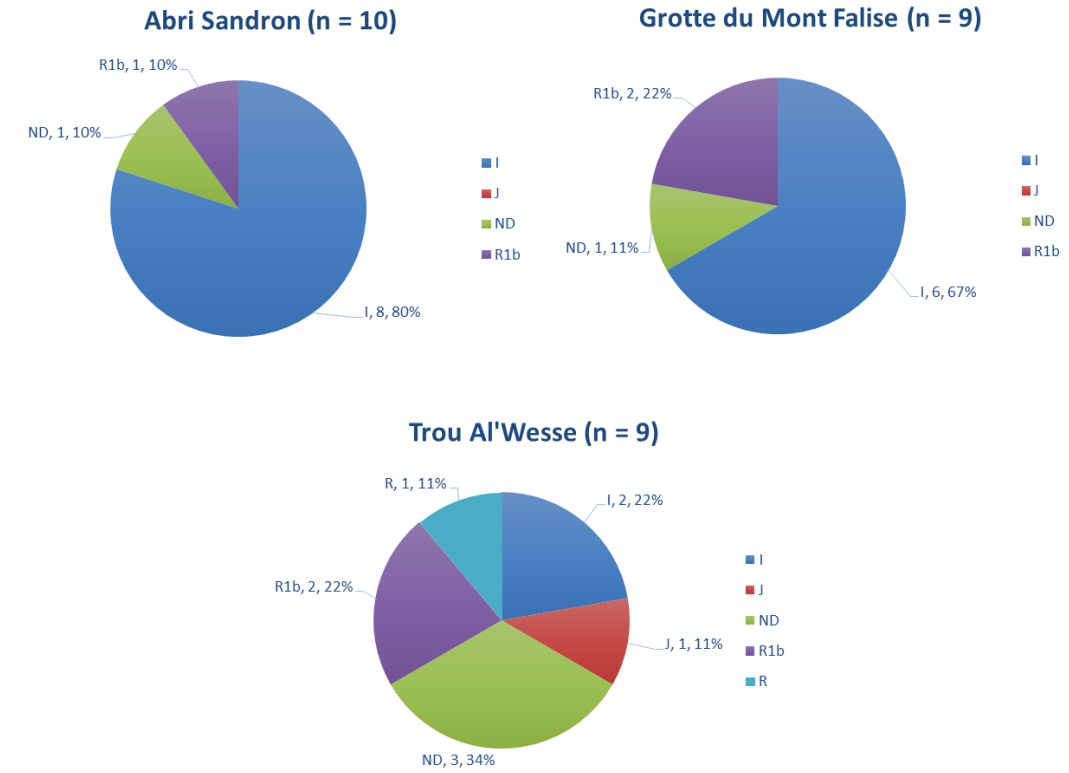


Figure 36. Y chromosome haplogroup distribution of the 28 ancient Belgian males, separated according to the archaeological site of origin. See legend for colour codes.

3.1.6 Dietary isotopes

The analysis of the ratio of dietary isotopes $\delta^{13}C$ and $\delta^{15}N$ can give a view of the dietary habits of animals, including humans, in a specific area or timeframe. Figure 37 shows a plot combining $\delta^{13}C$ values ranging from -24 to -19 and $\delta^{15}N$ values ranging from 4 to 14 (Figure 37). The ratio of carbon isotopes $^{12}C/^{13}C$ ($\delta^{13}C$) reflects the kind of plants consumed. Herbivores tend to show low values of $\delta^{13}C$ due to intake of wild C_3 plants, while higher values of $\delta^{13}C$ reflects intake of marine plants, C_4 plants or cultivated cereals (Figure 37). The ratio of nitrogen isotopes $^{14}N/^{15}N$ ($\delta^{15}N$), instead, reflects the source of animal protein in the diet and trophic level position. Herbivores show lower values of $\delta^{15}N$ compared to omnivores and carnivores (Figure 37).

The isotopic data of 21 ancient Belgian individuals, Mesolithic and Neolithic, from the three sites was merged with data from different species of animals published in Bocherens *et al.* 2007 to have a reference view of the food sources present in Belgium during the Mesolithic and Neolithic (Figure 37). The Mesolithic individuals show the highest levels of $\delta^{15}N$, reflecting a diet more abundant in meat compared to the Neolithic samples. Belgian Neolithic individuals across all three sites show comparable values of $\delta^{13}C$ and $\delta^{15}N$, suggesting same dietary habits.

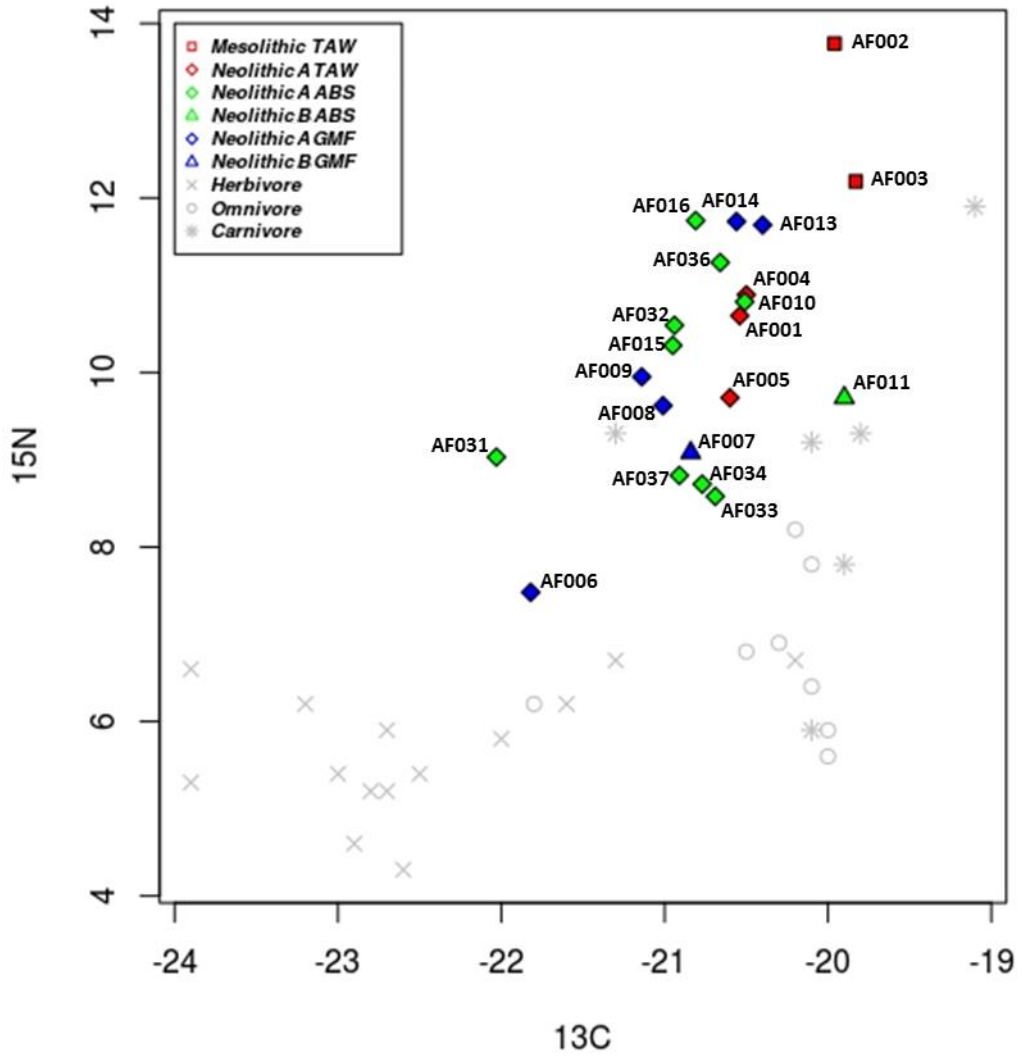


Figure 37. Isotopes analysis plot used for dietary reconstruction of the ancient Belgian individuals. On the x axis the ratio of carbon isotopes 12C/13C (δ13C) reflecting the intake of C3 or C4 plants. On the y axis, instead, the ratio of nitrogen isotopes 14N/15N (δ15N) reflecting the source of animal protein in the diet. See legend for colour codes.

3.1.7 Phenotype prediction

Due to the low coverage of the samples, the phenotype prediction for skin pigmentation, eye and hair colour, performed using the HiriSplexS platform (Walsh *et al.* 2014; Walsh *et al.* 2017; Chaitanya *et al.* 2018), was successful for six individuals only: AF010, AF011, AF012, AF014, AF015 and AF034 (Table 16). The six individuals had the majority of the 41 SNPs necessary for the prediction covered at least by one read, apart from AF014 where, due to the high number of missing SNPs, it was not possible to predict the colour of the hair.

All individuals have high probability (~90%) of having blue eyes, with the exception of AF015 who more likely (52%) carried brown eyes (Table 16). Two individuals, AF012 and AF015, both Neolithic-A individuals from Abri Sandron, have high probability (~70%) of having

brown hair; AF034 and AF011, respectively Neolithic-A and Neolithic-B individuals from Abri Sandron, more likely carried blond hair; AF010, instead, has similar values for both blond and brown hair, 51% and 44% respectively. There is absence of red or black hair phenotypes (Table 16). Four individuals from Abri Sandron, belonging to both Neolithic groups, have high probability of having light hair; one Neolithic-A from Abri Sandron and one Neolithic-A from Grotte du Mont Falise, instead, more likely carried darker hair. Three individuals have high probability of having an intermediate colour of the skin; two individuals have similar values for different categories: AF010 for pale and intermediate skin and AF012 for very pale and intermediate skin. AF015 more likely carried a dark to black skin (Table 16).

Table 16. Phenotypic prediction for eyes, hair and skin colour using HirisPlexS system (Walsh et al. 2014; Walsh et al. 2017; Chaitanya et al. 2018).

	Neolithic-A Abri Sandron								Neolithic-A Grotte du Mont Falise		Neolithic-B Abri Sandron	
	AF010		**AF012**		**AF015**		**AF034**		**AF014**		**AF011**	
	P-value	AUC loss	P-value	AUC loss	P-value	AUC loss	P-value	AUC loss	P-value	AUC loss	P-value	AUC loss
Blue eye	**0.91**	0.03	**0.93**	0.01	0.36	0.00	**0.91**	0.01	**0.90**	0.02	**0.91**	0.01
Intermediate eye	0.06	0.07	0.06	0.04	0.12	0.00	0.06	0.04	0.06	0.06	0.06	0.04
Brown eye	0.03	0.03	0.01	0.01	**0.52**	0.00	0.04	0.01	0.04	0.02	0.04	0.01
Blond hair	**0.51**	0.06	0.08	0.02	0.10	0.05	**0.65**	0.02	/		**0.75**	0.01
Brown hair	**0.44**	0.06	**0.71**	0.02	**0.67**	0.04	0.31	0.01	/		0.16	0.00
Red hair	0.01	0.01	0.09	0.08	0.02	0.01	0.01	0.03	/		0.07	0.01
Black hair	0.04	0.02	0.12	0.01	0.21	0.01	0.03	0.01	/		0.02	0.00
Light hair	**0.92**	0.03	0.21	0.01	**1.00**	0.03	**0.95**	0.01	0.17	0.05	**0.98**	0.00
Dark hair	0.08		**0.79**		0.00		0.05		**0.83**		0.02	
Very pale skin	0.03	0.06	**0.43**	0.02	0.00	0.02	0.03	0.03	0.02	0.08	0.03	0.01
Pale skin	**0.50**	0.03	0.20	0.03	0.00	0.01	0.34	0.01	0.09	0.05	0.21	0.01
Intermediate skin	**0.47**	0.05	**0.37**	0.03	0.34	0.02	**0.61**	0.02	**0.65**	0.10	**0.76**	0.01
Dark skin	0.00	0.03	0.01	0.01	0.00	0.00	0.02	0.02	0.22	0.03	0.00	0.01
Dark to black skin	0.00	0.01	0.00	0.00	**0.66**	0.00	0.00	0.00	0.02	0.01	0.00	0.00

P-value: Probability of prediction
AUC loss: Loss in the accuracy of the prediction

The genotyping of two SNPs located in the *LCTa* (rs4988235) and *LCTb* (rs182549) genes associated with lactase persistence was successful for 14 individuals (Table 17). When the alleles at both SNPs are present in the ancestral state, the lactase enzyme most likely will not be produced during the adulthood causing lactose intolerance in the individual. When one or both alleles are present in the derived state the chances to be able to digest milk in the adulthood increase.

Most of the individuals had only one of the two SNPs covered by at least one read, only AF010, AF015 and AF036 had both positions covered (Table 17). All the Mesolithic and Neolithic-A individuals have one (or both when covered) SNPs in the ancestral state (Table 17), suggesting that more likely they were not able to digest milk in the adulthood. AF029, one of the two Neolithic-B individuals, instead carries a mutation in the *LCTa* gene (rs4988235), which probably allowed him to digest milk (Table 17).

Table 17. Genotypes at SNPs associated with lactase persistence in LCTa gene (rs4988235) and LCTb (rs182549). The number between parentheses represents the number of reads that covers that position.

Period	Sample	rs4988235[a]	rs182549[b]
Mesolithic	**AF002**	-	C (x1)
	AF003	-	C (x1)
Neolithic-A	**AF001**	-	C (x1)
	AF010	G (x1)	C (x1)
	AF012	G (x4)	-
	AF014	-	C (x1)
	AF015	G (x1)	C (x1)
	AF031	G (x2)	-
	AF032	G (x1)	-
	AF033	G (x1)	-
	AF034	G (x1)	-
	AF036	G (x3)	C (x2)
Neolithic-B	**AF011**	-	C (x5)
	AF029	A (x1)	-

[a] *Ancestral allele / Derived allele: G / A*
[b] *Ancestral allele / Derived allele: C / T*

3.2 Modern mitochondrial data

3.2.1 Modern Belgium

Out of the 166 modern Belgian mitogenomes, it was possible to confidently classify into known mitochondrial haplogroups 124 of them. Despite the small sample size, these sequences belong to Belgian individuals from different provinces, covering the entire country (Figure 38).

Figure 38. Geographical distribution of 124 modern Belgian mitogenomes according to the Belgian province of origin.

Around 50% of the sequences belong to haplogroup H (Figure 39), the remaining part of the macro-haplogroup HV is represented exclusively by HV0 (4%). Haplogroups J and K, here represented only by specific sub-lineages, together they achieve a frequency of 5% and 9% respectively; similarly, the frequency of these two haplogroups in France are of 5.1% and 10.5% respectively (Karachanak *et al.* 2011). The frequency of haplogroup T2 is higher than T1, with comprable frequencies found across Western Europe. Haplogroup U is found mainly as lineages U5a and U5b and with a smaller percentage the lineage U2e (Figure 39).

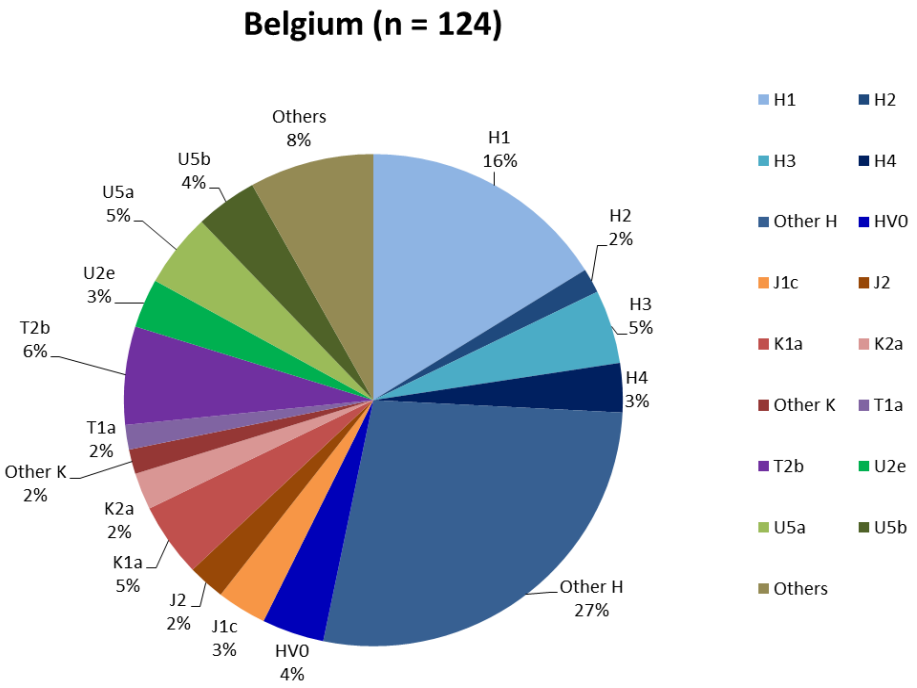


Figure 39. Mitochondrial haplogroup distribution of 124 complete modern Belgian mitogenomes. See legend for colour codes.

3.2.2 Haplogroup HV

The age estimation for mitochondrial haplogroup HV (excluding H) based on a total of 1,101 sequences, of which 388 are unpublished (see 2.2 for further details), was calculated both using ρ statistics and maximum likelihood approach (Table 18).

The coalescence age for haplogroup HV is around 20-23 kya, in line with previous dating assigned to the Upper Palaeolithic (Pereira *et al*. 2005). However, this age might not be completely accurate since haplogroup H, which is the largest haplogroup within HV, is not considered in the calculation.

