# Genetic structure and trait estimation in ancient Europeans

Bruno Ariano

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

A thesis submitted to Trinity College Dublin for the degree of Doctor or Philosophy

2023

Smurfit Institute of Genetics, Trinity College Dublin, Ireland

# List of abbreviations

| | |
|---|---|
| Telomere-to-Telomere | T2T |
| before present | BP |
| polymerase chain reaction | PCR |
| base pair | bp |
| Next Generation Sequencing | NGS |
| uracil–DNA–glycosylase | UDG |
| Whole Genome Sequence | WGS |
| Principal Component Analysis | PCA |
| Linkage Disequilibrium | LD |
| anatomical modern human | AMH |
| last glacial maximum | LGM |
| western hunter-gatherer | WHG |
| eastern hunter-gatherer | EHG |
| caucasus hunter-gatherer | CHG |
| Linearbandkeramik | LBK |
| Uracil-DNA-glycosylase | UDG |
| single-end | SE |

| | |
|---|---|
| paired-end | PE |
| Revised Cambridge Reference Sequence | rCRS |
| International Society of Genetic Genealogy | ISOGG |
| Simon Genome Diversity Project | SGDP |
| identical by descent | IBD |
| variant calling format | VCF |
| Early European Farmers | EEF |
| North Western Anatolia Neolithic | NW-A-Neolithic |
| run of homozygosity | ROH |
| centiMorgans | cM |
| homozygous by descent | HBD |
| hidden markov model | HMM |
| Irish Centre for High-End Computing | ICHEC |
| estimated effective migration surface | EEMS |
| genome wide association study | GWAS |
| conditional and joint analysis | COJO |

| | |
|---|---|
| polygenic risk score | PRS |
| body mass index | BMI |


## Abstract

The study of ancient human populations have mostly been carried out, until recently, by historians through the study of written records and ancient manufacts. Thanks to recent technological and theoretical advancements in genetics, it is now possible to support these studies and uncover new insights into human prehistory. The field of ancient genomics is in a phase of continuous progress and recently has become able to describe phenomena such as fine-population structure, phenotypic traits and demographic analyses. By investigating these features of a large set of ancient individuals I will address in this thesis events that shaped the structure, and traits of ancient European populations.

In the first chapter I present the analysis of novel genomic data derived from three ancient Maltese. In this analysis I describe how these individuals were very similar to other Neolithic European populations. I also show how these ancient Maltese possessed one of the lowest western hunter-gatherer ancestry components among contemporary populations. This is of great interest as it points to recent isolation of this group.

In the second chapter I will use a prediction method to increase the quality of ancient genetic data. I then exploit these diploid genome wide data to uncover the fine-structure and demography of ancient European populations. This shows that ancient Neolithic Europeans show a structure with similarities to modern populations, a recapitulation probably as a result of geographic barriers. I also perform a demographic analysis which identifies inbreeding and restricted population sizes in specific Neolithic populations.

In the third chapter I study phenotypic traits in ancient Western Eurasian populations. First I demonstrate how traits such as height can be reliably estimated in ancient populations. I then use regression analyses to describe how this trait changed across time. Finally I describe a body mass index analysis in ancient Maltese and I compare the results with other Neolithic Europeans.

## Summary

The field of ancient genomics has experienced great progress during the last two decades thanks to advancement in both wet and dry-lab techniques. The application of SNP capture methods and a steep reduction in the price of genome sequencing has allowed the analysis of hundreds of ancient individuals that lived thousands of years ago. While this represents great progress, the low coverage of these samples still represents an obstacle to the study of ancient populations. Genotype imputation represents a useful tool that can be used to face this challenge and increase the quality of ancient genome sequences. In this thesis I will demonstrate how genotype imputation can be successfully applied to both whole-genome and SNP capture sequenced ancient samples. Once imputed I will also show how this new information can be used to investigate the fine structure, population size, inbreeding and phenotypic traits in ancient European populations, particularly using haplotype-informed analyses.

By sampling petrous bones and exploiting advanced sequencing technology chapter 2 will describe how three late Neolithic individuals from Malta were sequenced to medium coverage. Given the vicinity of Malta near to both the Northern coasts of Africa and southern coast of Italy I first analysed whether the island was used as a population bridge to connect both these places. Autosomal DNA analysis suggests that this was not the case, likely because of the greater distance that spans between North Africa and Malta. I then looked genetically at the degree of connection between Neolithic Malta and other southern Italian sites. This is because the presence of imported obsidian on Malta, during the temple period, suggests that it was part of a trading network that connected places such as

Sicily, Sardinia, the Eolian islands and mainland Italy. By using shared genetic drift information it was discovered that late Neolithic Maltese were more genetically close to southern Mediterranean populations. Finally I investigated the presence of Steppe ancestry in late Neolithic individuals. During the time when our samples lived European populations became heavily influenced by Steppe migrations commencing at the beginning of the Bronze Age. Using both single haplotype and autosomal DNA I could not detect any sign of Steppe ancestry in our group, suggesting a Bronze Age influx was not responsible for the disappearance of the Temple people.

In the third chapter I used the haplotype information to investigate the fine structure, inbreeding and effective population size of ancient European populations. While pseudohaploid data has allowed the study of genetic structure and admixture in Neolithic Europeans little is known about their demography and inbreeding. By imputing the genomes of three ancient late Neolithic Maltese we were able to detect a sharp drop in their ancestral population size at the end of the Temple Period which coincides also with what archaeological records suggest. Through an inbreeding analysis I could also observe that the Xaghra Circle individuals were extremely inbred compared to contemporary populations. A second part of this chapter used shared haplotype information to investigate the genetic structure of European Neolithic populations at an unprecedented resolution. The main finding was that the genetic structure of ancient Neolithic Europeans mirrored their geographical location. This level of stratification was likely formed because of obstacles, such as large sea gaps, that hindered admixture between nearby populations and exacerbated founder effects during colonisation. Another contribution could be that different levels of admixture with local hunter-gatherer populations played a role in this diversification.

In chapter 4 I used the same imputation approach to investigate height and body mass index in ancient Europeans. The first aim was to prove the possibility of using imputed genetic data to predict phenotypic traits such as height in ancient humans. By focusing on a large set of ancient Europeans with annotated height I demonstrate that genetic imputed data can be used to predict the height trait with good accuracy. Then genetic and osteological results were compared to highlight

trends in stature across time periods. This showed that Viking individuals were unusually tall compared to preceding and Mediaeval individuals both genetically and physically. Finally, using isotope information to represent the level of protein intake it was investigated whether diet was detectably the height of ancient populations.

The same approach was used to estimate genetic predisposition for body mass index in Neolithic Europeans. In particular ancient Maltese that were rich in human repsesentations of obese figurines. Nonetheless, the distribution of body mass index across Europe was uniform and the Maltese were within its confidence interval.

## Acknowledgements

I would like to thank Prof. Daniel Bradley for guiding me through my entire Ph.D. His insightful ideas and patience were vital in motivating me through all my projects.

I would also like to thank all my colleagues of the lab and the members of Prof. Russel Mc Laughlin lab. Thanks to their experience and friendship I have had a wonderful time both in Dublin and outside when travelling for conferences. In particular I would like to thank Prof. Lara Cassidy and Dr. Valeria Mattiangeli who greatly contributed to this work.

I am grateful to all the members of the Fragsus project that allowed me to be part of the work on the Maltese. Their knowledge and passion for history have profoundly inspired me through my entire work on ancient DNA.

To my family and my girlfriend, a special thanks for their encouragement and understanding. Their continuous support during my PhD gave me the strength to succeed in it.

My gratitude goes also to all my friends, in particular Martina, Flavia, Nikos and Riccardo with whom I have many wonderful memories of my time in Dublin.

# Table of contents

# 1. Introduction

## 1.1 Population genetic analysis of modern genomes

Population genomics is a branch of science that offers powerful tools to investigate the history and evolution of groups of organisms through the study of their genomes. With the word genome we refer to the complete set of genetic information that is inherited from parents to offspring. In humans the first draft of a complete sequenced genome was obtained around two decades ago by the Human Genome Project and the Celera Genomics One groups (Venter *et al.*, 2001). This draft, obtained using samples from multiple individuals, covered almost 85% of the total human genome. Since then many efforts were taken to fill the remaining 15% of the missing genome and recently, thanks to the Telomere-to-Telomere (T2T) Consortium, completion was reached (Nurk *et al.*, 2021).

The Human Genome Project cost approximately 300 million dollars for the first draft to be released in 2000 and another 150 million to obtain a second draft in 2003. Before 2008 the cost of sequencing a genome using Sanger technology was around 25 million dollars (Schwarze *et al.*, 2020). For this reason, prior to the recent advancement in next generation sequencing (NGS) technologies, microarrays were governing the market to characterise, in a cost effective way, the genotypes of many individuals. A famous example of use of these technologies is in the HapMap project that helped characterise the genotypes for more than 1 million common variants present in 4 different populations (International HapMap Consortium, 2003). Notwithstanding the cost-effective advantage of this technology, microarrays commonly target variants present in high frequency among populations and they are mostly limited to specific regions of the genome.Since 2008 the price for sequencing a genome has been in steady decline with NGS technologies able to be used now on a population scale. An important project that took advantage of this was the “1000 Genome Project” (1000 Genomes Project Consortium *et al.*, 2015). This project, completed in 2015,

described the genotypic variation for 2504 individuals divided in 26 groups. Approximately 88 million variants were released from this project, including those of high and low frequency and they characterise the structure and disease burden present in the studied populations. Another example of a resource that used NGS to uncover population structure was the Simons Genome Diversity Project (Mallick *et al.*, 2016). This project produced one of the largest freely available datasets of high quality genomes (with a coverage higher than 30X) for 260 individuals from 127 populations. This, together with the 1000 Genomes project have also been used to increase the quality of subsequent genomes in the imputation of missing genotypes.

## 1.2 Features of the ancient DNA

Having samples of good quality is essential in genetics to achieve accurate results. Unfortunately, when dealing with ancient specimens, time and exposure to environmental conditions can strongly compromise the quality of a sample. When it comes to extracting DNA, it has been shown that it is theoretically possible, in optimal conditions, for a molecule of DNA to survive for hundreds of thousands of years after the death of an organism (Lindahl, 1993). However environmental conditions such as temperature and PH can highly affect this estimate. For example, a recent work successfully sequenced DNA from a mammoth dated more than a million preserved in a perfarmost environment (van der Valk *et al.*, 2021). Among the damage that typically befalls an ancient DNA molecule, most prominent is the degradation to short fragments that are usually a maximum of 50 to 100 base pairs (bp) in length.

Another common damage pattern that occurs in ancient DNA molecules is the deamination of cytosine bases to uracil (C ->T) and through opposite strand pairing, guanine to adenine (G->A). It has been clearly shown that C to T deamination occurs mainly at the 5′ end of a DNA fragment while G to A occurs mainly towards the 3′ end (Briggs *et al.*, 2007).