Haplogroup HV0 (14-15 kya) is entirely European. It divides in two main sub-haplogroups: HV0a, which includes V, and a node defined by a transition in position 195 (HV0+195). Individuals branching from the root are mainly from South East Europe, Italy in particular. A small clade with South Eastern and Eastern European samples hosts an Early Bronze Age individual from Bavaria (Knipper *et al.* 2017) (Appendix 17).

The nodes HV0+195 and HV+16311 are defined by a single hotspot mutation and particularly the case of HV+16311 it groups a big section of the HV phylogeny with a very heterogeneous composition in the geographic origin of the samples (Appendix 17). Most

probably the haplogroups nested within this node have a very different origin and history and they could be split in several parallel clades. However, the tree was built following a maximum parsimony approach,where the topology with the minimum number of mutations is preferred. Therefore, it is not convenient the presence of this recurrent mutation in many parallel clades. HV+16311 is characterised by the presence of six ancient samples at the root of the node from Hungary and Germany spanning from 5 to 3 kya (Haak *et al.* 2015). HV0+195 includes mainly North Western and South Eastern European complete sequences. A Neolithic individual from Ireland (~5 kya) harbours from the root of the haplogroup (Cassidy *et al.* 2016).

Table 18. Age estimation (years) calculated using ρ statistics (pink columns) and Maximum likelihood (blue columns) of the human mitochondrial haplogroup HV based on modern complete published and unpublished sequences. Only the lineages containing at least 10 individuals are listed below.

		ρ			ML		
Node	**Sample size**	**Age (years)**	**Lower bound (years)**	**Higher bound (years)**	**Age (years)**	**Lower bound (years)**	**Higher bound (years)**
HV	1101	20,458	14,597	26,480	23,086	19,365	26,869
>HV0	857	15,069	8,872	21,469	13,866	9,186	18,664
>>HV0+195	82	9,792	7,616	11,996	10,957	7,735	14,239
>>>HV0c	11	4,975	2,055	7,951	5,350	1,699	9,086
>>>HV0f	10	4,951	1,997	7,962	5,489	372	10,776
>>>HV0e	16	6,381	3,929	8,872	7,599	3,625	11,669
>>HV0a	646	12,915	7,884	18,084	11,174	7,392	15,037
>>>HV0a1	17	6,942	3,469	10,491	6,119	2,720	9,592
>>>V	560	10,737	8,662	12,837	8,877	7,624	10,140
>>>>V+72	15	4,863	2,859	6,893	6,611	2,125	11,223
>>>>V1	83	10,478	3,724	17,499	8,095	6,296	9,913
>>>>>V1a	80	7,690	3,115	12,392	6,611	4,012	9,251
>>>>>>V1a1	62	5,473	2,821	8,171	4,792	3,030	6,574
>>>>>>>V1a1a	20	5,084	82	10,249	2,582	316	4,884
>>>>>>>>V1a1a1	13	1,985	426	3,560	1,694	100	3,305
>>>>V2	40	8,229	5,068	11,451	8,024	5,677	10,404
>>>>>V2b	13	5,012	2,025	8,057	5,629	2,240	9,091
>>>>V3	28	8,567	4,880	12,336	7,740	5,534	9,976
>>>>V7	59	10,564	3,261	18,178	7,953	5,744	10,191
>>>>>V7a	50	5,057	878	9,350	5,839	2,177	9,586
>>>>>>V7a1	39	1,652	762	2,548	1,557	626	2,495
>>>>V9	20	21,168	8,326	34,805	7,740	2,415	13,240
>>>>>V9a	19	10,754	3,195	18,647	5,419	570	10,421
>>>>V10	16	10,781	5,378	16,353	8,237	6,437	10,056
>>>>>V10b	10	9,790	4,877	14,845	7,528	5,188	9,901

>>>>V18	10	7,622	1,224	14,275	5,559	1,370	9,861
>>>>V25	15	6,638	1,202	12,260	6,119	2,182	10,154
>>>>V29	40	6,549	1,438	11,824	6,752	3,880	9,675
>HV1	125	21,336	14,105	28,811	20,044	15,358	24,833
>>HV1a'b'c	121	18,446	13,521	23,489	18,313	13,667	23,065
>>>HV1a	46	17,279	10,604	24,178	18,313	11,385	25,479
>>>>HV1a1	26	11,667	5,817	17,710	12,188	7,551	16,945
>>>>>HV1a1a	17	6,942	3,442	10,519	9,663	5,366	14,069
>>>>HV1a2	11	10,610	6,631	14,680	13,354	7,435	19,464
>>>HV1b	55	15,506	10,182	20,977	16,746	12,136	21,462
>>>>HV1b2	23	7,190	450	14,214	9,234	643	18,274
>>>>>HV1b2a	21	1,842	907	2,783	1,898	699	3,107
>>>HV1e	11	12,111	5,593	18,867	15,707	6,100	25,804
>HV2	51	40,280	26,756	54,416	22,627	18,178	27,165
>>HV2a	43	38,546	25,065	52,656	21,560	17,279	25,925
>>>HV2a1	34	23,088	15,809	30,606	18,014	13,518	22,609
>>>>HV2a1a	17	19,349	10,885	28,163	14,601	9,060	20,304
>>>>>HV2a1a2	14	14,566	8,377	20,959	13,281	7,364	19,389
>>>>HV2a1b	10	13,370	7,544	19,382	12,261	3,647	21,295
>HV4	85	17,136	10,043	24,485	17,640	11,585	23,878
>>HV4a	66	15,532	8,379	22,952	17,640	9,890	25,692
>>>HV4a1	54	12,292	5,888	18,926	13,720	6,819	20,879
>>>>HV4a1a	42	6,549	3,862	9,280	6,330	3,600	9,106
>>>>>HV4a1a4	11	4,735	576	9,006	4,862	1,356	8,447
>>>HV4a2	12	11,985	5,846	18,336	12,989	4,884	21,459
>HV7	11	15,148	9,121	21,367	18,613	10,542	27,006
>HV12	34	17,003	11,020	23,167	19,440	13,761	25,274
>>HV12a	14	15,167	7,478	23,169	14,233	7,729	20,964
>>HV12b	16	10,953	5,206	16,889	12,916	5,772	20,343
>>>HV12b1	15	8,793	5,733	11,909	10,381	5,929	14,948
>HV13	18	19,649	12,162	27,407	20,800	14,939	26,822
>>HV13a	10	19,013	10,039	28,382	18,163	9,269	27,455
>HV14	23	15,193	6,674	24,098	15,042	7,953	22,398
>>HV14a	21	10,480	5,196	15,925	11,319	6,704	16,054
>>>HV14a1	11	6,670	1,164	12,368	7,882	2,956	12,957
>HV18	19	7,892	5,174	10,654	9,377	5,635	13,202
>HV+16311	173	9,458	7,795	11,136	11,319	9,346	13,313
>>HV6	19	3,965	1,662	6,302	3,476	1,596	5,380
>>HV9	23	9,660	5,737	13,672	9,306	5,702	12,986
>>HV15	10	10,884	5,359	16,585	10,094	6,749	13,503
>>HV16	15	3,983	737	7,299	5,769	1,707	9,936

The Founder Analysis shown in Figure 36 includes all the modern individuals belonging to haplogroup HV excluding H and HV0.

When the Near East is used as source area of migration, and Europe as sink area (Figure 40b), the migration profile, which plots the percentage of incoming lineages against time displays a smaller signal dating to postglacial (~11.5 kya) and a larger peak during the Neolithic (~7.5 kya) (Figure 40a).

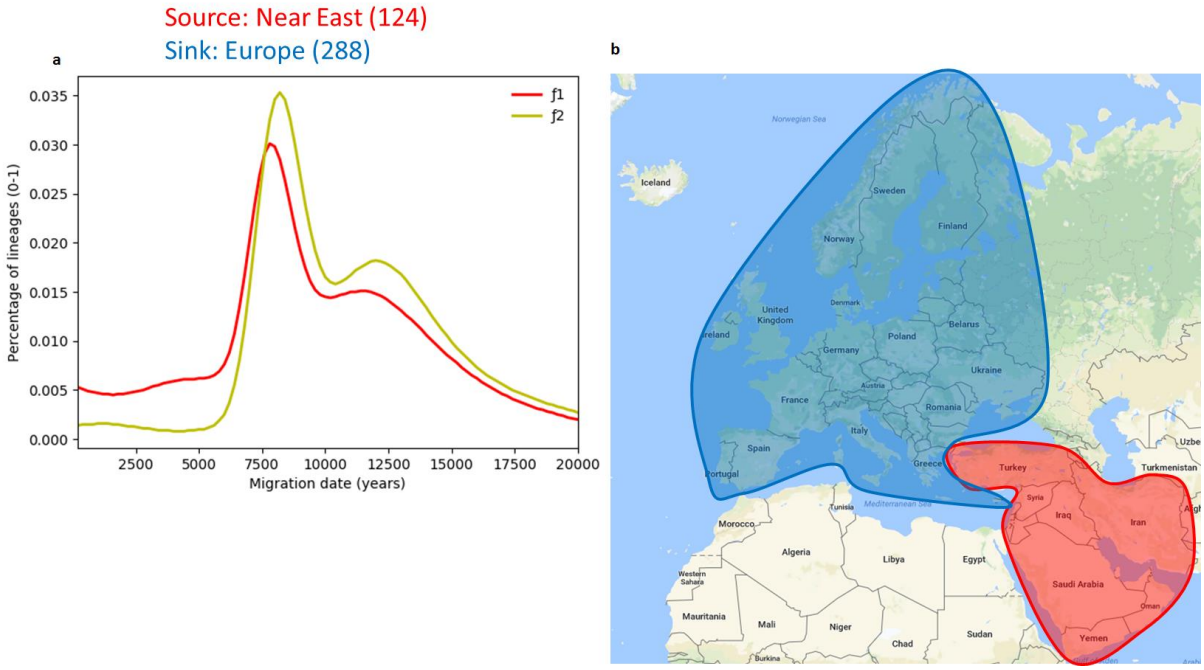


Figure 40. Founder Analysis (FA) result of complete modern mitochondrial DNA sequences belonging to haplogroup HV (excluding H and HV0) using the Near East as source area and Europe as sink area of migration. (a) Graphical representation of the FA result with a timeline spanning from 0 to 20 kya and the percentage of lineages involved in the migrations. (b) Map of the area of the world used as "source" and "sink", respectively in red and blue with the sample size used in the analysis.

To further investigate the role of Southern Europe in the movements that shaped the maternal genetic composition of the continent, a second test was performed, exclusively involving European HV lineages (excluding H) and using Southern Europe as source area and the rest of the continent as sink area (Figure 41b). The results show the presence of a larger peak during the Neolithic (~9 kya) and a smaller signal during the Bronze Age (~4-5 kya) (Figure 41a).

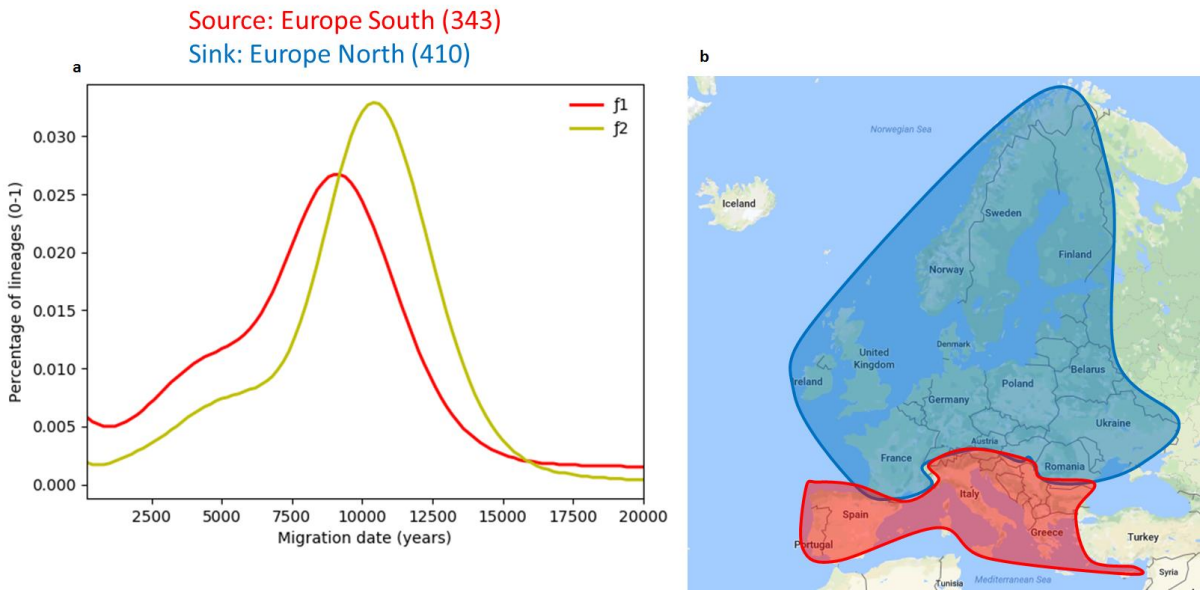


Figure 41. Founder Analysis (FA) result of complete modern mitochondrial DNA sequences belonging to haplogroup HV using Southern Europe as source area and the rest of Europe as sink area of migration. (a) Graphical representation of the FA result with a timeline spanning from 0 to 20 kya and the percentage of lineages involved in the migrations. (b) Map of the area of the world used as "source" and "sink", respectively in red and blue with the sample size used in the analysis.

4.1 Belgium

The ancient Belgians described in this thesis represent the first documented genetic transect, from the Mesolithic to the Neolithic, in Belgium. The dataset included 38 individuals from three archaeological sites in Belgium in close proximity to each other: Trou Al'Wesse (14), Grotte du Mont Falise (11) and Abri Sandron (13). Archaeological data on the three sites suggest human occupation from the Early Mesolithic until the late Middle Neolithic for Trou Al'Wesse and late Middle Neolithic occupation for Grotte du Mont Falise and Abri Sandron (see 2.1.1). After the first screening, six individuals were chosen for deep shotgun sequencing, whilst the remaining 32 were sent to the Department of Genetics, Harvard Medical School, Boston in the laboratory led by Professor David Reich, for SNP capture. Of the 32 samples, 27 were successfully sequenced, giving a total of 33 sequenced samples (six with whole genome shotgun sequencing and 27 with SNP capture). Despite the low coverage sequencing data for all the samples, it was possible to perform uniparental markers analyses, whole genome analysis and phenotype reconstruction. A striking difference was observed between the number of males and females in the dataset, with 28 males (85%) and only five females (15%) out of 33 individuals sequenced (Table 15). This over-representation of male individuals is unusual for Neolithic cave burials.

The presence of shared mitochondrial and Y chromosome haplogroups and high genetic affinity among individuals belonging to different sites suggest movement of people and genetic homogeneity in the area (Figures 33, 34 and 36). However, according to the IBD analysis there was no close degree of relationship (first, second or third degrees) between the individuals.

Despite the lack of direct radiocarbon dating on the samples, the archaeologists identified two Mesolithic individuals from Trou Al'Wesse and 36 Neolithic individuals from three sites in total. However, the genetic analyses performed revealed a more complex structure within the population.

4.1.1 Mesolithic Belgians (ACOF layer)

According to the archaeological context in which they were excavated, two individuals from Trou Al'Wesse, AF002 and AF003, were assigned to the Mesolithic period. The samples were retrieved from the ACOF layer, located at the bottom of the excavated portion of the terrace (Figure 16). Although no direct dates have been produced for this layer so far, the terrace stratigraphy suggests that the deposition of the ACOF layer preceeded the overlapping Early Neolithic layer, 4b-LaH, dated 7,800-6,630 BP (Figure 16 and Table 3) and might have followed the deposition of the Late Mesolithic layer 4b-delta, dated 7,830-7,580 BP. In the latter case, however, the relationship between the two layers is difficult to interpret, as they lie adjacent one to the other, with the 4b-delta closer to the cave opening, and the ACOF layer slightly below, closer to the river.

4.1.1.1 Autosomal evidence

The chronologically intermediate nature of the ACOF samples is reflected by the genetic results, albeit with some very striking differences: when the ADMIXTURE plots are considered (Figures 23 & 24) both AF002 and AF003 display major WHG-related ancestry (yellow) although in fact this seems to characterise some of the Neolithic samples as well (see following section for details). However, this is much higher in AF002, where it reaches >80%, compared to the ~60% observed in AF003. Intriguingly, both samples also display a secondary ancestral component, associated with the Neolithic (blue), but again with very different proportions: ~40% in AF003 (values mirrored in other Neolithic samples included in this study), and a striking <10% in AF002 (Figure 24). In addition to the small Neolithic component, AF002 displays a curious <10% of Pontic-Caspian ancestry (Figure 24). The PCA analysis plots (Figures 20-22) confirm, and further highlight these differences: AF002 clusters unequivocally with the WHG, well away from AF003, who in turn clusters closely among the other Neolithic samples from Belgium (Figure 21). Considering that AF002 has been contextually dated to Mesolithic times and clusters with other WHGs, it would be improbable to find a CHG component so early in time in western Europe. Therefore, we assume that the CHG component, and the Neolithic as well due to the comparable fraction (<10%), are artifacts, probably results of the poor quality of the sample and sequencing. Currently, we attempted to perform hydroxyproline dating on AF002 but due to the high degradation of the sample, the procedure failed. However, we are in the process of

analysing additional libraries of AF002 to verify these hypotheses. The genetic results regarding AF003, instead suggest that probably this individual belonged to a Neolithic population and was wrongly classified as Mesolithic. The wrong archaeological classification could be attributed to the close position of the location of the sample in the ACOF layer to the Early Neolithic layer 4b-LaH above (Figure 16).