Both DNA breaks and deamination damages were described in one of the first ancient genomics papers (Pääbo, 1989) and were then subsequently used as one approach to assess the authenticity of the DNA extracted from ancient specimens (Krause *et al.*, 2010). To correct for deamination damage, ancient DNA samples can be treated experimentally with a mix of uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII) enzymes during the library's preparation. This mix of enzymes removes the uracils at the end of the DNA fragments reducing also the length of the molecules. When samples are not or only partially treated with this mix, *in-silico* methods such as MapDamage (Jónsson *et al.*, 2013) or PMDtools (Skoglund, Northoff, *et al.*, 2014) can be used to quantify and correct for these de-amination damages.

## 1.3 The study of ancient genomes

When the price of sequencing was still too high to approach an ancient whole genome, the first ancient DNA experiments focused on studying mitochondrial DNA. This genome possesses uniquely interesting features: it does not undergo recombination, it is passed only through maternal inheritance and it contains a high quantity of information, in terms of variation present, due to its high mutation rate.

The first study that used mitochondrial DNA to study ancient specimens was led by Russ Higuchi and focused on an ancient extinct zebra-relative species (Higuchi *et al.*, 1984). One year later, another work applied similar methods to extract and study ancient mitochondrial DNA from(Pääbo, 1985) an Egyptian mummy aged approximately 2400 years.

Since then, thanks also to the progress in the polymerase chain reaction(PCR) techniques, ancient DNA has been successfully extracted from many different species and tissue such as bone (Hagelberg, Sykes and Hedges, 1989), hair (Gilbert *et al.*, 2004) and parchment (Teasdale *et al.*, 2015). With the drop in the cost of sequencing a genome and the advancement in DNA extraction techniques it became then possible to study ancient genomes on a population scale. For

example, silica based approaches to extract the nuclear genome from ancient humans show efficient yields(Yang *et al.*, 1998).

With regards to DNA extraction in human tissues a group of researchers in Trinity College Dublin tested a variety of different bone tissues in ancient specimens to understand which part of the human body yields the highest quantity of DNA. For this type of analysis 23 types of bones from 13 different ancient individuals were tested. The types of bones tested were dental crowns and roots, ribs, metacarpal and metatarsal portions and petrous bones. Among these tissues tested the petrous bone was found to be the part of the body storing the highest quantity of DNA probably because of its high density which acts as a time-capsule for DNA(Gamba *et al.*, 2014).

Another recent progression that made it possible to expand the number of ancient samples studied was the application of in-solution hybridization capture techniques(Avila-Arcos *et al.*, 2011). These methods target specific loci across a genome by using short DNA probes of length between 50 and 80 bp (Orlando *et al.*, 2021). The main advantage of this technique is that it allows the enrichment of DNA even in samples with low endogenous content as is typical of ancient specimens. However their application comes at the cost of a higher ascertainment bias compared to whole genome sequence (WGS) methods. The main factors that drive this bias are the difference in the affinity of the probes for certain variants and the use of a panel of probes that are not representative of the population investigated(Lachance and Tishkoff, 2013). For example, a recent work discovered that the application of a SNP capture panel which was not representative of the target population led to a biassed estimation of heterozygosity in cod populations(Bradbury *et al.*, 2011).

Despite the recent discoveries, extracting DNA in good quality and quantity still remains a challenge for ancient specimens. In fact, both time and environmental conditions are the two main factors that affect the quality of a DNA sample. When a sample does not yield enough DNA, the cost of sequencing it to high coverage becomes exponentially high.

## 1.4 Bioinformatic analysis of ancient DNA

In addition to sequencing and wet-lab technical advancements, *in silico* methods have also been subject to several advances. When just few loci were considered during the analyses, summary statistics methods such as Fst were commonly used to study the divergence between populations. The way the Fst method works is to calculate the genetic distance within and between populations to give an overview of the degree of differentiation among groups. For example Luca Cavalli Sforza used Wright's Fst method to build a tree that divided 15 populations based only on 20 loci (Cavalli-Sforza, Menozzi and Piazza, 1994). With an increase in the number of individuals and genotyped variants more statistically advanced models have been applied to understand migration and genetic drift in ancient and modern populations. For example Luca Cavalli Sforza et al. (Menozzi, Piazza and Cavalli-Sforza, 1978) introduced the use of principal component analysis (PCA) to study the genetic structure of populations by combining analysis of allele frequencies of many variants. A PCA is a method that allows the reduction of dimensionality of a dataset without losing too much information. By doing so it allows an user to easily visualise the variation within a dataset. Since it was first used by Cavalli Sforza this method, together with others that make use of allele frequencies, have been extensively used to understand phenomena such as population structure and selection. The main advantage of these approaches is that they can be applied on low coverage genomes, where diploid genotype information is uncertain. For example, the program *smartpca* implemented in the package EIGENSOFT (Patterson, Price and Reich, 2006) allows the use of PCA on ancient data in the form of pseudo-haploid genotyped SNPs. A suite of methods - *F-statistics* - have been developed to extract information from these types of sparse data. These methods are particularly focused on understanding events of admixture that shaped populations in the past. The most commonly used ones for pseudo-haploid analysis and also the ones described in this thesis are the F4/F3-statistics and the qpAdm methods. All these methods work by comparing the pattern of allele sharing between two or more target populations and an outgroup one. When the amount of shared alleles between the targets is higher than each of them with the outgroup we can assume that there was a recent admixture event between the targets. However for many investigations, such as

inferences of recent demography, use of allele frequency information alone is usually not sufficient and incorporating haplotype segregation is desirable. The genotype concatenated in a haplotype can be used in a population to study demography, gene flow and selection. For example, long tracts of genome inherited by common ancestors can be used to estimate the size of a population (Browning and Browning, 2015). Also loci in linkage (LD) can be incorporated in long chunks of genome to investigate the fine structure of populations (Lawson *et al.*, 2012). When analysing sequenced ancient DNA samples, missingness of SNP calls is a strong factor that often compromises the identification of haplotypes. Imputation of missing genotypes can alleviate this and these are typically used for low coverage samples. Despite that, imputation is a technique that has been used for many years (George and Elston, 1987; Keavney *et al.*, 1998); its application to ancient genomes has only been recently implemented. The main reasons for this are two: first, up until recently, only few reference dataset had enough information to accurately be used to impute low coverage genomes. The second reason is due to the high computational cost that these imputation algorithms require until newer and faster ones have been recently developed. The first work that successfully applied genotype imputation investigated population size and selection in a small group of ancient European individuals (Gamba *et al.*, 2014). Since then further work expanded the set of ancient imputed genomes to investigate fine genetic structure, inbreeding and phenotypic traits in European populations (Jones *et al.*, 2017; Martiniano *et al.*, 2017; Cassidy *et al.*, 2020; Margaryan *et al.*, 2020; Ariano *et al.*, 2022). To answer all these questions about demographic events that shaped populations we used the haplotype information that is shared within and between groups. For haplotype we refer to a set of genetic variants that are located in a particular chunk of genome and are inherited by a recent ancestor. The size and frequency of these haplotype blocks can be used as information to programs such as Chromopainter (Lawson *et al.*, 2012) to investigate the fine-grained population structure of a group of samples. As the name suggests, the program Chromopainter uses the haplotype information as a set of colours that come from reference populations and use these to “paint” the genome of target groups. A further software called Finestructure (Lawson *et al.*, 2012) uses the output of Chromopainter to reveal the fine population structure of specific groups.

## 1.5 The arrival of humans in Eurasia

In prior studies that supported the origin of modern humans in Africa, anthropological models, based on fossil evidence, were proposed. One of these models, the "Recent African Origin" (RAO)(Stringer and Andrews, 1988), placed the origin of anatomical moderns humans (AMHS) in Africa somewhere around 100 thousand years ago. In support of this a study focused on phylogenetic analysis of modern mitochondrial DNA proved that individuals outside Africa appear as a subset of the variation present in populations from Africa (Cann, 1988). Using a maximum parsimony tree method with a molecular clock, the same method also dated the first split of AMHS from Africa around 200 thousands years ago in line with the RAO model. By looking at autosomal DNA data, after this first out of Africa wave a second migration event followed around 50 thousand years ago mainly contributing to the formation of modern day populations (Henn *et al.*, 2012). However the nature of this event including the region of Africa where this wave originated is still a matter of debate with two main routes proposed. The first one called the northern route connected Northern Africa to the Arabian peninsula through the Nile-Sinai-Land Bridge. The second one, called the southern route, involved seafaring through the Bab el Mandeb strait. In support of this second route of migration, between 65 and 55 thousands years ago favourable climate conditions resulted in a drop in the sea level of the strait and allowed the Arabian peninsula to be visible from the coast of Africa (Beyer *et al.*, 2021). Recent work has associated the expansion of the mitochondrial haplogroup L3 with the dispersal from Southern Africa crossing the Bab el Mandeb strait (Soares *et al.*, 2011; López, van Dorp and Hellenthal, 2015). The presence of marine food indicating settlements near the Western coast of the Red sea also support the possibility of a seafaring cross of the strait (Walter *et al.*, 2000). However, there is an absence of boat remains supporting this hypothesis (Lambeck *et al.*, 2011; Beyer *et al.*, 2021). After migrating towards Eurasia, early modern Eurasians mixed with local Neanderthal groups who contributed between 1 and 4% of their ancestry to present-day Eurasian populations (Green *et al.*, 2010).

## 1.6 The Palaeolithic and Mesolithic periods

Early modern humans in Europe started to share genetic ancestry with present day ones around 37 thousand years ago testified by two of the oldest individuals sequenced and buried in Russia and Belgium, named respectively Kostenki14 and GoyetQ116-1. From this lineage a hunter-gatherer group associated with the Gravettian culture appeared around 33 thousand years ago. Interestingly this group appeared to have been genetically homogeneous despite that individuals sequenced were widely dispersed around Europe (Fu *et al.*, 2016). Subsequently the last glacial maximum (LGM) period began, stretching between 25 and 19 thousands years ago. During this period the temperature fell around 10 degrees below the present level and much of northern and middle Europe was covered in ice. Due to this shift in climate the sea level experienced a steep drop of around 110 metres uncovering regions of the Mediterranean sea (Cunliffe, 2017). Given that most of Northern Europe was covered in ice, hunter-gatherer populations survived in refugia in the southern part of Europe that were shielded by the Alps and Pyrenees from the cold climate. The HG groups that survived this episode experienced a strong bottleneck as indicated by demography estimates obtained from mitochondrial (Posth *et al.*, 2016). Following the LGM, an interstadial period started, called Bølling-Allerød, with the retreat of the ice-sheet. Also around this time the sea levels started to rise again and much evidence shows that hunter-gatherer individuals living close to the Mediterranean sea made extensive use of boats for seafaring (Cunliffe, 2017). For example exogenous obsidian has been found on the island of Melos in a cave occupied during the Palaeolithic period (Davis, 2001). With this rise in temperature hunter-gatherer populations left their refugia in Southern Eurasia to migrate towards northern latitudes. For example a group of Magdalenian HG individuals that lived in Spain, France and Germany were found to have a high genetic similarity with a HG population that lived 20 thousands years before in Belgium, thus probably representing the expansion of this from their refugia (Fu *et al.*, 2016). By 14 thousands year ago this Magdalenian population mixed and was later replaced by another one represented genetically by the Italian sample Villabruna (Villalba-Mouco *et al.*, 2019). This sample represented an early example of a particular population defined as western hunter-gatherer (WHG) which seems to have originated in

southeastern Europe (Lazaridis *et al.*, 2016). During the Mesolithic period another population defined as eastern hunter-gatherer (EHG) is detected by its presence in Russia around 8 thousands years ago (Haak *et al.*, 2015). This group was genetically characterised by a mix of WHG and a Palaeolithic Siberian ancestry which is represented by the individual Mal'ta that lived in Siberia approximately 24000 years ago (Raghavan *et al.*, 2014). This EHG group contributed to the ancestry of individuals from Norway, Sweden (Skoglund, Malmström, *et al.*, 2014) and a few HG populations from Eastern Europe (Lazaridis, 2018). Genetically distinct from these two previous HG populations, another group named Caucasus hunter gatherer (CHG) has been recently characterised through ancient DNA analysis. Genetic analysis showed that this group split from the WHG genome approximately 45 thousand years ago, after the second wave of migration of the modern human in Eurasia (Jones *et al.*, 2015). One effective display of the split between these HG groups is their placing at the extremes of a PCA composed of ancient and modern Europeans (Lazaridis *et al.*, 2016).

## 1.7 The arrival of farming in Europe

The domestication of animals and the cultivation of crops started in a large area of south Asia spreading from Turkey to Iran between 9600 to 6900 B.C. From small groups of HG, individuals started to increase in number and adopt a more sedentary lifestyle thanks also to the higher abundance of food (Cunliffe, 2017). The first genetic evidence of a distinct group that appeared in the Near East during the Neolithic period was in 2005. In this work mitochondrial DNA sequence was produced from 24 ancient individuals from Europe dating around 7500 years ago. The haplogroup of these individuals was shown to be frequent in Asia but not in modern Europe, which suggested their origins in a farming population in the Near East that later spread into Europe (Haak *et al.*, 2015). A genomic analysis establishing this connection emerged in 2012 (Skoglund *et al.*, 2012). In this work low coverage shotgun sequenced farmer individuals from Scandinavia showed a high similarity with populations from Greece and the Balkans suggesting that farming started to spread from Southern European regions towards Northern ones.

According to archaeological evidence, the expansion of farming from the Near East reached the Balkans and developed in the Starčevo–Kőrös–Criş culture (Whittle 1996). From there two different streams of migration expanded into Europe: one, named Cardial for the characteristic type of pottery produced, followed the coast along the Mediterranean sea arriving in Iberia around 5500 years B.C. The other vector of migration involved a different group called Linearbandkeramik (LBK) for the typical banded decoration pattern of their pottery. This group expanded towards Central Europe following the Danubian river valley. One of the first works that found a genetic distinction between these two groups was published in 2015 by studying Neolithic individuals from Iberia and Central Europe. In this work, it was observed that Cardial individuals possessed higher levels of WHG ancestry compared to LBKs from Central Europe (Olalde *et al.*, 2015).

While many analyses later agreed with these results of a tight relation between the Near East and European early farmers, the sequencing of ancient Iranian farmers gave unexpected results. Despite these individuals, found in the Zagros mountains, adopted a farming lifestyle with cultural similarity to that in the western Fertile Crescent they were genetically distinct from their neighbour Anatolian farmers (Broushaki *et al.*, 2016).

## 1.8 Sea as a conduit during the Neolithic period

Archaeological consideration of maritime connectivity has ranged from a biogeographical perspective that considers the sea as a barrier to a view of seaways as ancient highways that facilitate exchange (Rainbird, 2007). For example, the spread of the Cardial culture, following the Mediterranean coast from Greece to Portugal, was faster than to the inland migration of the LBK culture (Henderson *et al.*, 2014; Manen *et al.*, 2019). The difference between the inland and coastal ways of migration is also testified by different variation of crop cultivated in each of these (de Vareilles *et al.*, 2020; Gaastra, de Vareilles and Vander Linden, 2022). For example, Cardial sites along the Mediterranean coasts showed a higher variety of cereals compared to LBK sites. Archaeological

analyses (mainly materials and radiocarbon) have demonstrated that many maritime excursions must have occurred in the Mediterranean sea during the Neolithic period (Guilaine, 2017, 2018). Starting from 6000 years BC the Cardial culture from Greece expanded towards Italy crossing the Adriatic and later the Tyrrenean sea reaching the islands of Sicily and Sardinia around 5500 years BC. One proof of this maritime connection was the interchange of materials that occurred between the Mediterranean islands. Around 9000 years ago an eruption happened on the island of Lipari produced a high amount of obsidian, a material used during the prehistoric time to create tools such as arrowheads and other sharp objects. Around this same period Lipari was inhabited by people migrated from Sicily and with the presence of obsidian new trade started to occur that connected the island of Lipari to different Italian regions, both coastal and inland (Tykot, Freund and Vianello, 2013). Although the sea worked as a facilitator for exchange, the limited capacity of early boats affected the initial settlement and expansion of some islands like Britain and Ireland (Cunliffe, 2018). For example, using ancient DNA data it was shown that Mesolithic individuals from Ireland formed a distinct group compared to other European hunter-gatherers, indicating that the Irish sea acted as a barrier to genetic exchange prior to the Neolithic period (Cassidy *et al.*, 2020).

## 1.9 Representation of common phenotypes in ancient cultures(e.g. obese figurines in Malta)

Human figurines have often been objects of great interest in archaeology for their association with particular cultures and periods. For example the first human figurines to be found were of obese shape during the Ice Age period (around 30 thousand years ago). These figurines are unlikely to have represented the real status of hunter-gatherers who probably suffered the lack of food, especially during the Ice Age (Johnson, Lanaspa and Fox, 2021). In contrast to male figurines that were slim, these obese figurines were always women and their proportions put more emphasis on breasts and hips compared to feet and arms. This was typically interpreted as a symbol of beauty and fertility from which these figurines acquired the name Venus or “Mother goddess” figurines (Nasab

and Kazzazi, 2011). Interestingly during the ice Age period these women were represented as naked which seems counterintuitive given the harsh cold climate present at that time. It has been shown that the presence of these figurines correlated with the moments of harsh climatic conditions and extreme nutritional stress periods during the Palaeolithic time (Johnson, Lanaspa and Fox, 2021). The interpretation of the presence of these figurines in some archaeological sites during the Neolithic period is still a matter of debate. For example in an initial excavation of the Çatalhöyük site the presence of obese women figurines led to a conclusion of a matriarchally structured society (Mellaart, 1967) . However, later excavation of the same site failed to prove this hypothesis by finding few female figures dispersed randomly across the site (Hodder, 2010). The Maltese islands were another region where the production of figurines flourished during the Neolithic period. During the Temple period different types of figurines (such as anthropomorphic or zoomorphic) were present. Among these, obese figurines of women in a seated position particularly stand out for their presence. The interpretation of these figures is still open to debate especially in the context of rituals celebrated in megalithic temples (Malone and Stoddart, 2016).

## 1.10 Genetic and environment contribution on the health of ancient individuals

From prehistory until recent times different conditions shaped the human body, such as variations in diet or sudden changes in climate. For example in Britain, at the onset of the Neolithic period, there was a sharp change in the lifestyle from a marine based diet to a terrestrial one probably as a consequence of the introduction of domesticated animals (Richards, Schulting and Hedges, 2003). Similarly, during the Neolithic period the consumption of milk began after the domestication of animals. This has been proved by archaeological evidence showing the presence of milk lipids on the surface of Neolithic pottery (Cramp *et al.*, 2014; Jessica Smyth and Richard P. Evershed, 2015). These changes in diet caused human digestive systems to adapt. For example from approximately 4500 to 1500 years before present there was in Europe an increase in the frequency of the alleles linked to the digestion of milk (Burger *et al.*, 2020). As a response to

these dramatic changes in lifestyle and diet during the Neolithic Revolution the overall health of populations was also affected. For example, an increase in the consumption of maize by prehistoric American agriculturalists was correlated with an increase in the frequencies of caries (Meller *et al.*, 2009; Latham, 2013). Also interestingly, a high consumption of protein from meat, testified by an isotope analysis made on Late Roman remains (Jørkov, Jørgensen and Lynnerup, 2010), was linked to an increase in height. In contrast, the nutritional stress that affected some periods, such as the so-called "Little Ice Age" at the end of the Mediaeval period was argued to have resulted in a decline of body mass index and stature in Northern European populations (Ruff, 2018).

# 2. The genetic makeup of the Maltese population during the Neolithic period

## 2.1 Introduction

The evidence supplied by archaeology, particularly the affinities between Għar Dalam and early Neolithic Impressed Wares of Southern Italy, strongly suggest that the source population of the Neolithic expansion into the Maltese islands was located in Southern Italy and Sicily. Theories of an earlier colonisation of Malta have been debated, but since hunter-gatherer populations require a large space for foraging, it seems unlikely that Malta would have been a viable long-term home before the advent of agriculture (Sagona 2015).

From the first evidence of human settlement, the early Maltese society evolved through different cultural phases: Għar Dalam, Grey Skorba, Red Skorba and finally Żebbuġ, signalling the start of the Temple Period and an increasingly distinctive culture. In this last phase, the use of rock-cut tombs, containing collective burials and distinctive pottery defined the island culture (Malone *et al.*, 1995). Subsequent cultural phases (the Temple Period) witnessed an unprecedented development in Maltese society, culminating in the Tarxien phase between 2900 and 2400 BC (McLaughlin, Stoddart and Malone, 2018). During the Tarxien phase, collective burial in the elaborate Xagħra Circle cave complex on Gozo and at the Hal Saflieni hypogeum in Malta represent exceptional mortuary sites. The Circle excavations unearthed the individuals analysed for this study in the early 1990s (Malone and Stoddart, 2009) and are the subject of additional study in this volume.