When the genetic affinity of AF002 towards other Mesolithic and Neolithic samples is assessed in the D-statistics test, AF002 appears genetically closer to other Mesolithic individuals from Europe than to the Neolithic individuals from Belgium (Figure 28). In particular, AF002 shows stronger genetic affinities with the Loschbour and Bichon samples (Lazaridis *et al.* 2014; Jones *et al.* 2015), two hunter-gatherers from Mesolithic Luxembourg and Late Glacial Switzerland respectively, compared to the Belgian Early Upper Palaeolithic individual from Goyet (Figure 28), thus suggesting discontinuity with this chronologically distant "fellow countryman". This result is not surprising and is in line with the results by Posth *et al.* 2016, where Goyet's mitochondrial DNA was found to belong to haplogroup M, and was originally analysed among other Palaeolithic and Mesolithic Europeans. Considering that M was not found in postglacial hunter-gatherer populations, Posth and colleagues concluded that its lineage must have become extinct during the LGM (Posth *et al.* 2016).

4.1.1.2 Uniparental markers

Loschbour and Bichon, dated respectively to 7.2 kya (Mesolithic) and 13.7 kya (Upper Palaeolithic), belong to the WHG cluster both in the PCA and ADMIXTURE analyses (Lazaridis *et al.* 2014; Jones *et al.* 2015). Both individuals share the same mitochondrial haplogroup U5b, and Y-chromosome haplogroup I with the Belgian Mesolithic individual AF002. AF002's mitochondrial haplotype is a 100% match with an Upper Palaeolithic Late Glacial individual from France, Iboussieres (Posth *et al.* 2016), inside a branch of haplogroup U5b1 (Figure 33). Mitochondrial haplogroup U5b1 probably originated between central and southern Europe around 17.7 kya (Malyarchuk *et al.* 2010). Its modern distribution is characterised by the presence of Mediterranean clusters, early in the phylogeny, and by deeper central European lineages (Malyarchuk *et al.* 2010). U5b1 is the main clade within U5b, the mitochondrial signature of Western European Upper Palaeolithic and WHGs (Posth *et al.* 2016), and it was the dominant mitochondrial haplogroup in Europe until the arrival of the early farmers during the Neolithic (Bramanti *et al.* 2009; Brandt *et al.* 2014). On the paternal perspective,

AF002 carried Y-chromosome haplogroup I, which represents the signature of Upper Palaeolithic and Mesolithic Europeans (Lazaridis *et al*. 2014; Fu *et al*. 2016).

AF003, instead, carries mitochondrial haplogroup K1a4a1; modern distribution of this haplogroup suggest most likely an European origin dated to ~11 kya (Costa *et al.* 2013). K1a4 is a typical European Neolithic mitochondrial lineage, which probably originated in the Late Glacial of the Near East (Brandt *et al.* 2013; Costa *et al.* 2013). Mitochondrial haplogroup K1 (mainly K1c but including some K1a) has been found in low frequency in Mesolithic hunter-gatherers from southeast Europe (Gonzales-Fortes *et al.* 2015; Hofmanova *et al.* 2015; Mathieson *et al.* 2018). The frequency of this haplogroup rose during the Neolithic with the arrival of early farmers from Anatolia; among its sub-clades, the most frequent during the Neolithic was K1a (Fernandez *et al.* 2014; Mathieson *et al.* 2015; Kilinc *et al.* 2016; Lazaridis *et al.* 2016).

4.1.1.3 Stable isotopes

Stable isotopes analyses seem to support a real hunter-gatherer diet for AF002 and a less clear diet for AF003, which is shifted towards the Neolithic group (Figure 37). However, among the 21 ancient individuals analysed, AF002 and AF003 show the highest values in $\delta^{15}N$, 13.77 and 12.19 respectively, and the lowest in $\delta^{13}C$, -19.96 and -19.83 respectively (Figure 37), suggesting a meat-based diet, probably riverine. Similar values of $\delta^{15}N$ and $\delta^{13}C$, in a range from 11.2 to 12.6 ($\delta^{15}N$) and from -21.5 to -20.1 ($\delta^{13}C$), have been documented among Late Mesolithic individuals from France and Luxembourg (including Loschbour) (Drucker *et al.* 2016). The dietary isotopes pattern seems to confirm the interpretation of the relationship between the two ACOF samples, with AF002 characterised by a meat-based diet, typical of hunter-gatherers, and AF003, displaying a more Neolithic diet based on cultivated cereals and less meat intake.

4.1.1.4 Phenotype reconstruction

Despite being partial and characterised by very low coverage of reads, the result of the phenotype reconstruction for the lactase persistence based on one SNP (rs182549) in the *LCTb* gene suggests that AF002 and AF003 carried the ancestral state of this polymorphism and most likely were not able to digest milk in the adulthood (Table 17). The result is in line with the scenario depicted by European Mesolithic hunter-gatherers carrying the ancestral alleles of rs182549 in the *LCTb* gene and rs4988235 in the *LCTa* gene (Burger *et al.* 2007;

Allentoft *et al.* 2015; Mathieson *et al.* 2015). The most common alternative alleles in modern Europeans, which allow individuals to more likely digest milk during the adulthood, started to increase in frequency in Europe with the Late Neolithic and Bronze Age (Allentoft *et al.* 2015; Cassidy *et al.* 2016).

4.1.2 Neolithic-A Belgians

The Neolithic-A Belgian cluster groups 24 individuals from Trou Al'Wesse (6), Grotte du Mont Falise (8) and Abri Sandron (10). All the Neolithic-A individuals from Trou Al'Wesse were found in the chimney burials, across the four layers (Figure 15 and Table 4). The use of this burial seems to be dated later than the use of the terrace. The sites of Grotte du Mont Falise and Abri Sandron displayed signs of human occupation dated to the Late Neolithic (Abri Sandron: 4765-4619 cal BP, 4769-4607 cal BP, 4966-4815 cal BP; Grotte du Mont Falise: 4768-4611 cal BP, 4892-4807 cal BP) (Toussaint 2002; Toussaint 2003).

4.1.2.1 Autosomal evidence

In the PCA plot they occupy an intermediate position between WHG and other European Neolithic populations, a position not occupied by any ancient population published until now (Figure 21), and as such they seem to fill a genetic gap in the current ancient DNA landscape. The shift towards the WHG cline suggests a stronger Mesolithic ancestry than other Neolithic groups, which is also supported by the high WHG-related component displayed in the ADMIXTURE analysis (around 60%). This value is much higher than those observed in any other European Neolithic population published so far, such as LBK individuals, whose values range from 10 to 20% (Figure 23). After the onset of the Neolithic revolution in Europe, the two populations, European hunter-gatherers and Neolithic farmers, started admixing and the Anatolian early farmer genetic component spread across European Neolithic populations in various degrees (Gamba *et al.* 2014; Haak *et al.* 2015; Gunther *et al.* 2015; Lazaridis *et al.* 2016). Considering this high level of WHG ancestry in the Belgian Neolithic-A group, we decided to further investigate, using D-statistics, if this population was genetically closer to WHGs or other European and Levant Neolithic populations. The D-statistic shows that an excess of shared ancestry is presentbetween the WHGs Loschbour (Luxembourg), Bichon (Switzerland), the Iron Gates Mesolithic (Serbia) and AF002 with the Neolithic-A Belgian group, with the highest scores achieved by AF002,

suggesting a genetic continuity in the area (Figure 29). When compared to other Neolithic populations from Europe, instead, they showed genetic affinity with Neolithic British individuals, followed by Middle Neolithic populations from Central Europe (Figures 29 and 30). The Neolithic individuals from Britain, as reported by Brace and collaborators (2019), are characterised by a high genetic affinity with Iberian Neolithics and the origin of their Neolithic component has been traced to Aegean farmers who followed the Mediterranean route of dispersal (Brace *et al.* 2019). Instead, the Neolithic-A group more likely were the result of an early admixture between Neolithic and Mesolithic, where the WHG component was still more prevalent (Figures 29 and 30). However, to support the hypothesis of this early admixture, more radiocarbon dates for the Neolithic-A Belgians are necessary. Currently, the only individual belonging to the Neolithic-A group who has been directly dated is AF004 (5,281-4,977 cal BP).

4.1.2.2 Uniparental markers

The frequencies of mitochondrial lineages found among the Neolithic-A Belgian individuals are in line with the European Neolithic trend, with the presence of haplogroups U5a, U5b, J, K and H (Figure 32) (Gamba *et al*. 2012; Hervella *et al*. 2012; Sanchez-Quinto *et al*. 2012; Fernandez *et al.* 2014; Mathieson *et al.* 2015; Hofmanová *et al*. 2016; Kilinc *et al.* 2016; Lazaridis *et al.* 2016; Jones *et al.* 2017; Brace *et al.* 2019). Two individuals, AF001 and AF010, belonged to the U5a2b4 clade, sharing the same haplotype (Figure 33). This lineage directly branches from a Middle Neolithic Bernburg individual from Germany, dated 5054-4869 BP (Haak *et al.* 2015). U5a2 is one of the main haplogroups within U5a and its expansion took place shortly after the LGM; it is frequent in modern day individuals from central and Eastern Europe (Malyarchuk *et al.* 2010). Four Neolithic-A Belgians, instead, belonged to haplogroup U5b2b (Figure 33), dated to 19 kya (Malyarchuk *et al.* 2010). Other Neolithic and Bronze Age individuals share the same clade, among those the most ancient ones are one Neolithic Sardinian dated to 6179-5930 BP (Olivieri *et al.* 2017) and one Middle Neolithic individual from Germany dated to 5735-5606 BP (Lipson *et al.* 2017). As suggested by both modern and ancient data distribution, the diversification of this haplogroup firstly occurred in Central and Southern Europe and then spread across the continent (Malyarchuk *et al.* 2010). With the exception of U5, the mitochondrial lineages J, K and H had a Late Glacial source in the Near East (Richards *et al.* 2000; Achilli *et al.* 2004; Soares *et al.* 2010; Pala *et al.*

2012; Costa *et al.* 2013; Pereira *et al.* 2017). Their major dispersals in Europe, however, took place during the Neolithic as part of the so-called "Neolithic package", these haplogroups, alongside N1a, T2, V, W and X, represented the majority of the Early European farmer maternal diversity (Haak *et al.* 2010). Two individuals, AF014 and AF036, belonged to haplogroup J1c5 and shared the same haplotype (Figure 33). No other ancient individuals cluster in the same lineage, however J1c5 is characterised by the presence of Neolithic and Bronze Age individuals from Central Europe (Mathieson *et al.* 2015; Fernandes *et al.* 2018). J1c5 has been dated to 9 kya and has a modern European distribution as its root J1c, in particular in central and Eastern Europe (Pala *et al.* 2012). As mentioned before, haplogroup K1a was the most frequent K lineage during the Neolithic in Europe and had a Near Eastern origin, probably as several of its sub-clades. Individual AF026 belonged to haplogroup K1a3a, dated to 14 kya (Costa *et al.* 2013), and shared the haplotype with one Anatolian Neolithic individual, dated to 8450-8150 BP (Mathieson *et al.* 2015) and two individuals from Scotland, Neolithic and Chalcolithic respectively dated to 5892-4987 BP and 4089-3865 BP (Olalde *et al.* 2018). The presence of the Anatolian individual at the root of K1a3a suggests that after originated in the Near East, this haplogroup spread in Europe with the migration of the Anatolian farmers. Haplogroup H, instead, is the most common mitochondrial haplogroup in modern Europeans, with a frequency around 45% (Achilli *et al.* 2004). The origin of H has been associated to the Near East and despite of its sub-clades show a Late Glacial or Post-Glacial coalescence date (Pereira *et al.* 2005; Soares *et al.* 2010), no evidence in ancient DNA prior to the Neolithic have been found in Europe. The Neolithic-A Belgians belonged to diverse H haplogroups, such as H1c (dated to 6448 BP; Behar *et al.* 2012), H1q (dated to 6231 BP; Behar *et al.* 2012), H7d (dated to 7601 BP; Behar *et al.* 2012) and H4a1a1a (dated to 5891 BP; Behar *et al.* 2012). Several Central Neolithic, Chalcolithic and Bronze Age individuals from Scotland, Iberia, Germany and Czech Republic have been found sharing the same haplogroups as the Neolithic Belgians (Knipper *et al.* 2017; Olalde *et al.* 2018), supporting a wide spread of these H sub-clades during the Neolithic across Europe.

The paternal lineages, by comparison, do not show the same degree of variation, with most of the individuals belonging to haplogroup I (Figures 35 and 36). This lineage dates to around 33 kya and has two major sub-haplogroups I1 limited to Scandinavia and Germanic-

speaking areas, dating to 4.6 kya, and the much older I2, dating to ~21 kya and wide spread across Europe (Semino *et al.* 2004; Rootsi *et al.* 2012; Poznick *et al.* 2016; Haber *et al.* 2019). Due to the low coverage, it was not possible to classify many of the samples more accurately. However, the individuals classified as I most likely belonged to the I2 clade, and not I1, as this would be in line with the distribution and age of the haplogroup (Lazaridis *et al.* 2013; Lazaridis *et al.* 2014; Haak *et al.* 2015; Fu *et al.* 2016). Moreover, individuals AF031, AF032 and AF036 belonged to the clade I2a1a1a, dated to 11.7 kya (Francalacci *et al.* 2015). The presence of Y chromosome haplogroup I in Neolithic individuals reflects a measure of genetic continuity from the Mesolithic populations (Haak *et al*. 2015; Mathieson *et al*. 2015). AF018, by contrast, belonged to Y-chromosome haplogroup J, which is absent from the Neolithic in Europe with the exception of two J2 LBK individuals from Austria and one from Croatia; it has also been seen in a Bronze Age individual from Croatia and a Neolithic individual from Anatolia (Mathieson *et al.* 2015; Lazaridis *et al.* 2016; Lazaridis *et al.* 2017; Lipson *et al.* 2017; Mathieson et al. 2018). The origin of haplogroup J has been associated to Palaeolithic times in the Caucasus region and its initial spread across Eurasia took place prior to Mesolithic times, as suggested by finding of Upper Palaeolithic and Mesolithic individuals from Georgia, Russia and Iran (Di Giacomo *et al.* 2004; Jones *et al.* 2015; Mathieson *et al.* 2015; Lazaridis *et al.* 2016; Finocchio *et al.* 2018); it was associated with the Early Neolithic spread into Europe by Semino *et al.* (2000). Its (rare) Early Neolithic presence in Central Europe supports a link to the spread of the LBK. However, a deeper sequencing is necessary to classify AF018 further inside the Y chromosome phylogeny.

4.1.2.3 Stable isotopes

In order to further investigate continuity in the area, not only genetically but also regarding cultural and dietary habits, we performed stable isotopes analyses on the 17 Neolithic-A individuals (Figure 37). The results presented here point to a change in the diet of the Neolithic-A individuals compared to the Mesolithic AF002. In particular, the results seem to include AF003 (Mesolithic) in the Neolithic group. Values ranging from 7.48 to 11.74 ($\delta^{15}N$) and from -22.03 to -20.4 ($\delta^{13}C$) suggest that the Neolithic individuals were consuming a higher quantity of cereals and C3 plants with a lower intake of meat, supporting the scenario of a Neolithic agricultural influence (Fraser *et al.* 2013). Similar values, from 5.8 to 10.2 ($\delta^{15}N$) and from -22.5 to -19.7 ($\delta^{13}C$), have been recorded among LBK individuals from

Germany whose diet was mainly based on domesticated cereals (Fraser *et al.* 2013). However, it is noticeable the increase in the $\delta^{15}N$ of Neolithic Belgians compared to the LBK, probably reflecting traces of riverine diet.

4.1.2.4 Phenotype reconstruction

Another change that probably characterised the transition from the Mesolithic to the Neolithic in Europe was the frequency of some phenotypic features, due to the arrival of new populations from the southeast (Olalde *et al.* 2014; Fu *et al.* 2016). Recent studies have hypothesized that alleles associated with a lighter pigmentation of the skin in Europeans and green and brown eyes were introduced by Early Neolithic farmers from Anatolia, while Mesolithic hunter-gatherers were characterised by a darker skin tone and blue eyes (Olalde *et al.* 2014; Mathieson *et al.* 2015; Fu *et al.* 2016; Brace *et al.* 2019). However, EHG and SHG apparently already possessed lighter skin pigmentation, prior to the onset of the Neolithic (Mathieson *et al.* 2018). The phenotype prediction results suggest that the Neolithic-A Belgians displayed, overall, an intermediate tone of skin colour and blue eyes (Table 16). This interpretation is consistent with the scenario described previously, with the Neolithic-A Belgians carrying a pigmentation phenotype closer to hunter-gatherer populations as result of their high WHG ancestry. Like the Mesolithic individuals, Neolithic-A Belgians carried the ancestral state of rs4988235 (*LCTa*) and rs182549 (*LCTb*) (Table 17) suggesting that most likely they were not able to digest milk in the adulthood.