The ancient DNA work we report here was undertaken in collaboration with the FRAGSUS project (2013 - 2018) as part of a programme of environmental and archaeological research, including an extensive re-assessment of the Xagħra site (details in Figure 2.1), applying additional radiocarbon dating and stable isotope studies. The overall aim of this research has been to better understand the cultural,

economic and environmental dynamics of prehistoric Malta (Malone, Cutajar and McLaughlin, 2019). The first part of this work involved the analysis of the genetic variation of the Maltese population in comparison with other European groups that lived during the same period. The aim in this case was to highlight previous archaeological findings that described a strict relation between the Neolithic Maltese and other Southern Mediterranean populations. The second part of this work focused on investigating the genetic isolation of the Maltese population through an analysing of HG ancestry.

Since ancient times the Mediterranean Sea has represented one of the most important routes for migration in southern Europe. For example, during the late Neolithic period there is proof of both a cultural and a direct genetic connection between Portuguese and Greek Neolithic populations (Hofmanová *et al.*, 2016). Despite this evidence, the prehistoric population history of South Europe remains under-explored in terms of genetic studies. In contrast, most aDNA publications have focused on the history of Central and Northern European populations, with little attention paid to southern Europe. The reason for this absence is because of the particularly warm climate conditions that tend to accelerate the degradation process of aDNA samples. Importantly, the Maltese work we are reporting here is the genetic analysis of one of the most southerly archipelagos of the Mediterranean.

This chapter is focused on uniparental genetic data (mitochondrial DNA and Y chromosome haplotypes) as well as autosomal variant calls from 3 ancient individuals that lived in Malta during the transition between the Neolithic and Bronze Age periods. Thanks to these data the question is addressed of whether the Maltese were genetically more similar to Neolithic or to Bronze Age populations in Eurasia. The three sequenced samples described in this work all derive from the megalithic burial Circle on the Xagħra plateau between the temples of Ġgantija and Santa Verna, excavated between 1993 and 1994. The oldest sample (Xaghra6) derives from a deeper stratified area of stacked burials that also contained rich ceremonial objects higher in the stratigraphy. The two later samples were found in shallower deposits to the west. Xaghra5 was part of a display area of initially articulated human remains placed with portable figurines that was intentionally

dismembered, most probably, at least in part to the slightly deeper location of Xaghra9 slightly to the north. Xaghra6 was placed as the use of the site began to intensify whereas the other two samples date to the period of most intensive activity some four hundred years later (c. 2500 BC; Figure 2.1B).

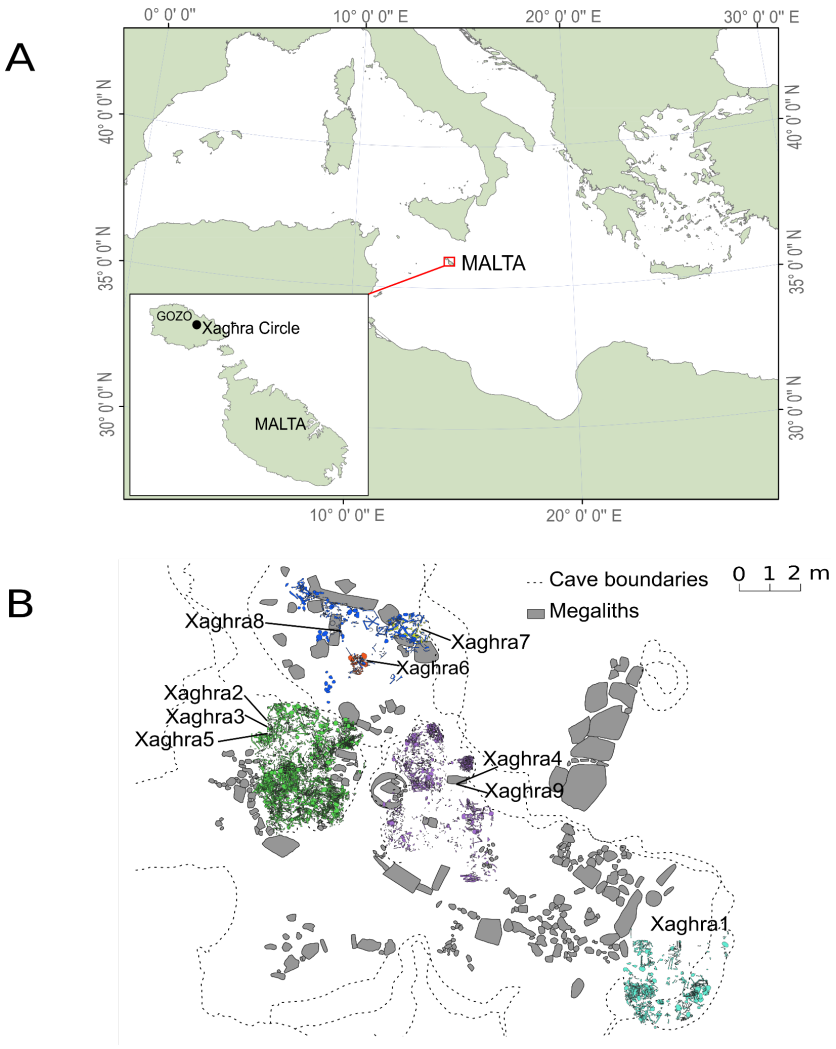


***Figure 2.1: Location of the samples within the Maltese Xagħra Circle site.*** *A) Location of the Maltese archipelago within southern Europe. B) Plan of the Xaghra Circle site showing skeletal remains from the archaeological contexts studied. Colours represent different archaeological layers (green: 783, blue: 951, lilac: 960, yellow: 111, turquoise: 1241, orange 1307)*

## 2.2 Methods

### 2.2.1 Sampling and DNA extraction

For this project, 5 petrous bones and 4 teeth from the Xagħra Circle archaeological site in Malta were processed in the clean room facilities of the Smurfit Institute, Trinity College, Dublin. Full body suits, face masks, hairnets and gloves were worn during the work. All tools and surfaces were cleaned with bleach, DNA-ExitusPlus, ethanol and exposure to UV light. Samples were photographed extensively prior to any alterations, and were then exposed to UV light for 30 minutes on either side to remove surface contaminants. Sample drilling was carried out in a fume hood lined with bleached tinfoil. The surface of each bone was cleaned using a drill bit. A triangular wedge section of the otic capsule region of each petrous bone and the root of each tooth were cut using a Dremel diamond wheel saw. Each sampled bone part was pulverised in a Mixer Mill MM 400 (Retsch). An aliquot of ~0.1g of this bone powder was used for DNA extraction, and the rest of the powder was stored in a separate tube. The DNA extraction procedure followed the same protocol described in (Yang *et al.*, 1998) with modifications presented elsewhere (MacHugh *et al.*, 2000). One sample subsequently sequenced at high coverage was re-extracted using an initial washing step by 0.5% bleach solution as described in (Boessenkool *et al.*, 2017). For each of the samples approximately 150 mg of bone powder were mixed with a lysis buffer containing: 20 mM of Tris HCL, 47.5 mM of EDTA and 0.65 U/ml Proteinase K. After an incubation of 48 the supernatants of each of the solutions was transferred into an Amicon Filter tube. Each sample was diluted in 3ml of 10mM Tris-EDTA Buffer plus 1mL of WEX and centrifuged at 5,000 rpm. The resulting extract, which is approximately 100 µl, is transferred in a QIAgen column together with 500 µl of binding buffer. This is spun for 1 minute and the resulting flow is discarded. After this 750 µl of PE is added to the column and spin again for 1 minute. After discarding the flow from this step 55 µl of EBT are added into the QIAgen column and after a spin of 1 minute the subsequent DNA flow is taken from the tube.

### 2.2.2 Radiocarbon analysis

All the samples considered in this work have been excavated around 1993 in the megalithic burial Circle on the Xagħra plateau. The sample Xaghra6 is the oldest one among the three samples and was found in a context rich in ceremonial ornaments. The remains of the other two samples, Xaghra5 and Xaghra9 were discovered in the western part of the site. In order to estimate the dates for our samples a Bayesian chronological model was used that included 117 radiocarbon dates already estimated from similar stratigraphic levels. A 95% confidence interval from this model was considered to date the samples (Malone, Cutajar and McLaughlin, 2019). It was estimated that the samples Xaghra6 dated between 2900 and 2650 BC (OxA-27837, 4198±26 BP) (Malone, Cutajar and McLaughlin, 2019). Xaghra5 and Xaghra9 were found in association with other 23 radiocarbon measurements from the same burial contexts to date approximately between 2550 BC and 2350 BC.

### 2.2.3 Library preparation

The initial screening of each sample and blank controls was performed by constructing a double-stranded DNA NGS library, priorly treated with Uracil-DNA-glycosylase (UDG), using the method outlined in Meyer & Kircher (Meyer and Kircher, 2010) and modified as described in (Gamba *et al.*, 2014). Libraries were amplified with AccuPrime Pfx Supermix (Life Technology) using 12-14 cycles of PCR, assigned with unique indexes and quantified with a TapeStation 2200 (Agilent Technologies). The same libraries were also used for further amplifications required for high coverage sequencing.

### 2.2.4 DNA Sequencing

The initial screening to assess the endogenous DNA was performed by sequencing all the libraries with the Illumina HiSeq 2500 platform (100bp SE) at Macrogen (Republic of Korea). Subsequently, 3 samples with high endogenous DNA were further sequenced to high coverage using the HiSeq 2500 Illumina

platform (100bp single-end(SE)) at Macrogen (Republic of Korea). The endogenous DNA quantity for each of these samples was estimated by taking the ratio of the reads that aligned to the human genome versus the total amount of reads that were sequenced from the first run. One sample was further sequenced using NovaSeq (50bp paired-end (PE)) Illumina platforms at TrinSeq (Ireland).

### 2.2.5 Reads processing

For samples sequenced in SEmode, reads were trimmed of their adapters and filtered based on their length using the software cutadapt v.1.9.1(Martin, 2011) (cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -O 1 -m 34). For PEibraries, adapters were trimmed and reads were filtered using AdapterRemoval v2.1.1 (Schubert, Lindgreen and Orlando, 2016) (--trimns --trimqualities --minquality 25 --collapse). Reads that passed these qualities and length filters were aligned to the human reference genome (hg19/GRCh37) with the mitochondrial sequence replaced by the Revised Cambridge Reference Sequence (rCRS, NC_012920.1) using the software BWA v.0.7.5a(Li and Durbin, 2009) with relaxed parameters (-l 1024 -n 0.01 -o 2). Aligned reads that came from PCR duplication or with a mapping quality below 20 were removed using the software SAMtools v.1.7(H. Li *et al.*, 2009) and Picard Tools v.1.101 (http://broadinstitute.github.io/picard/). The coverage of each completed aligned file was calculated using the tool Qualimap v.2.1.1 (Okonechnikov, Conesa and García-Alcalde, 2016). Indels were locally realigned using The RealignerTargetCreator and IndelRealigner tools from GATK v.2.4 (McKenna *et al.*, 2010). Additionally 2bp were soft clipped at the start and end of each read.