4.1.3 Neolithic-B Belgians

The Neolithic-B group includes seven individuals across the three sites, four from Trou Al'Wesse, two from Grotte du Mont Falise and one from Abri Sandron. Stratigraphic information was available only for the samples from Trou Al'Wesse: two individuals, AF023 and AF025, were found in the deepest layer of the chimney burial (4), AF020 was found in layer 3 and AF017 in layer 2 (Figure 15 and Table 4). The sites of Grotte du Mont Falise and Abri Sandron displayed signs of human occupation dated to the Late Neolithic (Abri Sandron: 4765-4619 cal BP, 4769-4607 cal BP, 4966-4815 cal BP; Grotte du Mont Falise: 4768-4611 cal BP, 4892-4807 cal BP) (Toussaint 2002; Toussaint 2003).

4.1.3.1 Autosomal evidence

Neolithic-B Belgians overall, resemble Late Neolithic/Early Bronze Age individuals from Europe, Single Grave and Battle Axe (Figure 23). However, given the newly obtained radiocarbon dates for individuals AF007 and AF017, respectively dated 4,828-4,626 cal BP and 1,929-1,827 cal BP, the Neolithic B group includes now individuals spanning the Late Neolithic (AF007) to the Iron Age (AF017). Genetically, in addition to the WHG and Neolithic components, Neolithic-B individuals display a substantial third genetic component (CHG), ranging from 20% to 35% (Figure 24). This CHG component, associated with the Pontic-Caspian Steppe populations, was not present in comparable amount in any of the Neolithic-A samples (Figure 24). Late Neolithic/Bronze Age populations from Europe overall show a decreasing cline of Steppe ancestry from east to west mirroring the direction of the migration (Juras *et al.* 2018). The presence of this Steppe ancestry in the PCA plot pushes the Neolithic-B group towards Yamnaya populations, overlapping Late Neolithic/Bronze Age individuals from Europe (Figure 21). To further investigate the origin of this Steppe component in Neolithic-B Belgians, a D-statistics test was performed (Figure 31). As expected, the results assign the source of the Steppe ancestry carried by Neolithic-B Belgians to the Yamnaya population from Russia(Figure 31). The confirmation of the presence of this component in Neolithic individuals from Belgium is surprising, since, considering the age of the sites, it could represent an early appearance of CHG ancestry in Western Europe. In particular, the newly obtained date for AF007 (4,828-4,626 cal BP), makes this the oldest ancient individual from Western Europe having Steppe component and carrying R1b-M269 as Y-chromosome haplogroup marker. The oldest sample from the whole Europe, instead, is represented by an Eneolithic individual from Croatia, dated 4,833-4,615 cal BP (Mathieson *et al.* 2018). Currently, we are in the progress of radiocarbon dating the remaining Neolithic-B individuals to add more data supporting this hyphothesis.

4.1.3.2 Uniparental markers

Some genetic continuity between the Neolithic-A and the Neolithic-B groups is supported by the presence of shared mitochondrial haplogroups, and indeed haplotypes in the case of a K1a4a1 sequence shared by Neolithic-A individuals AF003 and AF015 and Neolithic-B individual AF007 (Figures 32 and 33). In the Neolithic-B group it is possible to observe the presence of haplogroup T1a1n carried by two individuals from Trou Al'Wesse (Figures 32 and 33), not seen among Neolithic-A Belgians. Haplogroup T1a1, thought to be associated

with the spread of Late Neolithic/Bronze Age populations into north-western Europe (Brandt *et al.* 2013; Pereira *et al.* 2017), has been found exclusively in Late Neolithic/Bronze Age individuals from Europe, Bell Beaker samples, Yamnaya-related populations and Bronze Age Caucasus individuals (Allentoft *et al.* 2015; Haak *et al.* 2015; Mathieson *et al.* 2015; Krzewinska *et al.* 2018; Mathieson *et al.* 2018; Mittnik *et al.* 2018; Olalde *et al.* 2018).

Analysis on the Y chromosome supported the scenario, described previously, of shared ancestry between Yamnaya-related populations and Neolithic-B Belgians, since five Neolithic-B males out of six belong to the Y chromosome haplogroup R1b-M269. The coverage of the remaining individual, AF025, was too low to obtain an accurate haplogroup classification, currently is classified as R. The R1b-M269 lineage has been already described as the marker of the Yamnaya migrations and the genetic replacement of Y chromosome haplogroups G, J and I to mainly R1b (Haak *et al.* 2015; Mathieson *et al.* 2015; Fu *et al.* 2016; Olalde *et al.* 2018; Olalde *et al.* 2019).

4.1.3.3 Stable isotopes

The current stable isotopes data presented here do not show significant difference in the dietary habits between Neolithic-A and Neolithic-B Belgians, with marginally lower values of $\delta^{13}C$ for the Neolithic-B (Figure 37). However, this assumption is based on a very limited sample-size, two individuals. Currently we are in the process of analysing the stable isotope data for the remaining Neolithic-B individuals, to eventually assess a possible change in the diet.

4.1.3.4 Phenotype reconstruction

Interestingly, one Neolithic-B individual from Grotte du Mont Falise, AF029, carried a mutation in the SNP rs4988235, in the *LCTa* gene, suggesting that AF029 most likely was able to digest milk in adulthood (Table 17). This particular mutation is under-selection in European populations (Bersaglieri *et al.* 2004). rs4988235 increased its frequency among Europeans after the Neolithic when human populations started to consume milk and dairy products as a consequence of animal husbandry (Burger *et al.* 2007; Allentoft *et al.* 2015; Mathieson *et al.* 2015). This particular finding, combined with the presence of CHG ancestry in AF029, could place this individual, and possibly the entire Neolithic-B group, later in time compared to the Neolithic-A group.

4.1.4 Modern Belgium

From a starting dataset of 166 modern Belgian mitogenomes, I was able classify into known mitochondrial haplogroups 124 of them. Despite the limited sample size, these sequences belong to Belgian individuals from all the provinces of the country, giving a full coverage of the area (Figure 38). The comparison between the modern and the ancient Belgian mitochondrial data suggest the absence of genetic continuity of particular mitochondrial lineages in the specific area of Belgium. However, all the modern Belgian individuals display mitochondrial haplogroup frequencies comparable with the rest of Europe (Figure 39) (Karachanak *et al.* 2011).

4.2 Mitochondrial haplogroup HV

Mitochondrial haplogroup HV includes the root of the most common haplogroup of modern Europeans. However, its origin and routes of dispersal are still debated. Haplogroup HV as whole, including the main clades H and HV0, has an overall modern geographical distribution mainly based in Europe. However, the most basal clades within HV and the root of HV itself are characterised by a majority of South East European and Near Eastern individuals (Appendix 17). The new ML age calculations place the origins of the haplogroup at 23 (19.3; 26.9) kya (Table 18). This estimate is slightly younger than the previous estimate by Soares and collaborators (2009), which dated the origin of HV back to 27.1 kya. However, this latter calculation also included haplogroup H, which has been excluded from this new calculation. For this reason, the newly calculated coalescence age of ~23 kya, may be a minimum estimate. In any case, the new estimate confirms an LGM origin for haplogroup HV as a whole. Within HV, several sub-haplogroups appear also as dating back to the LGM or Late Glacial, such as HV0 (13.8 kya), HV1 (20.0 kya), HV2 (22.6 kya), HV4 (17.6 kya), HV7 (18.6 kya), HV12 (19.4 kya), HV13 (20.8 kya) and HV14 (15.0 kya). These ages are in line with previous estimates (Richards *et al*. 2000; Achilli *et al*. 2004; De Fanti *et al.* 2015).

Haplogroup HV0 is one of the largest branches within HV (excluding H). Almost all of the individuals, both ancient and modern, within haplogroup HV0 are European (with non-Europeans generally at the tips rather than at the root) and therefore we can assume an

origin within the European continent (Appendix 17). In contrast with HV0, the other sister clades within HV, such as HV1, HV2, HV4, HV7, HV12, HV13, HV14 and HV18, are almost entirely Near Eastern, with Near Eastern and Caucasian individuals branching from the roots. Some of these Near Eastern haplogroups, however, include European sub-haplogroups, such as: HV4a1a characterised by Spanish, Basque and French individuals, suggesting an expansion from the Franco-Cantabrian region (Gomez-Carballa *et al.* 2012), the European HV7a and HV7b which differ from the sister clade HV7c entirely Caucasian, or the many sub-haplogroups within HV+16311 characterised by younger coalescence ages, HV6 (3.4 kya), HV8 (4.3 kya), HV9 (9.3 kya) HV10 (3.3 kya), HV16 (5.7 kya).

Considering the ages of HV and its sub-haplogroups, and the geographic distribution of modern individuals, the involvement of HV (excluding H and HV0) in migratory events from the putative source (the Near East) to Europe was tested using the founder analysis approach. The results suggest that HV was involved in two main migratory events from the Near East to Europe (Figure 40): a smaller-scale event during the postglacial (~11.5 kya), and a major migration during the Neolithic (~7.5 kya). The second event may coincide with the migration of early farmers to Europe from the Fertile Crescent and/or Anatolia during the Neolithic transition (Figure 40).

To further investigate the possibility that southern Europe might have acted as a source of Near Eastern HV lineages, and the place of differentiation of European lineages (HV0) that were then spread into central and northern Europe, the combination EuropeSouth-Source/EuropeNorth-Sink was then tested (Figure 41). The results show that the majority of the spreading lineages (in this case mostly HV0) tend to peak at ~9.0 kya, with a secondary peak at ~4.0-5.0 kya.

These dates seem to imply that some of the older European HV lineages might have “hitch-hiked” the Neolithic dispersal from its source in the Near East, to southern Europe, and then from there within Europe. If we accept this scenario, then the Mediterranean area seems to have acted as a source of some HV lineages (mainly HV0) that were first introduced into Europe during the Late Glacial and others in the Neolithic, and spread across the rest of the continent during the Neolithic (~9 kya) and later during the Bronze Age (~4.3 kya) (Figure

41). This appears to be analogous to the situation for haplogroups J and T (Pereira *et al.* 2017) and may also be the case for some of haplogroup H.

Because of its rarity, at present, ancient DNA does not clarify the picture regarding HV very much, although they provide some evidence regarding the subsequent spread of HV0/V. The oldest Near Eastern appearance, so far, of haplogroup HV is dated to the Mesolithic time in Iran, in the individual I1293, dated to 7,250 ± 40 BP; I1293 was further classified as HV2a1 (Lazaridis *et al.* 2016). However, an even older individual has been found in Europe, an LBK individual from Germany, dated to 7,450-6,750 BP, belonging to the clade HV+16311 (Haak *et al.* 2015).

Considering that the coalescence age of haplogroup V calculated using ML is around 8.8 kya (Table 18) and the presence at the root of V of one Early Neolithic Spanish individual dated around 7 kya (Haak *et al.* 2015), the origin of haplogroup V could potentially be located in Iberia, as originally suggested by Torroni and collaborators in 2001, albeit with an earlier chronology which now seems less plausible. However, an origin in south-east Europe now seems more likely, given the early dates in south-east and central Europe, but with a secondary Neolithic expansion from Iberia. More ancient DNA data, and more refined founder analysis, will be necessary to test these scenarios in detail.

5. CONCLUSIONS

The aim of this thesis was to shed light on the Neolithic transition on ancient and modern European populations. To achieve this, two approaches were taken: i) the analysis of human remains from three prehistoric archaeological sites in Belgium; and ii) the phylogeographic analysis of modern sequences belonging to mitochondrial haplogroup HV.

i) The ancient Belgian samples were identified as target for analysis for two main reasons.First, they represented an area in Europe that chronologically, and geographically, has been scarcely sampled for ancient DNA studies. Second, because of what we know about the archaeology of the area, in particular in association with the appearance of controversial pottery traditions, such as La Hoguette (present at one of the sites) and Limburg. Their distinctive appearance at the boundary between the Mesolithic and Neolithic opens the possibility that something unique might have occurred in the area at the Mesolithic-Neolithic transition, perhaps with the active involvement of local hunter-gatherers.

   Very excitingly, my results seem to suggest that in Belgium the Mesolithic-Neolithic transition might have gone through a multi stage process. First, although we must treat this result very cautiously until more data is analysed, it is possible that Neolithic-related ancestry might have infiltrated Mesolithic communities very early on (AF002). Secondly, by the Middle Neolithic at least (although we cannot confidently attribute this to earlier La Hoguette influence without further research), the fully-fledged Neolithic communities of the area (from all three sites) had higher level of Mesolithic assimilation than has been seen to date during the western European Neolithic ("Neolithic-A"). The genetic analysis showed that one individual (AF003) attributed to the Mesolithic by the excavators in fact belonged to this cluster, showing the power of the archaeogenetic analysis to challenge archaeological interpretations.

   The uniparental markers show that there may also have been a sex bias in the Neolithic settlement of the area: although many of the Neolithic-A mtDNA lineages can be attributed to Neolthic expansion from southern Europe, there is only a single Y-

chromosome lineage of potentially southern Neolithic origin amongst these samples, with the remainder belonging to the indigenous European haplogroup I.

There is also a second cluster of individuals, referred to as “Neolithic-B”. In these samples, a new genome-wide Steppe (or CHG) component is evident, and the indigenous Mesolithic and Neolithic components appear to have been diluted to reach values comparable with other similar ancient populations after newcomers arrived carrying Yamnaya-related ancestry. This cluster also carries exlusively the Yamnaya-associated Y-chromosome haplogroup R1b-M269, as well as a number of mtDNA haplogroups associated with Late Neolithic/Bronze Age arrivals from the east. Given that the radiocarbon dates from one of the sites cluster to the late Middle Neolithic, and the others to the early Late Neolithic, this clustering into two distinct populations (one unique but related to other European Neolithic and Mesolithic populations, and the other closer to Corded Ware/Chalcolithic/Bronze Age populations) was unexpected and again demonstrates the power of genetic analysis.

Despite this unexpected clustering, overall all three sites under analysis are genetically very similar. This is not only evident from the autosomal analysis but is further supported by the presence of shared mitochondrial and Y-chromosome haplogroups and even haplotypes. This fits with their close geographic proximity, and might suggest that the rivers connecting the sites might have acted as prehistoric “highways” by promoting contact among the sites during the Neolithic. The work presented here also emphasizes the continuing importance of the use of uniparental marker systems. Despite giving only a partial understanding of the history if analysed separately (the maternal ancestry, in the case of mtDNA, and paternal for the Y chromosome) when combined with whole-genome analysis they become a very powerful tool which can enrich the interpretation of autosomal data. This is especially the case when they point to the existence of sex-specific dispersals, which are well-established for the Late Neolithic/Bronze Age but have been less clear to date for the spread of the earlier Neolithic.

ii) The results presented in this thesis, obtained using a dataset of 1,101 complete mitogenomes, published and unpublished, confirmed the proposed Upper Palaeolithic

Near Eastern origin of HV (Pereira *et al*. 2005) and its involvement in migratory events, from the Near East to Europe during the Late Glacial, and the Neolithic expansion from southern Europe to the rest of the continent. The findings suggest that HV lineages may have been introduced from the Near East into the Mediterranean area continuously between the Late Glacial and the Neolithic. The Mediterranean area might therefore have functioned as reservoir of HV lineages and area of differentiation of important European HV variants, such as HV0; during the Neolithic, and later on during the Bronze Age, Near Eastern and European HV lineages were then carried into the rest of Europe by subsequent dispersals.

6. REFERENCES

- Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A. (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. AJHG. 75, 910-918.
- Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti AS, Bandelt H, Semino O, Torroni A. (2005) Saami and Berbers – An Unexpected Mitochondrial DNA Link. AJHG. 76, 5:883-886.
- Alexander DH, Novembre J, Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 19:1655-1664.
- Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E, Zhang G, Scofield RP, Holdaway RN, Bunce M. (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B: Biological Sciences. 279, 1748:4724-4733.
- Allentoft ME, Sikora M, Sjögren K, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, Malaspinas A, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L, Baron J, Della Casa P, Dąbrowski P, Duffy PR, Ebel AV, Epimakhov A, Frei K, Furmanek M, Gralak T, Gromov A, Gronkiewicz S, Grupe G, Hajdu T, Jarysz R, Khartanovich V, Khokhlov A, Kiss V, Kolář J, Kriiska A, Lsak I, Longhi C, McGlynn G, Merkevicius A, Merkyte I, Metspalu M, Mkrtchyan R, Moiseyev V, Paja L, Pálfi G, Pokutta D, Pospieszny Ł, Price TD, Saag L, Sablin M, Shishlina N, Smrčka V, Soenov VI, Szeverényi V, Tóth G, Trifanova SV, Varul L, Vicze M, Yepiskoposyan L, Zhitenev V, Orlando L, Sicheritz-Pontén T, Brunak S, Nielsen R, Kristiansen K, Willerslev E. (2015) Population genomics of Bronze Age Eurasia. Nature. 522. 167-172.
- Ammerman AJ, Cavalli-Sforza LL. (1984) The Neolithic Transition and the Genetics of Populations in Europe. Princeton University Press. Princeton.
- Amorim A. (1999) Archaeogenetics. Journal of iberian archaeology. 1, 15-26.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. (1981) Sequence and organization of the human mitochondrial genome. Nature. 290:457-465.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 23:147.
- Anthony DW. (2007) Herding and gathering during the Late Bronze Age at Krasnosamarskoe, Russia, and the end of the dependency model of steppe pastoralism. Social Orders and Social Landscapes. Cambridge Univ Press.
- Anthony DW. (2010) The horse, the wheel and language: how Bronze-Age riders from the Eurasian steppes shaped the modern world. Princeton Univ Press.
- Ashcroft MB. (2010) Identifying refugia from climate change. Journal of Biogeography. 37, 8:1407-1413.
- Avery OT, MacLeod CM, McCarty M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J Exp Med. 79, 2:137-158.
- Avise JC. (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge.
- Awadalla P, Eyre-Walker A, Maynard Smith J. (1999) Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA. Science. 286, 5449:2524-2525.
- Bailey G, Spikins P. (2008) Mesolithic Europe. Cambridge Univ Press.
- Baldia MO. (2006) The Earliest Bandkeramik. The Comparative Archaeology Web.
- Bandelt HJ, Forster P, Sykes BC, Richards MB. (1995) Mitochondrial portraits of human populations using median networks. Genetics. 2. 743-53.
- Bandelt HJ, Forster P, Röhl A. (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evo. 1. 37-48.