### 2.2.6 Contamination estimation and sex determination

To determine the sex of each sample we applied two methods, one outlined in (Skoglund *et al.*, 2013) and the other described in (Cassidy *et al.*, 2020). In the first method the amount of reads aligned on the Y chromosome was calculated as a ratio versus the reads aligned to both the X and Y chromosomes. If this ratio

with its confidence interval is below 0.022 or higher than 0.75 a human sample is classified respectively as a female or a male.. Similarly to this method the second one (Cassidy *et al.*, 2020) first divides the amount of reads that align on the X-chromosome versus its total length. This value is then further divided by the same type of measure only considering this time the overall autosomal genome. This final ratio called Rx is used to classify an individual as male if its value is lower than 0.6 and female if it's higher than 0.9. We only considered sex assignments where both methods agreed. For three Maltese samples analysed in this study we estimated contamination using the haploid information contained in the mitochondrial genome and  in the X chromosome for two males, applying the same methods outlined in (Cassidy *et al.*, 2016) (Table 2.2 and Table 2.3).

### 2.2.7 Y and Mitochondrial haplogroups

Samples that were identified as male were evaluated for Y-chromosome haplogroup lineage. This task was executed using the software Yleaf v2 (Ralf *et al.*, 2018) and the International Society of Genetic Genealogy 2019 database(ISOGG) as reference (https://isogg.org/tree/ISOGG_YDNA_SNP_Index.html). Haplogroups annotated with the "~" label, which indicates a distinct group associated with unknown phylogenetic position, were excluded from this analysis. In the case of the mitochondrial genome analysis, our first step consisted of aligning the Fastq files of each sample to the humanrCRS(NC_012920.1) using the tool mpileup from the software samtools (H. Li *et al.*, 2009). Only SNP calls with a base quality above 30 (parameter -Q30) were then retained for further analyses. A consensus mitochondrial Fasta sequence was first obtained for each sample using bcftools software (Li, 2011)(parameter -c) and then given to the software Haplofind (Vianello *et al.*, 2013) for the haplogroup assignment. From this analysis, we considered as valid only the haplogroups that were at the most terminal part of a branch, had an assignment score of at least 0.9 and where the assignment did not derive from a transition SNP.

### 2.2.8 Collection of publicly available data for haplogroup analysis

To contextualise the haplogroup results with other published ancient samples, a well curated dataset of ancient DNA metadata from AmtDB was downloaded (Ehler *et al.*, 2019). This was used to compare the geographical distribution of all sample haplogroups (both mitochondrial and Y-chromosome), focusing in particular on Neolithic, and Bronze Age periods. The samples were finally filtered for latitude and longitude thus restricting our analysis to Eurasia.

### 2.2.9 Population structure analysis

Pseudohaploid genotypes were called at approximately 600,000 autosomal sites from the Human Origins panel (Lazaridis *et al.*, 2014) for a set of 276 ancient samples used in the representative of hunter-gatherers, Bronze Age and Neolithic farmers populations(Haak *et al.*, 2015; Jones *et al.*, 2015; Mathieson *et al.*, 2015; Broushaki *et al.*, 2016). Read bases were determined at each site using the Pileup tool from GATK v2.4 (McKenna *et al.*, 2010), filtered for a quality of 30, with bases not matching either the reference or alternate allele removed. A single base was then randomly selected to generate the pseudohaploid genotype. This ancient dataset was then merged with a subset of the Human Origins panel from Western Eurasia using the software PLINK v1.9(Chang *et al.*, 2015). A PCA was then carried out on the 604 modern individuals from Human Origins, with the genetic variation of the ancient samples projected onto this using the SmartPCA v.16000 algorithm implemented in EIGENSOFT (Patterson, Price and Reich, 2006; Price *et al.*, 2006) with parameters (killr2: YES, r2thresh: 0.2, numoutlieriter: 0, lsqproject: YES, autoshrink: YES).

### 2.2.10 *D*-statistics

To test for Admixture with African populations *D*-statistics were employed. This

test is used for verifying the gene flow between four populations under a null hypothesis of incomplete lineage sorting. In case this hypothesis is rejected if a score (D-score) together with a significance value (Z-score) indicates an admixture and a direction of the gene flow between these four populations

As a tool for performing this test, we used the AdmixTools package (Patterson *et al.*, 2012). Four ancient North African representatives were selected from (Fregel *et al.*, 2018). Tests were constructed in the form of: *D*(Chimp, Ancient North-Africa, Malta Late Neolithic, X) where X represents Neolithic populations that fall close in the PCA to the ancient Maltese (Figure 2.4; Table 2.S3).

Similarly to test for admixture between the Maltese and Caucasus hunter-gatherer (CHG) or Steppe populations we built our *D*-statistics test in the form of *D*(Mbuti, CHG/Yamnaya, Malta Late Neolithic, X). In this analysis the CHG population is represented by two individuals published in (Jones *et al.*, 2015) (Tabled 2.S1 and 2.S2).

### 2.2.11 *F3*-outgroup and qpAdm

As a reference dataset for this kind of analysis we used the "1240K" dataset containing approximately 1.2 million SNPs. This dataset is commonly used in population genetics analysis as it contains a curated list of SNPs that well represent the genetic variability of many different modern human populations (Fu *et al.*, 2016). Using the same set of ancient samples described in the previous paragraph and transversion sites only from the "1240K" panel, we estimated the amount of drift that the Maltese shared with each other population using the *outgroup-f3* statistics method implemented in the ADMIXTOOLS package v.7.0.2. This analysis was carried out in the form of *f3*(Mbuti; Ancient Maltese, X) where X represents different populations tested (Figure 2.6; Table 2.S4). The outgroup population, Mbuti, is represented by four individuals collected from the Simon Genome Diversity Project(SGDP) dataset (Mallick *et al.*, 2016).

To estimate the amount of WHG ancestry we used the method qpAdm. We first divided the individuals into groups according to their archaeological site of origin.

Each group was furthermore subdivided in bins of 1000 years and only sub-groups with at least 2 individuals were considered for this analysis. The reference group was comprised of the following genomes: (Mbuti.DG, Ust-Ishim, MA1, Villabruna, GoyetQ116-1, Han.DG, Papuan.DG, Mixe.DG, Karitiana.DG, AHG, Iran_Neolithic, CHG, EHG). The source population is Anatolian_Neolithic represented by individuals from Barcin and WHG individuals represented by Loschbour and KO1 (Figure 2.5; Table 2.S5 and 2.S6). Only groups with a p-value higher than 0.05 were included for this analysis.

### 2.2.12. Kinship analysis

To explore the kinship relations between individuals in our dataset, the software Lcmlkin was used (Lipatov *et al.*, 2015).This software uses the identical by descent (IBD) information from high and low coverage data to estimate the relatedness between individuals. Specifically, as an input this software takes as the genotype likelihood information stored in a variant calling format (VCF) file and returns as output the probability that two random alleles from two individuals are in IBD. The dataset for this test consisted of: 3 Neolithic Maltese individuals plus a set of 17 unrelated of Italian origin that lived in the same time period all with high coverage genome (Antonio *et al.*, 2019). To create the VCF file the program SNPbam2vcf was used as recommended. The SNP were called on the target dataset using the 1000 Genomes dataset (1000 Genomes Project Consortium *et al.*, 2015) filtered using PLINK v1.9 by: minor allele frequency (--maf 0.01) and 'transversion only' obtaining approximately 3 million SNPs. After SNP calling the dataset was further filtered for genotype missingness (geno 0.98), individual SNP missingness (0.6) and linkage disequilibrium with parameters 200, 25 and 0.4 (Martiniano *et al.*, 2017). The VCF file was used as input for Lcmlkin with default parameters.

## 2.3 Results

### 2.3.1 Mitochondrial contamination and history

A common method for estimating DNA contamination of a sample is to check the rate of heterozygous sites present in the mitochondrial DNA. The contamination percentages of the high coverage samples, not considering sites that can derive from transitions, range from values of 0.3% to 0.78% (Table 2.1). These values can be considered as acceptable for a no-contamination hypothesis. Once assured about the quality of our samples, the software **Haplofind** was used to investigate mitochondrial haplogroups, with the following results (Table 2.3):

- Xaghra5 belongs to the haplogroup K1a which is a subgroup of the major branch K. This branch has already been described in individuals that come from Anatolia during the Pottery and pre-Pottery Neolithic period ((Mathieson *et al.*, 2015)).
- The individual Xaghra6 belongs to the haplogroup V which, although low in frequency, has been found in populations from central Europe associated with LBK, Únětice and Pitted ware culture, and from Neolithic populations in Portugal (Haak *et al.*, 2015).
- Xaghra9 belongs to the haplogroup H4a1, which is a derived branch of haplogroup H. This major group evolved first in the Near East during the Neolithic period and afterward spread into western Europe (Torroni *et al.*, 1998). It appears in fact to be frequent in France during the Middle Neolithic period and Iberia during the Epi-Cardial Neolithic period.

| Sample ID | Date B.C. | Genomic Sex | Endogenous DNA % | Genome Coverage |
|---|---|---|---|---|
| Xaghra1 | 2575-2520 | Female | 1.9 | 0.05 |
| Xaghra2 | 2550-2350 | Unknown | 0.06 | <0.01 |
| Xaghra3 | 2550-2350 | Male | 0.42 | <0.01 |
| Xaghra4 | 2535-2475 | Female | 1.7 | 0.03 |
| Xaghra5 | 2550-2350 | Male | 37 | 1.24 |
| Xaghra6 | 2900-2750 | Female | 12 | 0.98 |

| | | | | |
|---|---|---|---|---|
| Xaghra7 | 2875-2615 | Female | 0.16 | <0.01 |
| Xaghra8 | 2575-2470 | Female | 0.03 | <0.01 |
| Xaghra9 | 2530-2400 | Male | 15 | 7.52 |

*Table 2.1: Summary of all samples from Late Neolithic contexts at the Xagħra circle that have been analysed in this work. Date ranges have been estimated using the 95% confidence interval of Bayesian chronological models (Malone, Cutajar and McLaughlin, 2019). The sex of each individual was predicted using the methods outlined in paragraph 2.2.6.*

| Sample ID | Only ChrY | Ratio ChrY/ ChrY+ChrX | SE | Sex Assignment | Haplogroup | Contamination X% | SE | P-value |
|---|---|---|---|---|---|---|---|---|
| Xaghra1 | 56 | 0.0041 | 1e-06 | Female | // | // | // | // |
| Xaghra2 | 4 | 0.02 | 0.0003 | Unknown | // | // | // | // |
| Xaghra3 | 79 | 0.07397 | 0.000126 | Male | // | // | // | // |
| Xaghra4 | 7 | 0.00515 | 0 | Female | // | // | // | // |
| Xaghra5 | 208312 | 0.1162 | 0.0002 | Male | H2 | 0.6 | 0.0014 | 6.789e-11 |
| Xaghra6 | 43469 | 0.0178 | 0.0001 | Female | // | // | // | // |
| Xaghra7 | 5 | 0.0049 | 1e-05 | Female | // | // | // | // |
| Xaghra8 | 4 | 0.001 | 1e-06 | Female | // | // | // | // |
| Xaghra9 | 177879 | 0.1224 | 0.0003 | Male | G2a2a1a3 | 1.1 | 0.0017 | 1.128e-08 |