- Bandelt HJ, Kloss-Brandstatter A, Richards MB, Yao YG, Logan I. (2013) The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies. Journal of Human Genetics. 1-12.
- Barnett WK. (2000) Cardial pottery and the agricultural transition in Mediterranean Europe. Europe's first farmers. Cambridge Univ Press.
- Batini C, Hallast P, Zadik D, Delser PM, Benazzo A, Ghirotto S, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Dupuy BM, Eriksen HA, King TE, de Munain AL, Lopez-Parra AM, Loutradis A, Milasin J, Novelletto A, Pamjav H, Sajantila A, Tolun A, Winney B, Jobling MA. (2015) Large-scale recent expansion of European patrilineages shown by population resequencing. Nat com. 6, 7152.
- Beckerman F. (2015) Corded Ware Coastal Communities: Using Ceramic Analysis to Reconstruct Third Millennium BC Societies in the Netherlands. Sidestone Press.
- Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali E, Silva NM, Kivisild T, Torroni A, Villems R. (2012) A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. AJHG. 90, 4:675-684.
- Bergström A, Nagle N, Chen Y, McCarthy S, Pollard MO, Ayub Q, Wilcox S, Wilcox L, van Oorschot RAH, McAllister P, Williams L, Xue Y, Mitchell RJ, Tyler-Smith C. (2016) Deep Roots for Aboriginal Australian Y Chromosomes. Current Biology. 26, 6:809-813.
- Birks HH, Ammann B. (2000) Two terrestrial records of rapid climatic change during the glacial-Holocene transition (14,000- 9,000 calendar years B.P.) from Europe. PNAS. 97, 4:1390-1394.
- Birky CW. (2001) The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models. Annual Review of Genetics. 35, 125-148.
- Boattini A, Martinez-Cruz B, Sarno S, Harmant C, Useli A, Sanz P, Yang-Yao D, Manry J, Ciani G, Luiselli D, Quintana-Murci L, Comas D, Pettener D and The Genographic Consortium. (2013) Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. PLoS One. 8.
- Bogenhagen D, Clayton DA. (1977) Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. Cell. 11, 4:719-727.
- Bollongino R, Nehlich O, Richards MP, Orschiedt J, Thomas MG, Sell C, Zuzana F, Powell A, Burger J. (2013) 2000 Years of Parallel Societies in Stone Age Central Europe. Science. 342, 6157:479-481.
- Boore JL. (1999) Animal mitochondrial genomes. Nucleic Acids Res. 27:1767-1780.
- Bower B. (2010) Humans: Dating the dawn of copper making: Find may alter views on time, place of metallurgy's origins. Science News.
- Bowcock AM, Hebert JM, Mountain JL, Kidd JR, Rogers J, Kidd KK, Cavalli-Sforza LL. (1991) Study of an additional 58 DNA markers in five human populations from four continents. Gene geogr. 5(3):151-73.
- Bowmaker M, Yang MY, Yasukawa T, Reyes A, Jacobs HT, Huberman JA, Holt IJ. (2003) Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. J Biol Chem. 278(51):50961-9.
- Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind CJ, Lueth F, Terberger T, Hiller J, Matsumura S, Forster P, Burger J. (2009) Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. Science. 326, 5949:137-140.
- Brandt G, Haak W, Adler CJ, Roth C, Szecsenyi-Nagy A, Karimnia S, Moeller-Rieker S, Meller H, Ganslmeier R, Friederich S, Dresely V, Nicklisch N, Pickrell JK, Sirocko F, Reich D, Cooper A, Alt KW, The Genographic Consortium. (2013) Ancient DNA reveals key stage in the formation of central European mitochondrial genetic diversity. Science 342, 257-261.
- Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, Pääbo S. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. PNAS. 104, 37:14616-14621.
- Brooks SJ, Birks HJB. (2001) Chironomid-inferred air temperatures from Lateglacial and Holocene sites in north-west Europe: progress and problems. Quaternary Science Reviews. 20, 16-17:1723-1741.
- Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, Cooper A. (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of *post mortem* miscoding lesions. Nucleic Acid Research. 35, 17:5717-5728.
- Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Adler CJ, Richards SM, Der Sarkissian C, Ganslmeier R, Friederich S, Dresely V, van Oven M, Kenyon R, Van der Hoek MB, Korlach J, Luong K, Ho SYM, Quintana-Murci L, Behar DM, Meller H, Alt KW, Cooper A, and The Genographic Consortium. (2013) Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat. Commun. 4, 1764.

- Broushaki F, Thomas MG, Link V, López S, van Dorp L, Kirsanow K, Hofmanová Z, Diekmann Y, Cassidy LM, Díez-del-Molino D, Kousathanas A, Sell C, Robson HK, Martiniano R, Blöcher J, Scheu A, Kreutzer S, Bollongino R, Bobo D, Davoudi H, Munoz O, Currat M, Abdi K, Biglari F, Craig OE, Bradley DG, Shennan S, Veeramah KR, Mashkour M, Wegmann D, Hellenthal G, Burger J. (2016) Early Neolithic genomes from the eastern Fertile Crescent. Science. 353, 6298:499-503.
- Brown FH, McDougall I, Fleagle JG. (2012) Correlation of the KHS Tuff of the Kibish Formation to volcanic ash layers at other sites, and the age of early *Homo sapiens* (Omo I and Omo II). Journal of Human Evolution. 63, 4:577-585.
- Cabana GS, Hulsey BI, Pack F. (2013) Molecular Methods. Research methods in Human Skeletal Biology. 449-480.
- Cao L, Shitara H, Horii T, Nagao Y, Imai H, Abe K, Hara T, Hayashi J, Yonekawa H. (2007) The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells. Nature Genetics. 39, 386-390.
- Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, Rasmussen M, Gravel S, Guillen S, Nekhrizov G, Leshtakov K, Dimitrova D, Theodossiev N, Pettener D, Luiselli D, Sandoval K, Moreno-Estrada A, Li Y, Wang J, Gilbert MTP, Willerslev E, Greenleaf WJ, Bustamante CD. (2013) Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. AJHG. 93, 5:852-864.
- Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, Hartwell B, Bradley DG. (2016) Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. PNAS. 113, 2:368-373.
- Caut S, Angulo E, Courchamp F. (2008) Discrimination factors (Δ^{15}N and Δ^{13}C) in an omnivorous consumer: effect of diet isotopic ratio. Functional Ecology. 255-263.
- Chandler H, Sykes B, Zilhão J. (2005) Using ancient DNA to examine genetic continuity at the Mesolithic-Neolithic transition in Portugal. Monografías del Instituto Internacional de Investigaciones Prehistóricas de Cantabria 1. In: Arias P, Ontanon R, Garcia-Monco C (Eds.), Actas del III Congreso del Neolítico en la Península Iberica, Santander, pp. 781-786.
- Chauvet J, Brunel Deschamps E, Hillaire C. (1996) Dawn of art : the Chauvet Cave : the oldest known paintings in the world. HN Abrams.
- Childe VG. (1925) The Dawn of European Civilisation. Kegan Paul, London.
- Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. (2009) The Last Glacial Maximum. Science. 325, 5941:710-714.
- Clayton DA. (1982) Replication of animal mitochondrial DNA. Cell. 28(4):693-705.
- Cree LM, Samuels DC, de Sousa Lopes SC, Rajasimha HK, Wonnapinij P, Mann JR, Dahl HM, Chinnery PF. (2008) A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. Nature Genetics. 40, 249-254.
- Crombé P, Perdaen Y, Sergant J. (2005) La néolithisation de la Belgique: quelques réflexions. Bulletin de la Société Préhistoriques française XXXVI: 47-66.
- Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, Scozzari R. (2011) A Revised Root for the Human Y Chromosomal Phylogenetic Tree: The Origin of Patrilineal Diversity in Africa. AJHG. 88, 6:814-818.
- Cunliffe B. (2008) Europe between the oceans, 9000 BC to AD 1000. Yale Univ Press.
- Cunliffe B, Koch JT. (2012) Celtic from the West: Alternative Perspectives from Archaeology, Genetics, Language and Literature. Oxbow Books. 15.
- Czebreszuk J. (2004) Similar But Different. Bell Beakers in Europe. Adam M Univ.
- Dabney J, Meyer M, Paabo S. (2013) Ancient DNA Damage. Cold Sptring Harb Perspect Biol. 5,7.
- De Fanti S, Barbieri C, Samo S, Sevini F, Vianello D, Tamm E, Metspalu E, van Oven M, Hübner A, Sazzini M, Franceschi C, Pettener D, Luiselli D. (2015) Fine dissection of human mitochondrial DNA haplogroup HV lineages reveals Paleolithic signatures from European glacial refugia. PloS One 10, 12.
- De la Rua C, Izagirre N, Alonso S, Hervella M. (2015) Ancient DNA in the Cantabrian fringe populations: A mtDNA study from Prehistory to Late Antiquity. Quaternary International. 364, 306-311.
- Der Sarkissian C, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, Koshel S, Zaporozhchenko V, Gronenborn D, Moiseyev V, Kolpakov E, Shumkin V, Alt KW, Balanovska E, Cooper A, Haak W, the Genographic Consortium. (2013) Ancient DNA Reveals Prehistoric Gene-Flow from Siberia in the Complex Human Population History of North East Europe. PloS Genet. 9, 2.

- Derclaye C, López Bayón I, Collin F, Otte M. (1999) Contributions à la connaissance du Mésolithique récent en Ardennes. Étude archéologique de la couche 4 du Trou Al'Wesse (Petit-Modave, province de Liège, Belgique). Notae Praehistoricae. 19:85-95.
- Ermini L, Olivieri C, Rizzi E, Corti G, Bonnal R, Soares P, Luciani S, Marota I, De Bellis G, Richards MB, Rollo F. (2008) Complete mitohcondrial genome sequence of the Tyrolean Iceman. Curr Biol. 18, 21:1687-93.
- Feldman M, Fernández-Domínguez E, Reynolds L, Baird D, Pearson J, Hershkovitz I, May H, Goring-Morris N, Benz M, Gresky J, Bianco RA, Fairbairn A, Mustafaoğlu G, Stockhammer PW, Posth C, Haak W, Jeong C, Krause J. (2019) Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. Nature Communications. 10, 1218.
- Fernandes V, Triska P, Pereira JB, Alshamali F, Rito T, Machado A, Fajkošová Z, Cavadas B, Černý V, Soares P, Richards MB, Pereira L. (2015) Genetic stratigraphy of key demographic events in Arabia. PloS One. 10, 3.
- Fontijn DR. (2003) Sacrifical landscapes : cultural biographies of persons, obkects and "natural" places in the Bronze Age of the Southern Netherlands, c. 2300-600 BC. DANS.
- Forster P, Harding R, Torroni A, Bandelt HJ. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. AJHG. 59: 935-945.
- Fraipont J. (1897) La grotte du mont Falhise. Bulletin de l'Académie royale de Belgique. 1, 47-51.
- Fraipont J. (1898) L'ossuaire de l'abri-sous-roche de Sandron, à Huccorgne (province de Liège). Bulletin de la Société d'Anthropologie de Bruxelles. 311-332, 382-391.
- Fu Q, Rudan P, Pääbo S, Krause J. (2012) Complete Mitochondrial Genomes Reveal Neolithic Expansion into Europe. Plos ONE. 7, 3.
- Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, Pääbo S. (2013) DNA analysis of an early modern human from Tianyuan Cave, China. PNAS. 110, 6:2223-2227.
- Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Petri AA, Prüfer K, de Filippo C, Meyer M, Zwyns N, Salazar-Garcia DC, Kuzmin YV, Keates SG, Kosintsev PA, Razhev DI, Richards MP, Peristov NV, Lachmann M, Douka K, Higham TFG, Slatkin M, Hublin J, Reich D, Kelso J, Viola TB, Pääbo S. (2014) The genome sequence of a 45,000-year-old modern human from western Siberia. Nature. 514, 7523:445-449.
- Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland N, Lazaridis I, Nickel B, Viola B, Prüfer K, Meyer M, Kelso J, Reich D, Pääbo S. (2015) An early modern human from Romania with a recent Neanderthal ancestor. Nature. 524, 216-219.
- Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, Meyer M, Mittnik A, Nickel B, Peltzer A, Rohland N, Slon V, Talamo S, Lazaridis I, Lipson M, Mathieson I, Schiffels S, Skoglund P, Derevianko AP, Drozdov N, Slavinsky V, Tsybankov A, Grifoni Cremonesi R, Mallegni F, Gély B, Vacca E, González Morales MR, Straus LG, Neugebauer-Maresh C, Teschler-Nicola M, Constantin S, Moldovan OT, Benazzi S, Peresani M, Coppola D, Lari M, Ricci S, Ronchitelli A, Valentin F, Thevenet C, Wehrberger K, Grigorescu D, Rougier H, Crevecoeur I, Flas D, Semal P, Mannino MA, Cupillard C, Bocherens H, Conard NJ, Harvati K, Moiseyev V, Drucker DG, Svoboda J, Richards MP, Caramelli D, Pinhasi R, Kelso J, Patterson N, Krause J, Pääbo S, Reich D. (2016) The genetic history of Ice Age Europe. Nature. 534, 200-205.
- Gamba C, Fernandez E, Tirado M, Deguilloux MF, Pemonge MH, Utrilla P, Edo M, Molist M, Rasteiro R, Chikhi L, Arroyo-Pardo E. (2012) Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. Mol. Ecol. 21, 45-56.
- Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki L, Kővári I, Pap I, Anders A, Whittle A, Dani J, Raczky P, Higham TFG, Hofreiter M, Bradley DG, Pinhasi R. (2014) Genome flux and stasis in a five millennium transect of European prehistory. Nature Communications. 5, 5257.
- Gandini F, Achilli A, Pala M, Bodner M, Brandini S, Huber G, Egyed B, Ferretti L, Gómez-Carballa A, Salas A, Scozzari R, Cruciani F, Coppa A, Parson W, Semino O, Soares P, Torroni A, Richards MB, Olivieri A. (2016) Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. Scientific Reports. 6.
- Gilbert MTP, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, Tikhonov A, Dalén L, Kuznetsova T, Kosintsev P, Campos PF, Higham T, Collins MJ, Wilson AS, Shidlovskiy F, Buigues B, Ericson PGP, Germonpré M, Götherström A, Iacumin P, Nikolaev V, Nowak-Kemp M, Willerslev E, Knight JR, Irzyk

GP, Perbost CS, Fredrikson KM, Harkins TT, Sheridan S, Miller W, Schuster SC. (2007) Whole-Genome Shotgun Sequencing of Mitochondria from Ancient Hair Shafts. Science. 317, 5846:1927-1930.
- Giles RE, Blanc H, Cann HM, Wallace DC. (1980) Maternal inheritance of human mitochondrial DNA. PNAS. 77, 11:6715-6719.
- Graves JAM. (2006) Sex Chromosome Specialization and Degeneration in Mammals. Cell. 124, 5:901-914.
- Gronenborn D. (1999) A Variation on a Basic Theme: The Transition to Farming in Southern Central Europe. Journal of World Prehistory. 13, 2:123-210.
- Gronenborn D. (2003) Migration, acculturation and culture change in western temperate Eurasia. Documenta Praehistorica. 30.
- Güldoğan E. (2010) Mezraa-Teleilat settlement 'Impressed' Ware and transferring Neolithic life style? Proceedings of the 6th International Congress of the Archaeology of the Ancient Near East. 3.
- Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-Varela R, Sverrisdóttir OO, Daskalaki EA, Skoglund P, Naidoo T, Svensson EM, Bermúdez de Castro JM, Carbonell E, Dunn M, Storå J, Iriarte E, Arsuaga JL, Carretero JM, Götherström A, Jakobsson M. (2015) Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. PNAS. 112, 38:11917-11922.
- Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardon K, Fu Q, Mittnik A, Banffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szecsenyi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 522, 207-211.
- Haeck J. (1964) La grotte du Mont Falise à Antheit, vallée de la Méhaigne, province de Liège. Bulletin de la Société royale belge d'Anthropologie et de Préhistoire. 74, 43-58.
- Haber M, Jones AL, Connell BA, Arciero AE, Yang H, Thomas MG, Xue Y, Tyler-Smith C. (2019) A rare deep-rooting D0 African Y-Chromosomal Haplogroup and Its implications for the Expansion of modern humans out of Africa. Genetics. 212.4,1421-1428.
- Hallast P, Batini C, Zadik D, Delser PM, Wetton JH, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Bisol GD, Dupuy BM, Eriksen HA, Jorde LB, King TE, Larmuseau MH, Lopez de Munain A, Lopez-Parra AM, Loutradis A, Milasin J, Novelletto A, Pamjav H, Sajantila A, Schempp W, Sears M, Tolun A, Tyler-Smith C, Van Geystelen A, Watkins S, Winney B, Jobling MA. (2014) The Y-Chromosome Tree Bursts into Leaf: 13,000 High-Confidence SNPs Covering the Majority of Known Clades. Mol Bio and Evol. 32, 3:661-673.
- Hansen HB, Damgaard PB, Margaryan A, Stenderup J, Lynnerup N, Willerslev E, Allentoft ME. (2017) Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. PloS ONE. 12, 1.
- Harlan JR, Zohary D. (1966) Distribution of Wild Wheats and Barley. Science. 153, 3740:1074-1080.
- Hauswirth WW, Laipis PJ. (1982) Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. PNAS. 79, 15:4686-4690.
- Hauzeur A. (2009) First Appearance of Pottery in Western Europe: The Questions of La Hoguette and Limburg Ceramics. Early Farmers, Late Foragers, and Ceramic Traditions: On the Beginning of Pottery in the Near East and Europe. Newcastle upon Tyne: Cambridge Scholars Publishing, 167-188.
- He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA, Kinzler KW, Vogelstein B, Papadopoulos N. (2010) Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 464, 610-614.
- Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P. (2014) Library construction for next-generation sequencing: Overviews and challenges. Biotechniques. 56: 61.
- Hervella M, Izagirre N, Alonso S, Fregel R, Alonso A, Cabrera VM, de la Rua C. (2012) Ancient DNA from hunter-gatherer and farmer groups from Northern Spain supports a random dispersion model for the Neolithic expansion into Europe. PLoS One. 7.
- Higgins D, Rohrlach AB, Kaidonis J, Townsend G, Austin JJ. (2015) Differential Nuclear and Mitochondrial DNA Preservation in Post-Mortem Teeth with Implications for Forensic and Ancient DNA Studies. PloS ONE. 10, 5.
- Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. (1984) DNA sequences from the quagga, an extinct member of the horse family. Nature. 312, 282-284.

- Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-del-Molino D, van Dorp L, López S, Kousathanas A, Link V, Kirsanow K, Cassidy LM, Martiniano R, Strobel M, Scheu A, Kotsakis K, Halstead P, Triantaphyllou S, Kyparissi-Apostolika N, Urem-Kotsou DC, Ziota C, Adaktylou F, Gopalan S, Bobo DM, Winkelbach L, Blöcher J, Unterländer M, Leuenberger C, Çilingiroğlu Ç, Horejs B, Gerritsen F, Shennan S, Bradley DG, Currat M, Veeramah KR, Wegmann D, Thomas MG, Papageorgopoulou C, Burger J. (2016) Early farmers from across Europe directly descended from Neolithic Aegeans. PNAS. 113, 25:6886-6891.
- Hofreiter M, Jaenicke V, Serre D, Haeseler A, Pääbo S. (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acid Research. 29, 23:4793-4799.
- Hublin JJ, Neubauer S, Gunz P. (2015) Brain ontogeny and life history in Pleistocene hominins. Philos Trans R Soc Lond B Biol Sci. 370(1663).
- Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, Bergmann I, Le Cabec A, Benazzi S, Harvati K, Gunz P. (2017) New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. Nature. 546, 289-292.
- Hugh C. (1911) Encyclopaedia Britannica. Solutrian Epoch. Cambridge Univ Press. 25.
- Irwin JA, Saunier JL, Niederstatter H, Strouss KM, Sturk KA, Diegoli TM, Brandstatter A, Parson W, Parsons TJ. (2009) Investigation of Heteroplasmy in the Human Mitochondrial DNA Control Region: A Synthesis of Observations from More Than 5000 Global Population Samples. JME. 68, 5:516-527.
- James S. (2005) Exploring the World of the Celts. Thames and Hudson.
- Jeunesse C. (2000) Les composantes autochtone et danubienne en Europe centrale et occidentale entre 5500 et 4000 av. JC. Les derniers chasseurs-cueilleurs d'Europe occidentale. Besançon: Presses universitaires Franc-Comtoises. 361-378.
- Jeunesse C. (2002) Armatures asymétriques, regionalisation, acculturation. Contribution à l'Étude des relations entre le Rubané et la composante autochtone dans l'ouest de la sphère danubien. Préhistoire de la Grande Plaine du Nord de l'Europe. Les échanges entre l'Est et l'Ouest dans les sociétés préhistoriques. Actes di Colloque Chaire Francqui interuniversitaire, Université de Liège, 99: 147-165.
- Jobling MA, Tyler-Smith C. (2003) The Human Y chromosome: an evolutionary marker comes of age. Nature Rev Genet. 4, 598-612.
- Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, Meshveliani T, Bar-Yosef O, Müller W, Belfer-Cohen A, Matskevich Z, Jakeli N, Higham TFG, Currat M, Lordkipanidze D, Hofreiter M, Manica A, Pinhasi R, Bradley DG. (2015) Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nature Communications. 6, 8912.
- Karmin M, Saag L, Vicente M, Wilson Sayres MA, Jarve M, Talas UG, Rootsi S, Ilumae A, Magi R, Mitt M, Pagani L, Puurand T, Faltyskova Z, Clemente F, Cardona A, Metspalu E, Sahakyan H, Yunusbayev B, Hudjashov G, DeGiorgio M, Loogvali E, Eichstaedt C, Eelmets M, Chaubey G, Tambets K, Litvinov S, Mormina M, Xue Y, Ayub Q, Zoraqi G, Korneliussen TS, Akhatova F, Lachance J, Tishkoff S, Momynaliev K, Ricaut F, Kusuma P, Razafindrazaka H, Pierron D, Cox MP, Sultana GNN, Willerslev R, Muller C, Westaway M, Lambert D, Skaro V, Kovacevic L, Turdikulova S, Dalimova D, Khusainova R, Trofimova N, Akhmetova V, Khidiyatova I, Lichman DV, Isakova J, Pocheshkhova E, Sabitov Z, Barashkov NA, Nymadawa P, Mihailov E, Tien Seng JW, Evseeva I, Bamberg Migliano A, Abdullah S, Andriadze G, Primorac D, Atramentova L, Utevska O, Yepiskoposyan L, Marjanovic D, Kushniarevich A, Behar DM, Gilissen C, Vissers L, Veltman JA, Balanovska E, Derenko M, Malyarchuk B, Metspalu A, Fedorova S, Eriksson A, Manica A, Mendez FL, Karafet TM, Veeramah KR, Bradman N, Hammer MF, Osipova LP, Balanovsky O, Khusnutdinova EK, Johnsen K, Remm M, Thomas MG, Tyler-Smith C, Underhill PA, Willerslev E, Nielsen R, Metspalu M, Villems R, Kivisild T. (2015) A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res. 25:459-466.
- Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, Leidinger P, Backes C, Khairat R, Forster M, Stade B, Franke A, Mayer J, Spangler J, McLaughlin S, Shah M, Lee C, Harkins TT, Sartori A, Moreno-Estrada A, Henn B, Sikora M, Semino O, Chiaroni J, Rootsi S, Myres NM, Cabrera VM, Underhill PA, Bustamante CD, Vigl EE, Samadelli M, Cipollini G, Haas J, Katus H, O'Connor BD, Carlson MRJ, Meder B, Blin N, Meese E, Pusch CM, Zink A. (2012) New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nature Communications. 3, 698.
- Kipfer BA. (2000) Encyclopedic Dictionary of Archaeology. Springer.

- Kircher M. (2012) Analysis of High-Throughput Ancient DNA Sequencing Data. Ancient DNA. Methods in Molecular Biology (Methods and Protocols). 840, 197-228.
- Kivisild T. (2017) The study of human Y chromosome variation through ancient DNA. Human Genetics. 136, 5:529-546.
- Köhler-Rollefson I, Gillespie W, Metzger M. (1988) The fauna from Neolithic 'Ain Ghazal. The Prehistory of Jordan. British Archaeology Reports. 396, 1:423-430.
- Kornberg A, Lehman IR, Bessman MJ, Simms ES. (1956) Enzymic synthesis of deoxyribonucleic acid. Biochim et Biophys Acta. 21, 197-198.
- Kozlowski JK. (1999) Gravettian/Epigravettian sequences in the Balkans: environment, technologies, hunting strategies and raw material procurement. British School at Athens. 3, 319-329.
- Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP, Pääbo S. (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature. 464, 894-897.
- Kreuz A, Markle T, Marinova E, Rosch M, Schafer E, Schamuhn S, Zerl T. (2014) The Late Neolithic Michelsberg culture – just ramparts and ditches? A supraregional comparison of agricultural and environmental data. Praehistorische Zeitschrift. 89.1: 72-115.
- Kuijt I, Goring-Morris N. (2002) Foraging, Farming, and Social Complexity in the Pre-Pottery Neolithic of the Southern Levant: A Review and Synthesis. Journal of World Prehistory. 16, 4:361-440.
- Larrasoaña JC, Roberts AP, Rohling EJ. (2013) Dynamics of Green Sahara Periods and Their Role in Hominin Evolution. PloS ONE. 8, 10.
- Larsen CS. (1995) Biological Changes in Human Populations with Agriculture. Annual Review of Anthropology. 24:185-213.
- Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, Berger B, Economou C, Bollongino R, Fu Q, Bos KI, Nordenfelt S, Li H, de Filippo C, Prüfer K, Sawyer S, Posth C, Haak W, Hallgren F, Fornander E, Rohland N, Delsate D, Francken M, Guinet J, Wahl J, Ayodo G, Babiker HA, Bailliet G, Balanovska E, Balanovsky O, Barrantes R, Bedoya G, Ben-Ami H, Bene J, Berrada F, Bravi CM, Brisighelli F, Busby GBJ, Cali F, Churnosov M, Cole DEC, Corach D, Damba L, van Driem G, Dryomov S, Dugoujon J, Fedorova SA, Gallego Romero I, Gubina M, Hammer M, Henn BM, Hervig T, Hodoglugil U, Jha AR, Karachanak-Yankova S, Khusainova R, Khusnutdinova E, Kittles R, Kivisild T, Klitz W, Kučinskas V, Kushniarevich A, Laredj L, Litvinov S, Loukidis T, Mahley RW, Melegh B, Metspalu E, Molina J, Mountain J, Näkkäläjärvi K, Nesheva D, Nyambo T, Osipova L, Parik J, Platonov F, Posukh O, Romano V, Rothhammer F, Rudan I, Ruizbakiev R, Sahakyan H, Sajantila A, Salas A, Starikovskaya EB, Tarekegn A, Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Vasquez R, Villena M, Voevoda M, Winkler CA, Yepiskoposyan L, Zalloua P, Zamunik T, Cooper A, Capelli C, Thomas MG, Ruiz-Linares A, Tishkoff SA, Singh L, Thangaraj K, Villems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler EE, Burger J, Slatkin M, Pääbo S, Kelso J, Reich D, Krause J. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 513, 409-413.
- Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K, Connell S, Stewardson K, Harney E, Fu Q, Gonzalez-Fortes G, Roodenberg SA, Lengyel G, Bocquentin F, Gasparian B, Monge JM, Gregg M, Eshed V, Mizrahi AS, Meiklejohn C, Gerritsen F, Bejenaru L, Blueher M, Campbell A, Cavalleri G, Comas D, Froguel P, Gilbert E, Kerr SM, Kovacs P, Krause J, McGettigan D, Merrigan M, Merriwether DA, O'Reilly S, Richards MB, Semino O, Shamoon-Pour M, Stefanescu G, Stumvoll M, Tonjes A, Torroni A, Wilson JF, Yengo L, Hovhannisyan, Patterson N, Pinhasi R, Reich D. (2016) Genomic insights into the origin of farming in the ancient Near East. Nature. 536, 419-424.
- Leblois E. (2010) La nécropole hallstattienne d'Harchies "Maison Cauchies": un bilan et des perspectives (province de Hainaut, Belgique). Lunula, Archaeologia protohistorica. 18, 107-111.
- Li H, Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25, 14:1754-1760.
- Lightowlers RN, Chrzanowska-Lightowlers ZM. (2012) Exploring our origins – the importance of OriL in mtDNA maintenance and replication. EMBO Rep. 13(12): 1038-9.
- Lindahl T. (1993) Instability and decay of the primary structure of DNA. Nature. 362, 709-715.
- Lindgreen S. (2012) AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Research Notes. 5:337.
- Lipson M, Szécsényi-Nagy A, Mallick S, Pósa A, Stégmár B, Keerl V, Rohland N, Stewardson K, Ferry M, Michel M, Oppenheimer J, Broomandkhoshbacht N, Harney E, Nordenfelt S, Llamas B, Mende BG,

Köhler K, Oross K, Bondár M, Marton T, Osztás A, Jakucs J, Paluch T, Horváth F, Csengeri P, Koós J, Sebők K, Anders A, Raczky P, Regenye J, Barna JP, Fábián S, Serlegi G, Toldi Z, Nagy EG, Dani J, Molnár E, Pálfi G, Márk L, Melegh B, Bánfai Z, Domboróczki L, Fernández-Eraso J, Mujika-Alustiza JA, Alonso Fernández C, Jiménez Echevarría J, Bollongino R, Orschiedt J, Schierhold K, Meller H, Cooper A, Burger J, Bánffy E, Alt KW, Lalueza-Fox C, Haak W, Reich D. (2017) Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature. 551, 368-372.

- Ludwig MZ. (2002) Functional evolution of noncoding DNA. Genetics and Development. 12, 6:634-639.
- Luo S, Valencia CA, Zhang J, Lee N, Slone J, Gui B, Wang X, Li Z, Dell S, Brown J, Chen SM, Chien Y, Hwu W, Fan P, Wong L, Atwal PS, Huang T. Biparental Inheritance of Mitochondrial DNA in Humans. PNAS. 115, 51:13039-13044.
- Macaulay V, Richards MB, Sykes BC. (1999) Mitochondrial DNA recombination-no need to panic. Proc Biol Sci. 266.(1433)2037-9.
- MacHugh DE, Edwards CJ, Bailey JF, Bancroft DR, Bradley DG. (2000) The Extraction and Analysis of Ancient DNA From Bone and Teeth: a Survey of Current Methodologies. Ancient Biomolecules. 3:81-102.
- Marchand G, Arthuis R, Philibert S, Sellami F, Sicard S, Forré P, Lanoë S, Nauleau JF, Quesnel L, Querré G. (2009) Un habitat azilien en Anjou : les Chaloignes à Mozé-sur-Louet (Maine-et-Loire). Gallia-Préhistoire. 51, 1-113.
- Margaryan A, Derenko M, Hovhannisyan H, Malyarchuk B, Heller R, Khachatryan Z, Avetisyan P, Badalyan R, Bobokhyan A, Melikyan V, Sargsyan G, Piliposyan A, Simonyan H, Mkrtchyan R, Denisova G, Yepiskoposyan L, Willerslev E, Allentoft ME. (2017) Eight Millenia of Matrilineal Genetic Continuity in the South Caucasus. Current Biology. 27, 13:2023-2028.
- Mariën ME. (1999) Nécropole hallstattienne à tombelles dans le " Bois de la Taille des Vignes " à Havré (Hainaut, Belgique). Archéologie des Celtes. 227-242.
- Marks AE, Bieho N, Zilhão J, Ferring R. (1994) Upper Pleistocene Prehistory in Portuguese Estremadura: Results of Preliminary Research. Journal of Field Archaeology. 21, 1:53-68.
- Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardon K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JMB, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Guerra MAR, Roodenberg J, Verges JM, Krause J, Cooper A, Alt KW, Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, Reich D. (2015) Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528, 499-503.
- Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, Olalde I, Broomandkhoshbacht N, Candilio F, Cheronet O, Fernandes D, Ferry M, Gamarra B, González Fortes G, Haak W, Harney E, Jones E, Keating D, Krause-Kyora B, Kucukkalipci I, Michel M, Mittnik A, Nägele K, Novak M, Oppenheimer J, Patterson N, Pfrengle S, Sirak K, Stewardson K, Vai S, Alexandrov S, Alt KW, Andreescu R, Antonović D, Ash A, Atanassova N, Bacvarov K, Balázs Gustáv M, Bocherens H, Bolus M, Boroneant A, Boyadzhiev Y, Budnik A, Burmaz J, Chohadzhiev S, Conard NJ, Cottiaux R, Čuka M, Cupillard C, Drucker DG, Elenski N, Francken M, Galabova B, Ganetsovski G, Gély B, Hajdu T, Handzhyiska V, Harvati K, Higham T, Iliev S, Janković I, Karavanić I, Kennett DJ, Komšo D, Kozak A, Labuda D, Lari M, Lazar C, Leppek M, Leshtakov K, Lo Vetro D, Los D, Lozanov I, Malina M, Martini F, McSweeney K, Meller H, Menđušić M, Mirea P, Moiseyev V, Petrova V, Price TD, Simalcsik A, Sineo L, Šlaus M, Slavchev V, Stanev P, Starović A, Szeniczey T, Talamo S, Teschler-Nicola M, Thevenet C, Valchev I, Valentin F, Vasilyev S, Veljanovska F, Venelinova S, Veselovskaya E, Viola B, Virag C, Zaninović J, Zäuner S, Stockhammer PW, Catalano G, Krauß R, Caramelli D, Zarina G, Gaydarska B, Lillie M, Nikitin AG, Potekhina I, Papathanasiou A, Borić D, Bonsall C, Krause J, Pinhasi R, Reich D. (2018) The genomic history of southeastern Europe. Nature. 555, 197-203.
- Mayr E. (1982) The Growth of biological thought. The Belknap Press of Harvard University Press.
- McDougall I, Brown FH, Fleagle JG. (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature. 433, 733-736.
- Mellars P. (2006) Archeology and the dispersal of modern humans in Europe: Deconstructing the "Aurignacian". Evolutionary Anthropology. 15:167-182.
- Mendez FL, Krahn T, Schrack B, Krahn A, Veeramah KR, Woerner AE, Fomine FLM, Bradman N, Thomas MG, Karafet TM, Hammer MF. (2013) An African American Paternal Lineage Adds an Extremely Ancient Root to the Human Y Chromosome Phylogenetic Tree. AJHG. 92, 3:454-459.