*Table 2.2. Sex assignment and contamination levels for each sample. For male individuals also the Haplogroup is assigned using the ISOGG database as reference and a contamination level is given using the X chromosome in male individuals. The ratio of reads that align on chromosome Y versus the overall number of reads that align on both sex chromosomes is the defining estimate for the sex assignment(see paragraph 2.2.6). The contamination percentage represents the rate of heterozygous calls on the X chromosome versus the rate of adjacent monomorphic sites.*

| Sample ID | Mean coverage | Site contamination % | Site contamination % no-MD | Haplogroup | Haploscore | Assignment score |
|---|---|---|---|---|---|---|
| Xaghra5 | 128.26 | 1.422 | 0.533 | K1a | 0.8 | 0.96 |
| Xaghra6 | 106.8 | 1.548 | 0.787 | V | 1 | 0.98 |
| Xaghra9 | 184.87 | 0.563 | 0.340 | H4a1 | 1 | 0.99 |

*Table 2.3. Results from the contamination and haplogroup analyses. No sample shows significant traces of contamination, both excluding and including*

*Transition sites (MD). The assignment score from Haplofind gives a probability of a sample to be part of an haplogroup. The Haploscore gives an assignment score taking into account the previous major haplogroup from the same branch.*

By inspecting the distribution of ancient haplogroups, it appears that the Maltese belonged to mitochondrial branches that were particularly widespread during the Neolithic period. Interestingly, samples that matched the Maltese haplogroups during the Bronze Age period (Figure 2.2) tended to come from central Europe and the Iberian Peninsula and belonged to the Bell Beaker culture.

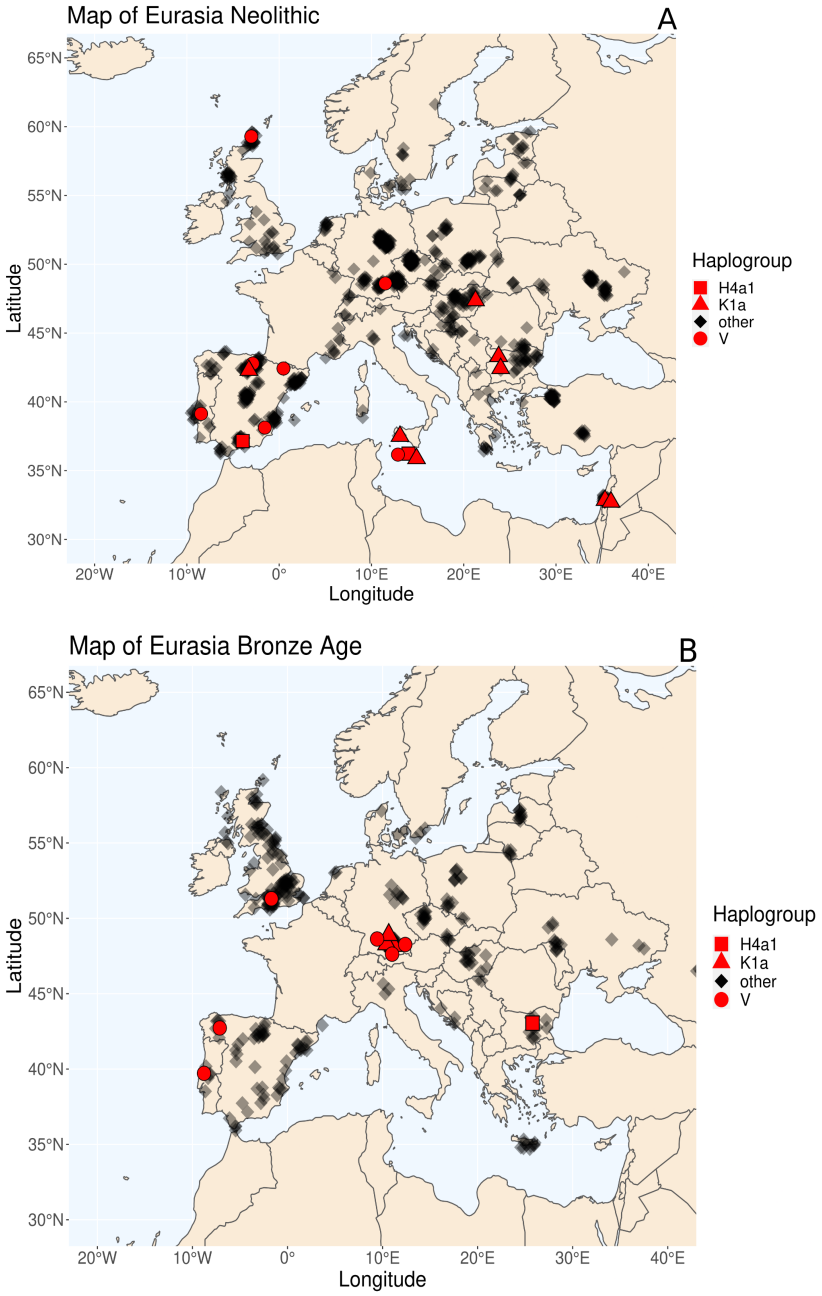


***Figure 2.2: Distribution of ancient mitochondrial haplogroups in Eurasia. Each point is a sample with the shape representing the haplogroup to which it belongs****. Red colour indicates a match with one of the Maltese haplogroups encountered in this work, dark grey points show the geographical distribution of unmatched samples. A) distribution of haplogroups during the Neolithic. B) distribution of haplogroups in Bronze and Iron Age samples. The haplogroup information was taken from the AMTdb (Ehler et al., 2019)*

### 2.3.2 Y-chromosome contamination and lineages

The results from Y chromosome screening indicated that two of the samples (Xaghra5 and Xaghra9) were male. SNP information from the ISOGG database was used to define haplogroups and the two individuals each belonged to one of two common European Neolithic haplogroup branches. Xaghra5 belongs to haplogroup H2. This haplogroup is rarely found in modern European populations and its earliest evidence dates back to a pre-pottery sample in the Levant between 7300-6750 BC (Lazaridis *et al.*, 2016). In more recent times this haplogroup was found in an Anatolian farmer and a European Neolithic sample belonging to the Starčevo culture. Xaghra9 has the haplogroup G2a2a1a3, one of the subclades of the major branch G commonly present in Europe during the Neolithic period (Broushaki *et al.*, 2016). From examination of the incidence of these haplogroups in ancient Eurasia, their prevalence during the Neolithic period compared with later times is clear (Figure 2.3). There is a trend for matches to follow a more southern distribution. In the post-Neolithic comparison, only two H2 matches were found, in an Early Bronze Age sample from Bulgaria. Haplogroup G2a2a1a3 was interestingly found in 3 samples from the Neolithic-Copper Age in Spain and Portugal. Other close subclades are common among Early European farmers and rarely feature in the Bronze period sample where they are mostly replaced by haplogroups R1a and R1b (Haak *et al.*, 2015).

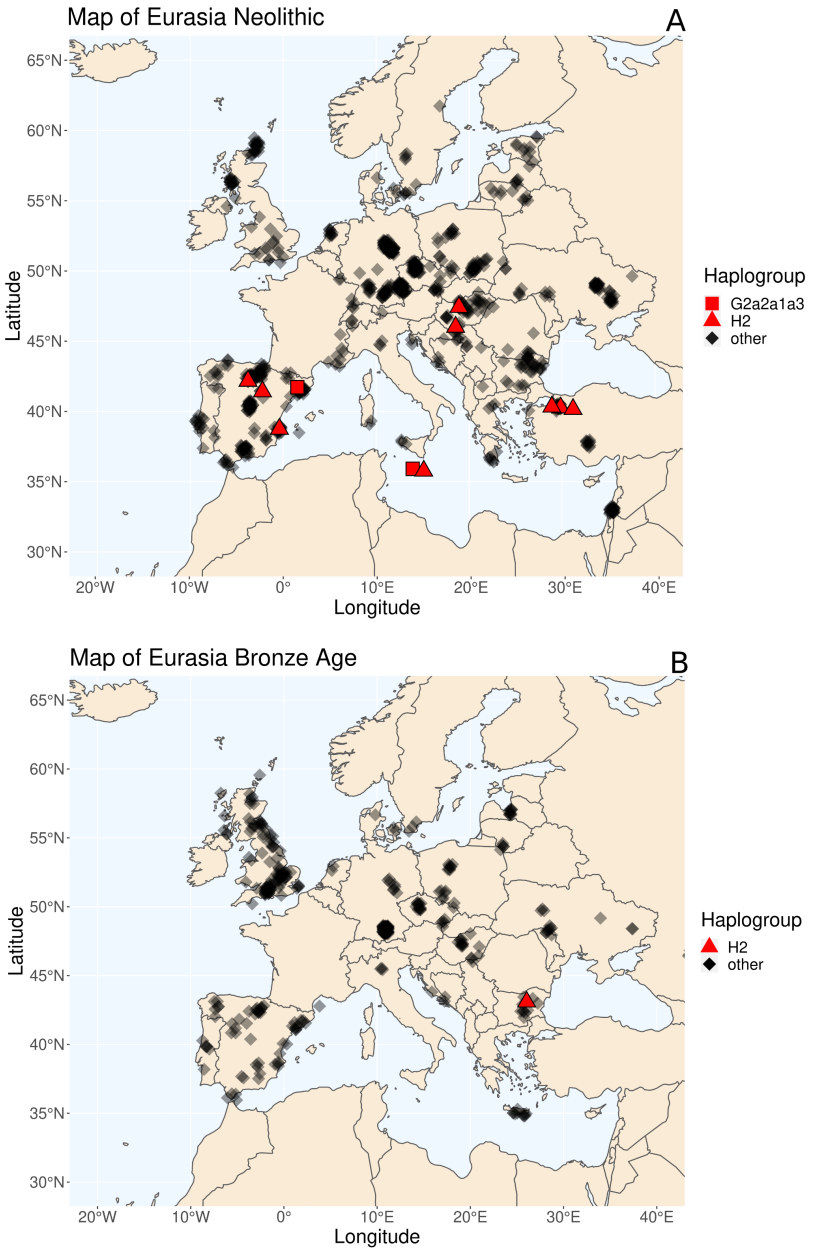


***Figure 2.3. Geographical distribution of ancient Y haplogroups in ancient Eurasia. Each point is a sample with the shape representing the haplogroup****. A red symbol indicates a match with one of the Maltese haplogroups encountered in this work. A) distribution of haplogroups during the Neolithic period. B) distribution of haplogroups in Bronze Age samples.*