- Menozzi P, Piazza A, Cavalli-Sforza L. (1978) Synthetic maps of human gene frequencies in Europeans. Science. 201, 4358:786-792.
- Meyer M, Kircher M. (2010) Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. CSH Protocols. 5.
- Miller R, Zwyns N, Otte M, Stevens C, Stewart J. (2012) La séquence mésolithique et néolithique du Trou Al'Wesse (Belgique): résultats pluridisciplinaires. L'anthropologie. 116, 99-126.
- Monnat RJ, Maxwell CL, Loeb LA. (1985) Nucleotide Sequence Preservation of Human Leukemic Mitochondrial DNA. AACR. 45, 4:1809-1814.
- Mullis KF, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. (1986) Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. Cold Spring Harbor Symposium in Quantitative Biology. 51:263-273.
- Neustupny E. (1969) Economy of the Corded Ware Cultures. Archeologické rozhledy. 21, 43-68.
- Noe-Nygaard N. (1989) Man-made trace fossils on bones. Human Evolution. 4, 6:461-491.
- Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. (2008) Genes mirror geography within Europe. Nature. 456, 98-101.
- O'Leary MH. (1988) Carbon Isotopes in Photosynthesis. BioScience. 38. 5:328-336.
- Okonechnikov K, Conesa A, GarcÍa-Alcade F. (2015) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 32, 2:292-294.
- Olalde I, Allentoft ME, Sanchez-Quinto F, Santpere G, Chiang CWK, DeGiorgio M, Prado-Martinez J, Rodriguez JA, Rasmussen S, Quilez J, Ramirez O, Marigorta UM, Fernandez-Callejo M, Prada ME, Vidal Encinas JM, Nielsen R, Netea MG, Novembre J, Sturm RA, Sabeti P, Marques-Bonet T, Navarro A, Willerslev E, Lalueza-Fox C. (2014) Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature. 507, 225-228.
- Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, Rohland N, Makkick S, Szécsényi-Nagy A, Mittnik A, Altena E, Lipson M, Lazaridis I, Harper TK, Patterson N, Broomandkhoshbacht N, Diekmann Y, Faltyskova Z, Fernandes D, Ferry M, Harney E, de Knijff P, Michel M, Oppenheimer J, Stewardson K, Barclay A, Alt KW, Liesau C, Ríos P, Blasco C, Vega Miguel J, Menduiña García R, Avilés Fernández A, Bánffy E, Bernabò-Brea M, Billoin D, Bonsall C, Bonsall L, Allen T, Büster L, Carver S, Castells Navarro L, Craig OE, Cook GT, Cunliffe B, Denaire A, Egging Dinwiddy K, Dodwell N, Ernée M, Evans C, Kuchařík M, Farré JF, Fowler C, Gazenbeek M, Garrido Pena R, Haber-Uriarte M, Haduch E, Hey G, Jowett N, Knowles T, Massy K, Pfrengle S, Lefranc P, Lemercier O, Lefebvre A, Heras Martínez C, Galera Olmo V, Bastida Ramírez A, Lomba Maurandi J, Majó T, McKinley JI, McSweeney K, Mende BG, Modi A, Kulcsár G, Kiss V, Czene A, Patay R, Endrődi A, Köhler K, Hajdu T, Szeniczey T, Dani J, Bernert Z, Hoole M, Cheronet O, Keating D, Velemínský P, Dobeš M, Candilio F, Brown F, Flores Fernández R, Herrero-Coral A, Tusa S, Carnieri E, Lentini L, Valenti A, Zanini A, Waddington C, Delibes G, Guerra-Doce E, Neil B, Brittain M, Luke M, Mortimer R, Desideri J, Besse M, Brücken G, Furmanek M, Hałuszko A, Mackiewicz M, Rapiński A, Leach S, Soriano I, Lillios KT, Cardoso JL, Parker Pearson M, Włodarczak P, Price TD, Prieto P, Rey P, Risch R, Rojo Guerra MA, Schmitt A, Serralongue J, Silva AM, Smrčka V, Vergnaud L, Zilhão J, Caramelli D, Higham T, Thomas MG, Kennett DJ, Fokkens H, Heyd V, Sheridan A, Sjögren K, Stockhammer PW, Krause J, Pinhasi R, Haak W, Barnes I, Lalueza-Fox C, Reich D. (2018) The Beaker phenomenon and the genomic transformation of northwest Europe. Nature. 555, 190-196.
- Olalde I, Mallick S, Patterson N, Rohland N, Villalba-Mouco V, Silva M, Dulias K, Edwards CJ, Gandini F, Pala M, Soares P, Ferrando-Bernal M, Adamski N, Broomandkhoshbacht N, Cheronet O, Culleton BJ, Fernandes D, Lawson AM, Mah M, Oppenheimer J, Stewardson K, Zhang Z, Jiménez Arenas JM, Toro Moyano IJ, Salazar-García DC, Castanyer P, Santos M, Tremoleda J, Lozano M, García Borja P, Fernández-Eraso J, Mujika-Alustiza JA, Barroso C, Bermúdez FJ, Viguera Mínguez E, Burch J, Coromina N, Vivó D, Cebrià A, Fullola JM, GarcÍa-Puchol O, Morales JI, Oms FX, Majó T, Vergès JM, Díaz-Carvajal A, Ollich-Castanyer I, López-Cachero FJ, Silva AM, Alonso-Fernández C, Delibes de Castro G, Jiménez EchevarrÍa J, Moreno-Márquez A, Berlanga GP, Ramos-García P, Ramos-Muñoz J, Vijande Vila E, Aguilella Arzo G, Esparza Arroyo A, Lillios KT, Mack J, Velasco-Vázquez J, Waterman A, Benítez de Lugo Enrich L, Sánchez MB, Agusti B, Codina F, de Prado G, Estalrrich A, Fernández Flores A, Finlayson C, Finlayson G, Finlayson S, Giles-Guzmán F, Rosas Am Barciela González V, GarcÍa Atiénzar G, Hernández Pérez MS, Llanos A, Carrión Marco Y, Collado Beneyto I, López-Serrano D, Sanz Tormo M, Valera AC, Blasco C, Liesau C, Ríos P, Daura J, de Pedro Michó MJ, Diez-Castillo AA, Flores Fernández R, Farré JF, Garrido-Pena R, Gonçalves VS, Guerra-Doce E, Herrero-Corral AM, Juan-Cabanilles J, López-Reyes D,

McClure SB, Merino Pérez M, Foix AO, Sanz Borràs M, Sousa AC, Vidal Encinas JM, Kennett DJ, Richards MB, Alt KW, Haak W, Pinhasi R, Lalueza-Fox C, Reich D. (2019) The genomic history of the Iberian Peninsula over the past 8000 years. Science. 363, 6432:1230-1234.

- Olenkovskiy M. (2010) The Eastern Epigravettian in the North Azov region (Ukraine). Atti Soc Preist Protost Friuli-VG, Trieste. 7-26.
- Olson D, DellaSala DA, Noss RF, Strittholt JR, Kass J, Koopman ME, Allnutt TF. (2012) Climate Change Refugia for Biodiversity in the Klamath-Siskiyou Ecoregion. Natural Areas Journal. 32, 1:65-74.
- Orlando L, Gilbert MTP, Willerslev E. (2015) Reconstructing ancient genomes and epigenomes. Nature Reviews Genetics. 16, 395-408.
- Ottoni C, Ricaut FX, Vanderheyden N, Brucato N, Waelkens M, Decorte R. (2011) Mitochondrial analysis of a Byzantine population reveals the differential impact of multiple historical events in South Anatolia. EJHG. 19, 571-576.
- Özdoğan M. (2011) Archaeological Evidence on the Westward Expansion of Farming Communities from Eastern Anatolia to the Aegean and the Balkans. Current Anthropology. 52, 4.
- Pääbo S. (1985) Molecular cloning of Ancient Egyptian mummy DNA. Nature. 314, 644-645.
- Pääbo S. (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. PNAS. 86, 6:1939-1943.
- Pääbo S, Wilson AC. (1991) Miocene DNA sequences – a dream come true? Current Biology. 1, 1:45-46.
- Parton A, Farrant AR, Leng MJ, Telfer MW, Groucutt HS, Petraglia MD, Parker AG. (2015) Alluvial fan records from southeast Arabia reveal multiple windows for human dispersal. Geology. 43, 4.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. (2012) Ancient Admixture in Human History. Genetics. 192. 3:1065-1093.
- Pereira L, Richards M, Goios A, Alonso A, Albarran C, Garcia O, Behar DM, Golge M, Hatina J, Al-Gazali L, Bradley DG, Macaulay V, Amorim A. (2005) High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Res. 15, 19-24.
- Pereira JB, Costa MD, Vieira D, Pala M, Bamford L, Harich N, Cherni L, Alshamali F, Hatina J, Rychkov S, Stefanescu G, King T, Torroni A, Soares P, Pereira L, Richards MB. (2017) Reconciling evidence from ancient and contemporary genomes: a major source for the European Neolithic within Mediterranean Europe. Proc R Soc B. 284: 20161976.
- Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, Alpasian-Roodenberg S, Gerritsen F, Moiseyev V, Gromov A, Raczky P, Anders A, Pietrusewsky M, Rollefson G, Jovanovic M, Trinhhoang H, Bar-Oz G, Oxenham M, Matsumura H, Hofreiter M. (2015) Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. PloS One. 10, 6.
- Pirson S, Flas D, Abrams G, Bonjean D, Court-Picon M, Di Modica K, Draily C, Damblon F, Haesaerts P, Miller R, Rougier H, Toussaint M, Semal P. (2012) Chronostratigraphic context of the Middle to Upper Palaeolithic transition: Recent data from Belgium. Quaternary International. 259, 78-94.
- Poznik GD, Henn BM, Yee M, Sliwerska E, Euskirchen GM, Lin AA, Snyder M, Quintana-Murci L, Kidd JM, Underhill PA, Bustamante CD. (2013) Sequencing Y Chromosome Resolves Discrepancy in Time to Common Ancestor of Males Versus Females. Science. 341, 6145:562-565.
- Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, Ayub Q, McCarthy SA, Narechania A, Kashin S, Chen Y, Banerjee R, Rodriguez-Flores JL, Cerezo M, Shao H, Gymrek M, Malhotra A, Louzada S, Desalle R, Ritchie GR, Cerveira E, Fitzgerald TW, Garrison E, Marcketta A, Mittelman D, Romanovitch M, Zhang C, Cheng-Bradley X, Abecasis GR, McCarroll SA, Flicek P, Underhill PA, Coin L, Zerbino DR, Yang F, Lee C, Clarke L, Auton A, Erlich Y, Handsaker RE, 1000 Genomes Project Consortium, Bustamante CD, Tyler-Smith C. (2016) Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat Genet. 48.(6):593-9.
- Price TD. (2000) Europe's First Farmers. Cambridge University Press, Cambridge.
- Price AL, Tandon A, Patterson N. (2009) Sensitive Detection of Chromosomal segments of Distinctive Ancestry in admixed populations. Plos genet. 5.
- Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Orlando L, Metspalu E, Karmin M, Tambets K, Rootsi S, Mägi R, Campos PF, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova LP, Fedorova SA, Voevoda MI, DeGiorgio M, Sicheritz-Ponten T, Brunak S, Demeshchenko S, Kivisild T, Villems R, Nielsen R, Jakobsson M, Willerslev E. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 505. 87-91.

- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MTP, Wang Y, Raghavan M, Campos PF, Munkholm Kamp H, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grønnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TFG, Bronk Ramsey C, Hansen TO, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature. 463, 757-762.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calí F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A, Bandelt HJ. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. AJHG. 67, 1251-1276.
- Richards MP, Schulting RJ, Hedges REM. (2003) Sharp shift in diet at onset of Neolithic. Nature. 425, 366.
- Renaud G, Stenzel U, Kelso J. (2014) leeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acid Research. 42, 18:141.
- Rootsi S, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barać L, Peričić M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioğlu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan J, Chaventre A, Ferak V, Füredi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzić R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Magri C, Semino O. (2004) Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe. AJHG. 75, 1:128-137.
- Rougier H, Crevecoeur I, Beauval C, Flas D, Posth C, Wissing C, Furtwängler A, Germonpré M, Gomez-Olivencia A, Semal P, van der Plicht J, Bocherens H, Krause J. (2016) La Troisième caverne de Goyet en Belgique : un site exceptionnel ayant livré des restes néandertaliens et des restes humains modernes du Paléolithique supérieur. FNRS.
- Rowley-Conwy P. (2011) Westward Ho! The spread of agriculture from Central Europe to the Atlantic. Curr. Anthropol. 52, 431-451.
- Saillard J, Magalhães PJ, Schwartz M, Rosenberg T, Nørby S. (2000) Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV. Hum Biol 72:1065-1068.
- Samuels DC, Wonnapinij P, Cree L, Chinnery PF. (2010) Reassessing evidence for a postnatal mitochondrial genetic bottleneck. Nature Genetics. 42, 6:471-472.
- Sanchez-Quinto F, Schroeder H, Ramirez O, Avila-Arcos MC, Pybus M, Olalde I, Velazquez AM, Marcos ME, Encinas JM, Bertranpetit J, Orlando L, Gilbert MT, Lalueza-Fox C. (2012) Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. Curr. Biol. 22, 1494-1499.
- Sanger F, Coulson AR. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 94:441-448.
- Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M. (1977) Nucleotide sequence of bacteriophage phi X174 DNA. Nature. 265:687-695.
- Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rashied KA, Willerslev E, Krogh A, Orlando L. (2012) Improving ancient DNA read mapping against modern reference genomes. BMC Genomics. 13:178.
- Schulting RJ. (1998) Slighting the sea: Stable isotopes evidence for the transition to farming in northwestern Europe. Documenta Praehistorica. 25, 203-218.
- Schwartz M, Vissing J. (2002) Paternal Inheritance of Mitochondrial DNA. N Engl J Med. 347:576-580.
- Schwarz C, Debruyne R, Kuch M, McNally E, Schwarcz H, Aubrey AD, Bada J, Poinar H. (2009) New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Research. 37, 10:3215-3229.
- Schwendler R. (2012) Diversity in social organization across Magdalenian Western Europe ca. 17-12,000 BP. Quaternary International, ScienceDirect. 272-273, 333-353.
- Scozzari R, Massaia A, D'Atanasio E, Myres NM, Perego UA, Trombetta B, Cruciani F. (2012) Molecular Dissection of the Basal Clades in the Human Y Chromosome Phylogenetic Tree. PloS ONE. 7, 11.

- Scozzari R, Massaia A, Trombetta B, Bellusci G, Myres NM, Novelletto A, Cruciani F. (2014) An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. Genome Res. 24:535-544.
- Semal P, Rougier H, Crevecoeur I, Jungels C, Flas D, Hauzeur A, Maureille B, Germonpré M, Bocherens H, Pirson S, Cammaert L, De Clerck N, Hambucken A, Higham T, Toussaint M, van der Plicht J. (2009) New data on the late Neandertals: Direct dating of the Belgian Spy fossils. AJPA.
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. (2000) The Genetic Legacy of Paleolithic *Homo sapiens sapiens* in Extant Europeans: A Y Chromosome Perspective. Science. 290. 1155-1159.
- Skoglund P, Storå J, Götherström A, Jakobsson M. (2013) Accurate sex identification of ancient human remains using DNA shotgun sequencing. 40, 12:4477-4482.
- Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, Günther T, Hall P, Tambets K, Parik J, Sjögren K, Apel J, Willerslev E, Storå J, Götherström A, Jakobsson M. (2014) Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. Science. 344, 6185:747-750.
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou S, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang S, Waterston RH, Wilson RK, Rozen S, Page DC. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. 423, 825-837.
- Sklenář K. (1983) Archaeology in Central Europe: the first 500 years. Leicester Univ Press.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. AJHG. 6, 84:740-759.
- Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt HJ, Torroni A, Richards MB. (2010) The Archaeogenetics of Europe. Current Biology. 20, 4:174-183.
- Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilova E, Macaulay V, Richards MB, Cerny V, Pereira L. (2012) The Expansion of mtDNA Haplogroup L3 within and out of Africa. Molecular Biology and Evolution. 29, 3:915-927.
- Soffer O. (1993) Upper Paleolithic Adaptations in Central and Eastern Europe and Man-Mammoth Interactions. Springer Science. 31-49.
- Srejovic D. (1972) Europe's First Monumental Sculpture: New Discoveries at Lepenski Vir. Thames and Hudson.
- Stanko VN. (2007) Fluctuations in the level of the Black Sea and Mesolithic settlement of the northern Pontic area. The Black Sea Flood Question: Changes in Coastline, Climate, and Human Settlement. Springer. 371-385.
- Stika HP. (1996) Cultivated plant remains of the late Neolithic Michelsberg Culture at Heilbronn-Klingenberg (southwest Germany)-a comparison of different features, find assemblages and preservation conditions relating to the representation of archaeobotanical remains. Vegetation history and archaeobotany. 5. 1-2: 57-64.
- Stringer C, Galway-Witham J. (2017) On the origin of our species. Nature. 546, 212-214.
- Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G. (2000) Ubiquitinated Sperm Mitochondria, Selective Proteolysis, and the Regulation of Mitochondrial Inheritance in Mammalian Embryos. Biology of Reproduction. 63, 2:582-590.
- Templeton AR, Routman E, Phillips CA. (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. Genetics. 140:767-782.
- Timmermann A, Friedrich T. (2016) Late Pleistocene climate drivers of early human migration. Nature. 538, 92-95.
- Torroni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC. (1993) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet. 53, 3:591-608.
- Torroni A, Achilli A, Macaulay V, Richards MB, Bandelt H. (2006) Harvesting the fruit of the human mtDNA tree. Trends in Genetics. 22, 6:339-345.

- Toussaint M. (2002) Problematique chronologique des sepultures du mesolithique mosan en milieu karstique. Notae Praehistorica. 22:141-66.
- Toussaint M. (2006) 1997-2005 Research in the caves of Goyet (Gesves, province of Namur, Belgium). Tongeren Neandertal symposium excursion.
- Trinkaus E, Moldovan O, Milota Ş, Bîlgăr A, Sarcina L, Athreya S, Bailey SE, Rodrigo R, Mircea G, Higham T, Bronk Ramsey C, van der Plicht J. (2003) An early modern human from the Peştera cu Oase, Romania. PNAS. 100, 20:11231-11236.
- Trinkaus E. (2006) Modern Human versus Neandertal Evolutionary Distinctiveness. Current Anthropology. 47, 4:597-620.
- Tykot RH. (2006) Isotope Analysis and the Histories of Maize. Histories of Maize in Mesoamerica: Multidisciplinary Approaches. 131-142.
- Van Oven M, Kayser M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 30, 2: 386-394.
- Wang DG, Fan J, Siao C, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES. (1998) Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. Science. 280, 5366:1077-1082.
- Watson JD, Crick FHC. (1953) A Structure for Deoxyribose Nucleic Acid. Nature. 171, 737-738.
- Wei W, Ayub Q, Chen Y, McCarthy S, Hou Y, Carbone I, Xue Y, Tyler-Smith C. (2013) A calibrated human Y-chromosomal phylogeny based on resequencing. Genome Res. 23:388-395.
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF. (2001) The Eurasian Heartland: A continental perspective on Y-chromosome diversity. PNAS. 98, 18:10244-10249.
- Whittle AWR. (1996) Europe in the Neolithic: The Creation of New Worlds. Cambridge University Press, Cambridge.
- Whittle AWR, Cummings V. (2007) Going Over: The Mesolithic-Neolithic Transition in North-West Europe. Oxford University Press, Oxford.
- Wood B. (2011) Did early *Homo* migrate “out of” or “in to” Africa? PNAS. 108, 26:10375-10376.
- Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR. (1998) Improved DNA extraction from ancient bones using silica-based spin columns. AJPA. 105, 4:539-543.
- Zhou H, Alexander DH, Lange K. (2009) A quasi-Newton method for accelerating the convergence of iterative optimization algorithms. Statistics and Computing.
- Zilhao J. (2000) From the Mesolithic to the Neolithic in the Iberian Peninsula. Europe’s first farmers. Cambridge Univ Press.
- Zimmermann WH. (2002) Kontinuität und Wandel im Hausbau südlich und östlich der Nordsee vom Neolithikum bis zum Mittelalter. The Rural House from the migration period to the oldest still standing buildings. Prag. 164-168.
- Zvelebil M, Rowley-Conwy P. (1984) Transition to farming in Northern Europe: A hunter-gatherer perspective. Norwegian Archaeological Review. 17, 2:104-128.
- Zvelebil M. (1986) Foragers and farmers in Atlantic Europe. Hunters in transition, Mesolithic societies of temperate Eurasia and their transition to farming. Cambridge University Press. 67-93.
- Zvelebil M. (1998) Agricultural frontier, Neolithic origins, and the transition to farming in the Baltic basin. Harvesting the sea, farming the forest, the emergence of Neolithic societies in the Baltic region. Sheffield Academic Press. 9-27.

7. APPENDIX

Appendix 1. Trimming and merging scripts using leeHom, followed by the description of the options used in the command.

```
~/path/to/leeHom –fq1 *.fastq.gz –fq2 *.fastq.gz –fqo *_leeHom_merged –ancientdna -f AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -s AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT –log *_leeHom.log
```

Option	Description
-fq1	First input file to be merged
-fq2	Second input file to be merged
-fqo	Output file
-f	Forward adapter sequence
-s	Reverse adapter sequence
-log	Creates a report file

Appendix 2. Scripts used to perform the mapping step against a reference genome with BWA, followed by the description of the options used in the commands. For the algorithm to work, BWA has to construct the index for the reference genome using the command bwa index.

```
bwa index /your/pathway/reference.fasta
```

```
bwa aln -n 0.01 -o 2 -l 16569 -t 6 /your/pathway/reference.fasta *.fq.gz > *_aligned.sai
```

```
bwa samse /your/pathway/reference.fasta *_aligned.sai *.fq.gz > *_aligned.sam
```

Option	Description
-n	Maximum number of mismatches allowed between the read and the reference
-o	Maximum number of gaps
-l	Random number, seed disabled
-t	Multi-threading mode

Appendix 3. Mapping quality filtering and sorting scripts using Samtools, followed by the description of the options used in the commands.

```
samtools view -bF 4 -q 20 *_aligned.sam > *_filtered.bam
```

```
samtools sort file.bam > sorted_file.bam
```

Option	Description
view	Prints the alignments of the specific input file (.sam, .bam or .cram)
sort	Sort the reads by chromosome rather by read number

-b	Set the output file in .bam format
-F 4	Excludes unmapped reads in the output alignments
-q	Excludes alignments with mapping quality below the desired value

Appendix 4. Marking duplicates script using Picard, followed by the description of the options used in the commands.

java -jar /your/pathway/picard.jar MarkDuplicates INPUT=*_sorted.bam OUTPUT=*_dedup.bam METRICS_FILE=metrics.txt

Option	Description
-jar	Allows the software to run using a .jar file
-INPUT	One or more input files, .sam or .bam
-OUTPUT	The output file will be generated
-METRICS_FILE	Report of the command in .txt format

Appendix 5. Read length filtering script using Samtools, followed by the description of the options used in the command.

samtools view -@6 -h *_dedup.bam | awk 'length($10) > 30 || $1 ~ /^@/' | samtools view -@6 -b > *_above30bp.bam

Option	Description
-view	Prints the alignments of the specific input file (.sam, .bam or .cram)
-@	Number of threads used in the analysis
-h	Includes the header of the file in the output
-b	Set the output file in .bam format

Appendix 6. Script used to obtain mapdamage report (mapdamage 2.0).

~/path/to/mapDamage –i *_sorted_file.bam –r /your/pathway/reference.fasta –d *_mapdamage

Option	Description
-i	Input file
-r	Reference sequence
-d	Output directory

Appendix 7. Script used to soft-clip the reads using the software bamUtil-master and the description of the options used in the command.

~/path/to/bamUtil-master/bin/bam trimBam input.bam output.bam –L X –R X --clip

Option	Description
-L X	Number of X bases to clip from the left end of the reads
-R X	Number of X bases to clip from the right end of the reads
--clip	The bases will be flagged instead of deleted (soft-clipping)

Appendix 8. Qualimap report script and the description of the options used in the command.

~/your/pathway/qualimap_v2.2.1/qualimap bamqc -bam *_above30bp.bam -outdir *_qualimap_results

Option	Description
-bam	Uses an input file in .bam format
-outdir	Creates an output directory

Appendix 9. Script used to merge bam files, followed by the description of the options used in the command.

java -jar ~/your/pathway/picard.jar MergeSamFiles I=input_1.bam I=input_2.bam O=*_merged.bam

Option	Description
-jar	Allows the software to run using a .jar file
-MergeSamFiles	Merges multiple .sam or .bam files in one file
-I	Input file in .sam or .bam format
-O	Output file in .sam or .bam format

Appendix 10. SNP calling script, followed by the description of the options used in the command.

```
~/path/to/GATK –T Pileup –R reference.fasta –I input.bam –o output –L intervallist.bed
```

Option	Description
-T	Which program inside GATK to use
-R	Reference in fasta format
-I	Input file
-o	Output
-L	Interval list of chromosome positions in bed format

Appendix 11. Command used to run PCA using smartpca and the description of the option applied.

~/path/to/smartpca -p ~/path/to/parameters_smartpca

Option	Description
-p	Set the path for a parameter file

Appendix 12. Script used to run ADMIXTURE analysis, followed by the description of the options used in the command.

for n in {2..15}; do ~/path/to/admixture –cv -s time ~/path/to/file.bed $n -j20 | tee ~/path/to/admixture_output_$n; done

Option	Description
-cv	Activates the cross-validation procedure
-s time	Each run will have a different seed since starting at different time of the day.
-j	Multi-thread option

Appendix 13. Command used to run IBD test with plink. Plink requires the input files in binary format (.bed, .bim and .fam).

```
~/path/to/plink --bfile <input> --Z-genome
```

Option	Description
--bfile	Enables input file as binary format
--Z-genome	The output file is compressed in a .gz file

Appendix 14. Command used to run D-statistics test. The parameter file (.txt) contains the path for the input files and for the output files.

```
~/path/to/qpDstat -p Dstat_param > Dstat.log
```

Option	Description
-p	Set the path for a parameter file

Appendix 15. List of Eurasian populations used for the Principal Component Analysis (PCA) with the corresponding number of individuals. Subset of the Human Origins dataset (Lazaridis et al. 2014).

Population ID	Number of individuals
Abkhasian	9
Adygei	16
Albanian	6
Armenian	10
Assyrian	11
Balkar	10
Basque	29
Bedouin	44
Belarusian	10

Bulgarian	10
Chechen	9
Croatian	10
Cypriot	8
Czech	10
Druze	39
English	10
Estonian	10
Finnish	7
French	61
Georgian	10
Greek	20
Hungarian	20
Icelandic	12
Iranian	46
Italian	25
Jewish (Ashkenazi, Georgian, Iranian, Iraqi, Libyan, Moroccan, Tunisian, Turkish and Yemenite)	67
Jordanian	9
Kumyk	8
Lebanese	28
Lezgin	9
Lithuanian	10
Maltese	8
Mordovian	10
North_Ossetian	10
Norwegian	11
Orcadian	13
Palestinian	38
Romanian	10
Russian	22
Saami	1
Sardinian	27

Saudi	8
Scottish	4
Sicilian	11
Spanish	58
Syrian	8
Turkish	56
Ukrainian	9
Yemeni	6

Appendix 16. List of Belgian individuals showing the relative ID, mtDNA coverage, mtDNA haplogroup classification and haplotype obtained mapping the genomes against rCRS reference. The mtDNA haplotypes have been manually checked and annotated loading the bam files on IGV (Integrative Genomic Viewer).

ID	MtDNA coverage	MtDNA haplogroup	MtDNA haplotype (rCRS)
AF001	90.1284	U5a2b4	73 263 750 1438 2706 3197 4769 7028 8860 9477 9548 11467 11719 12308 12372 13617 14766 14793 15301 15326 16192 16256 16270 16526
AF002	21.4299	U5b1+16189	73 150 263 750 1438 2706 3197 4769 5656 7028 7768 8860 9477 11467 11719 12308 12372 13617 13708 14182 14766 15326 16189 16192 16270
AF003	24.4952	K1a4a1	73 263 497 750 1189 1438 1811 2706 3480 4769 6260 7028 8860 9055 9698 10398 10550 11299 11467 11485 11719 11840 12308 12372 13740 14167 14766 14798 15326 16224 16311
AF004	11.5759	H4a1a1a	73 263 750 1438 3992 4024 4769 5004 8269 8860 9123 10044 14365 14582 15326
AF005	1.1557	H4a1a1a?	73 5004 9123 10044 14365 15326
AF006	91.2208	H3k1	152 263 750 1438 4769 6776 8860 11590 14687 15326
AF007	66.532	K1a4a1	73 263 497 750 1189 1438 1811 2706 3480 4769 6260 7028 8860 9055 9698 10398 10550 11299 11467 11485 11719 11840 12308 12372 13740 14167 14766 14798 15326 16224 16311
AF008	4.8254	U5b2b2	73 150 263 750 1438 1721 2706 3197 4616 4769 7028 7768 8027 8860 9477 11467 11653 11719 12308 12372 12634 13617 13630 13637 14182 14484 14766 15326 16270
AF009	24.3182	H	263 750 1438 4769 8860 15326
AF010	612.4758	U5a2b4	73 263 750 1438 2706 3197 4769 7028 8860 9477 9548 11467 11719 12308 12372 13617 14766 14793 15301 15326 16192 16256 16270 16526
AF011	335.0626	J1c3g	73 185 228 263 295 462 489 750 1438 2706 3010 4216 4769 7028 8860 9755 10398 11251 11719 11914 12612 13692 13708 13934 14022 14766 14798 15326 15452 16069 16126
AF012	136.9571	H1q+16188	263 750 1438 3010 4769 4859 8860 15326 16188
AF013	40.2859	H4a1a1a	73 263 750 1438 3992 4024 4769 5004 8269 8860 9123 10044 14365 14582 15326
AF014	40.6939	J1c5	73 228 263 295 462 489 709 750 1438 2706 3010 4216 4769 5198 7028 8860 10398 11251 11719 12612 13708 14766 14798 15326 15452 16069 16126
AF015	215.3144	K1a4a1	73 263 497 750 1189 1438 1811 2706 3480 4769 6260 7028 8860 9055 9698 10398 10550 11299 11467 11485 11719 11840 12308 12372 13740 14167 14766 14798 15326 16224 16311
AF016	16.8026	H4a1a1a	73 263 750 1438 3992 4024 4769 5004 8269 8860 9123 10044 14365 14582 15326
AF017	15.482	H1	263 723 750 1438 3010 4769 8860 15326 16093
AF018	8.4689	U5b2b2	73 150 263 750 1438 1721 2706 3197 4616 4769 7028 7768 8027 8860 9477 11467 11653 11719 12308 12372 12634 13617 13630 13637 14182 14766 15326 16270
AF020	22.4802	T1a1n	73 152 195 263 709 750 1438 1888 2706 4216 4769 4917 7028 7269 8697 8860 9899

			10463 11251 11719 12633 13368 14766 14905 15326 15452 15607 15928 16126 16163 16186 16189 16294 16304 16318
AF022	14.8472	H	263 750 1438 4769 8860 15326
AF023	15.9831	H1e2	263 750 1438 3010 4769 5460 8860 15326 15817
AF024	7.1164	H1c	263 477 750 1438 3010 4769 8860 15326
AF025	33.7931	T1a1n	73 152 195 263 709 750 1438 1888 2706 4216 4769 4917 7028 7269 8697 8860 9899 10463 11251 11719 12633 13368 14766 14905 15326 15452 15607 15928 16126 16163 16186 16189 16294 16304 16318
AF026	19.5213	K1a3a	73 263 497 750 1189 1438 1811 2706 3480 4769 7028 7559 8860 9055 9698 10398 10550 11299 11467 11719 12308 12372 13117 14167 14766 14798 15326 16093 16224 16311
AF028	11.305	K1a4a1	73 235 263 497 750 1189 1438 2706 3480 4769 6260 7028 8860 9055 9698 10398 10550 11299 11467 11485 11719 11840 12308 12372 13740 14167 14766 14798 15326 16224 16311
AF029	44.4747	U5a2a1	73 195 263 750 1438 2706 3197 4769 6641 7028 8860 9477 11467 11719 12308 12372 13617 13827 13928 14766 14793 15326 16114 16192 16256 16270 16294 16526
AF030	33.4928	H4a1a1a1	73 263 750 1438 3992 4024 4769 5004 8269 8860 9123 10034 10044 14365 14582 15326
AF031	149.7546	H3an	251 263 750 1438 4769 6776 8860 8975 15326
AF032	58.0474	H1q+16188	263 750 1438 3010 4769 4859 8860 15326 16188
AF033	76.2689	U5b2b2	73 150 263 750 1438 1721 2706 3197 4616 4769 7028 7768 8027 8860 9477 11467 11653 11719 12308 12372 12634 13617 13630 13637 14182 14766 15326 16270
AF034	559.7559	H7d	152 263 750 1438 4769 4793 8860 15326 15409
AF036	276.7217	J1c5	73 228 263 295 462 489 709 750 1438 2706 3010 4216 4769 5198 7028 8860 10398 11251 11719 12612 13708 14766 14798 15326 15452 16069 16126
AF037	205.2883	U5b2b3	73 150 263 517 750 1438 1721 2706 2755 3197 4769 7028 7768 8860 9477 9707 11467 11653 11719 12308 12372 12634 13617 13630 13637 14180 14182 14766 15326 15905 16224 16270