### 2.3.3 Principal component analysis

The first thing observable from the PCA result is that the first two principal components (PC1 and PC2) correlate strongly with modern population geographic locations (Figure 2.4). The modern populations used for this analysis consist of 677 individuals from 54 West Eurasian populations. These include population from the Caucasus (such as Abkhasian and Adygei) that fall at the bottom right of the plot; Eastern Europeans (such as Belarusian and Estonian) that fall on the middle right of the PCA; Western Europeans (such as French and Basque) that are located on the middle right of the plot and Southern Europeans (like Croatian and Greek) that are instead at the bottom-centre and bottom-right of the plot. In line with previous findings (eg. (Lazaridis *et al.*, 2016)), the ancient populations can be divided into four major clusters with the shape of a quadrangle. The WHG group is located in a leftward position compared to the modern variation while the Mesolithic individual from Russia, in line with a previous finding (Haak *et al.*, 2015), falls in a top left direction. The CHG population resides in a separate corner of the quadrangle cluster, together with other modern populations from the Caucasus. The Early European Farmers (EEF), as already described in a previous work (Gamba *et al.*, 2014), cluster tightly with the modern Sardinian population. This behaviour has been explained by the fact that modern Sardinians possess a higher Neolithic-farmer genetic component compared to other modern European populations (Skoglund *et al.*, 2012). This could be due to an isolation by distance of this population from further incursion, after an initial Neolithic settlement (Sikora *et al.*, 2014). From the PCA the ancient Neolithic assembly can be further divided into three minor population subgroups: a south-west group formed by Iberian populations, a middle group with prevailing populations from Hungary, Germany, and Italy, and a group from south-west Asia comprising ancient Western Anatolian and Peloponnese farmers. The ancient Maltese individuals are part of the middle group with the individuals Xaghra5, Xaghra6 and Xaghra9 nearly indistinguishable from each other. Unfortunately, due to low coverage, the other ancient Maltese individuals could not be included.

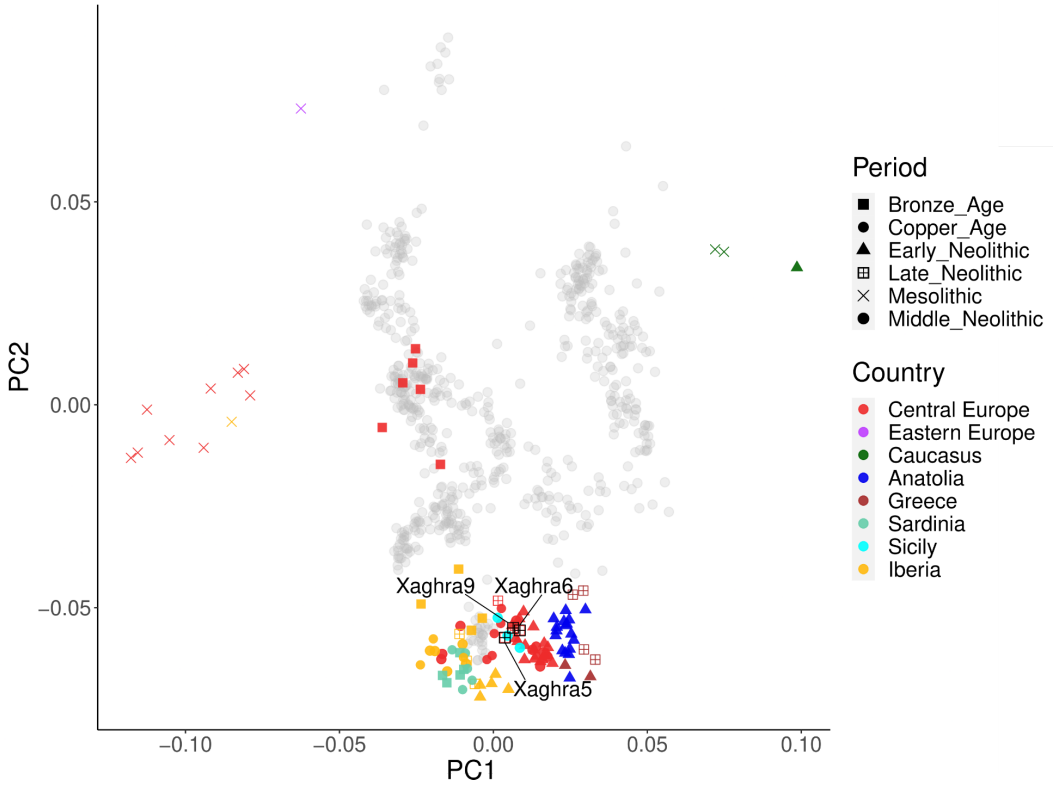


***Figure 2.4: PCA for ancient Maltese and other ancient European populations.*** *Maltese genomes as well as other published Neolithic and Mesolithic genomes were projected onto a principal component analysis plot using the Human Origins dataset as a modern reference. The ancient Maltese samples group together with Central and Southern European Neolithic genomes.*

### 2.3.4 Formal admixture tests

Malta was one of the final regions of Europe to be inhabited, with little evidence of human presence prior to the arrival of Neolithic communities, which were established on the archipelago by 5500 cal. BC (McLaughlin *et al.*, 2020). These were associated with a developed style of impressed pottery (Għar Dalam ware) that represented a regional variant of Sicilian and southwestern Italian ware. Accordingly, we find the genomes from Xagħra Circle share highest levels of drift with the Early Neolithic populations Italy and Greece, followed by Middle

isNeolithic and Chalcolithic populations from Italy and Sicily, as estimated using outgroup $f_3$-statistics (Figure 2.6; Table 2.S4).

Levels of WHG admixture have been shown to vary across European Neolithic samples(Gamba *et al.*, 2014; Skoglund, Malmström, *et al.*, 2014; Haak *et al.*, 2015; Mathieson *et al.*, 2015; Cassidy *et al.*, 2016; Seguin-Orlando *et al.*, 2021) particularly through time. To examine levels of WHG ancestry within our Neolithic sample we applied the qpAdm method to each site, binning genomes into 500 year intervals. We observe WHG ancestral components to increase significantly with time (Figure 2.5 and Table S5; $r^2$=-0.52, p-value = $2.8 \times 10^{-4}$). Interestingly the Xagħra Circle site shows the lowest amount of hunter-gatherer ancestry (6.8 ± 2.5%) among other groups from the later Neolithic (Figure 2.5 and Table 2.S5). This may reflect a shielding by its island context from the dissemination of admixtures with persisting WHG populations that widely influenced mainland populations and which have been estimated to occur as late as 3800 BC (Seguin-Orlando *et al.*, 2021). Using D-statistics , we also tested for gene flow related to North African, CHG and Neolithic Iranian farmers and Yamnaya-steppe groups into the Maltese populations, to the exclusion of the Greek and Italian Early Neolithic (Table 2.S1-3). In all these cases no significant results were obtained.

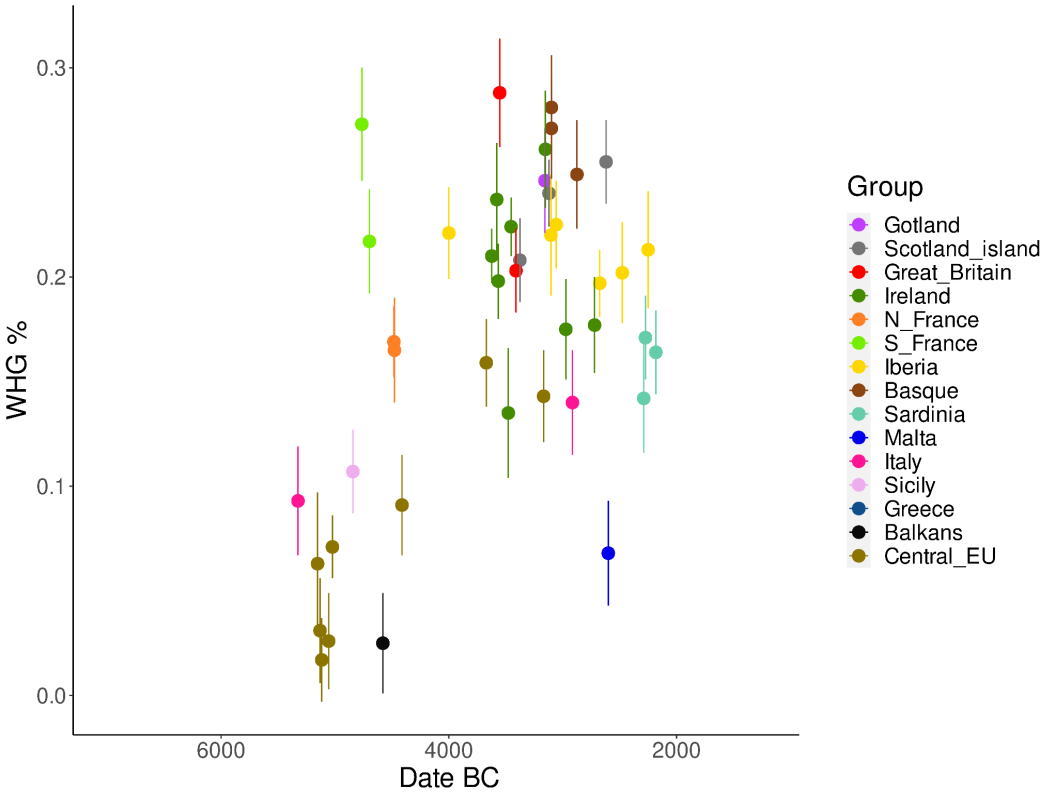


***Figure 2.5: Temporal distribution of hunter-gatherer ancestry in Neolithic Europeans.*** *qpAdm was used to measure the quantity of WHG populations, represented by the individuals KO1 and Loschbour, present in European Neolithic populations. Each point represents a group with at least 2 individuals from an archaeological site and time period. The WHG percentages are reported with 95% confidence intervals.*

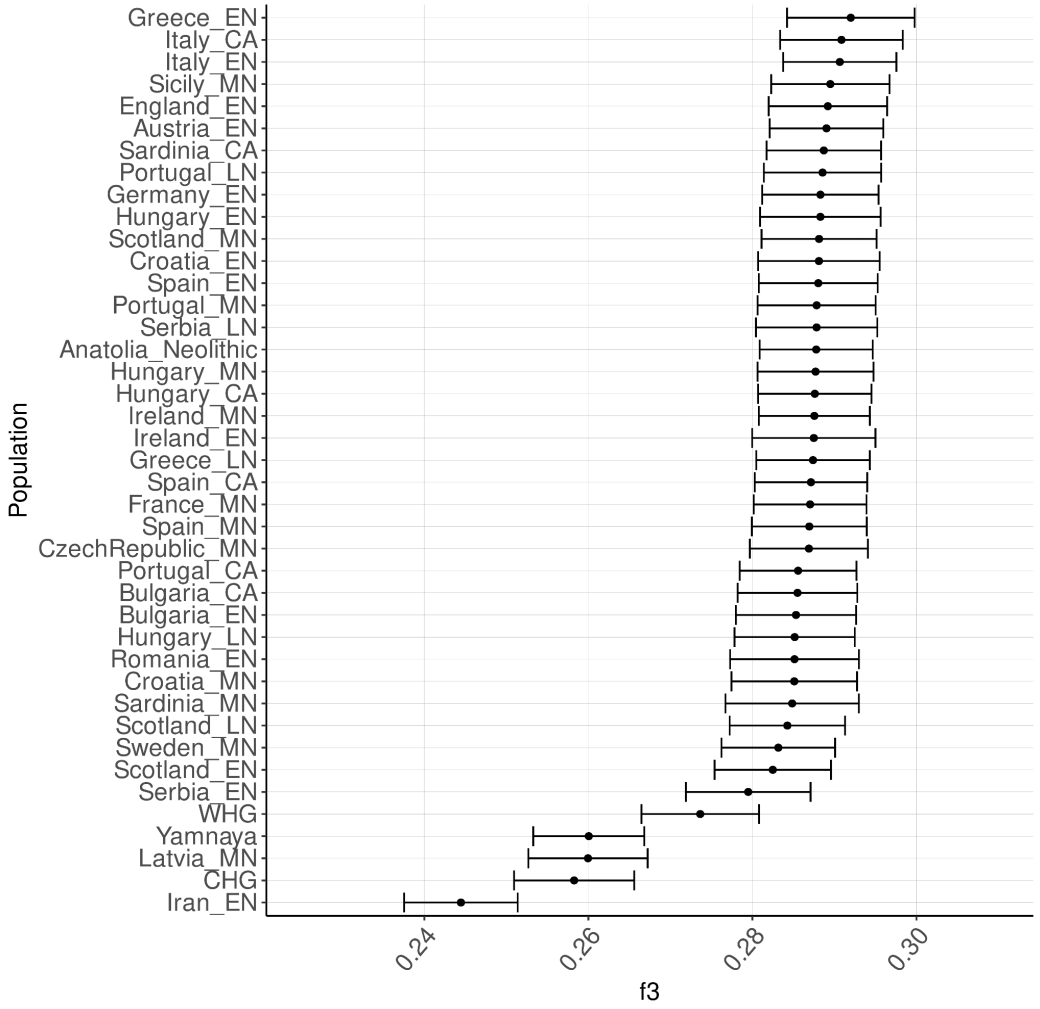


***Figure 2.6: Shared drift measured using the outgroup-f3 statistics.*** *The results using outgroup-f3 statistics in the form (Mbuti; X, Malta Neolithic) show the Maltese being closer to early Neolithic individuals from Greece and the Italian peninsula. (EN=Early Neolithic, MN=Middle Neolithic, LN=Late Neolithic, CA=Copper Age, WHG=Western hunter-gatherer, CHG=Caucasus hunter-gatherer, EHG=Eastern hunter-gatherer). The Anatolian Neolithic population is represented by individuals from Marmara, Barcın.*

## 2.4 Discussion

Ancient Maltese individuals lived during the Late Neolithic period were analysed using different population genetics methods. The aim was to investigate the impact that Neolithic and Hunter-Gatherer European populations had on the Maltese individuals. This analysis used 9 individual bones of which, five were petrosal and four were teeth. An initial screening of these samples confirmed that overall the petrous bone yielded a higher quantity of endogenous DNA compared to teeth. Given their high quality, three of these petrous bones were then retained for further sequencing and analyses. A preliminary analysis on the Y-chromosome proved that two of these three individuals (Xaghra5 and Xaghra9) were male, while the other was female. Y chromosome haplogroup information showed that Xaghra5 and Xaghra9 fall both within the European Neolithic haplogroup variation, with in particular Xaghra9 being more similar to the Italian Copper Age Otzi and Xaghra5 to a Linear Pottery lineage from Central Europe. The mitochondrial haplogroup confirmed these findings with Xaghra6 in particular belonging to the same branch as LBK individuals.

To explore the structure of our samples we projected its genetic variation, together with other known ancient samples, onto modern and European individuals using PCA. In this, the Maltese individuals cluster together with both present-day Sardinians and with Central and Southern European Neolithic populations. This result was validated then by an allele sharing analysis that showed an affinity between the Neolithic Maltese and populations from Sicily and Greece living in the same time period.

Given that our individuals lived in a period of transition between Neolithic and Bronze Age we also tested for an early hunter-gatherer incursion in Malta from Eastern Europe. However, genetic results did not show any significant contribution in our individuals from this population, highlighting instead a WHG influence. Surprisingly, the qpAdm test showed that the amount of WHG influence in our individuals was not as high as other European populations that lived in the same time period. An interesting feature of the European Neolithic, described in Central and Western European populations (Haak *et al.*, 2015; Lipson *et al.*, 2017; Rivollat *et al.*, 2020) is that later samples tend to show an increase in local hunter-gatherer

ancestry compared to initial colonists. This phenomenon can occur almost 20 centuries later. This points toward survival of hunter-gatherer communities, despite profound landscape-altering influxes of early farmers and their later incorporation into mating networks. This is not universal, e.g. in Britain and Ireland the local Mesolithic communities seem to have had little influence on later Neolithic ancestries (Brace *et al.*, 2019; Cassidy *et al.*, 2020). Our results clearly indicate a clear temporal trend of growing WHG influence within Neolithic populations across a broad European sample. A first explanation for the exception of the Maltese Neolithic from this could be an insular isolation of the Maltese populations from the continental shared influences due to their island context and the strong likelihood that hunter-gatherers could not have survived in a parallel society within such a small locale. Note that an early and strong finding of population genomics was the isolation of island Sardinian populations from the later major Bronze Age continent-wide migrations.

Given the high genetic affinity between our Neolithic Maltese we finally performed a kinship test using IBD estimation. The result did not highlight any particular family relationship between the individuals.

Finally we investigated whether our Maltese samples had recent African ancestry. This question was moved by the vicinity between Malta and the northern African coasts. In fact only 288 km span between Malta and the coast of Tunisia making the trip by sea a possibility.

## 2.5 Conclusion

The populations of the Maltese islands, located in the south of the Mediterranean Sea, were shaped by a succession of different cultures during the Neolithic period. The first group settled on the islands just after 6000 BC, probably as an Early Neolithic population. After an initial oscillation between growth and decline an apogee of culture and population density was reached during the Temple Period,

especially in the Tarxien phase between c. 2800 and 2400 bc, which saw the construction of unparalleled sophisticated megalithic structures. Then this culture seemingly collapsed, and a number of questions have vexed scholars of early Malta ever since: who were these ancient inhabitants of Malta, and which ancient population did they resemble the most? To answer these questions, we offer here a first assessment of Maltese ancient DNA data using three individuals that lived during the Tarxien phase of the Temple Period.

The culture of Neolithic farming spread from north-west Anatolia into western Europe following two main routes. One route was associated with the LBK and followed the Danube valley and spread north west towards northern Europe. The other route was associated with Impressa-Cardial pottery culture and followed a westward Mediterranean route reaching the Atlantic in France and Iberia. Shared drift analysis highlighted the vicinity of the Maltese population to other Mediterranean ones such as Greece and Sicily indicating a route through the coast similar to what was observed for the Cardial culture.

By the second millennium bc, the Bronze Age period populations from the steppe migrated from eastern to western Europe, displacing preceding local cultures (Olalde *et al.*, 2018). Exotic pottery coming from eastern Europe, even before the Bronze Age period, could suggest a connection between Maltese and other populations (Thermi for example, Beakers and the potential Balkan Cetina style). However, our shared drift results showed no genetic evidence of our samples mixing with eastern hunter-gatherer populations. Interestingly, these late Neolithic Maltese are outliers when compared with their continental contemporaries for WHG ancestry components. They have not experienced the resurgence of hunter-gatherer ancestry visible elsewhere, a result of the barrier to exchange from their island context and the prohibition from its small population carrying capacity to survival of a parallel WHG society.

# Supplementary Tables

| Outgroup | Pop1 | Pop2 | Pop3 | D | stderr | Z-score | BABA | ABBA | nsnps |
|---|---|---|---|---|---|---|---|---|---|
| Mbuti | EHG | Malta_LN | Austria_EN | -0.0046 | 0.004985 | -0.922 | 9604 | 9692 | 176279 |
| Mbuti | EHG | Malta_LN | Croatia_EN | -0.0129 | 0.005927 | -2.178 | 7953 | 8161 | 148353 |
| Mbuti | EHG | Malta_LN | Croatia_MN | -0.011 | 0.006147 | -1.793 | 8395 | 8582 | 154670 |
| Mbuti | EHG | Malta_LN | Czech_MN | -0.0033 | 0.005388 | -0.62 | 9404 | 9467 | 172710 |
| Mbuti | EHG | Malta_LN | Germany_EN | -0.0115 | 0.004969 | -2.314 | 10150 | 10386 | 187435 |
| Mbuti | EHG | Malta_LN | Greece_EN | -0.0078 | 0.006047 | -1.285 | 8785 | 8923 | 163693 |
| Mbuti | EHG | Malta_LN | Greece_LN | -0.0035 | 0.004904 | -0.722 | 10158 | 10230 | 186577 |
| Mbuti | EHG | Malta_LN | Hungary_EN | -0.0178 | 0.006586 | -2.697 | 7604 | 7879 | 141155 |
| Mbuti | EHG | Malta_LN | Hungary_LN | 0.002 | 0.006283 | 0.32 | 9521 | 9482 | 173043 |
| Mbuti | EHG | Malta_LN | Hungary_MN | -0.0067 | 0.004792 | -1.407 | 10190 | 10328 | 187438 |
| Mbuti | EHG | Malta_LN | Italy_CA | -0.0036 | 0.005104 | -0.715 | 10175 | 10250 | 187335 |
| Mbuti | EHG | Malta_LN | Italy_EN | -0.0042 | 0.004813 | -0.877 | 10133 | 10219 | 186291 |
| Mbuti | EHG | Malta_LN | Sicily_MN | 0.0018 | 0.005833 | 0.314 | 9495 | 9461 | 174081 |
| Mbuti | EHG | Malta_LN | Anatolia_Neolithic | -0.0105 | 0.00446 | -2.353 | 10131 | 10346 | 187438 |
| Mbuti | Yamnaya | Malta_LN | Austria_EN | 0.005 | 0.00341 | 1.474 | 10403 | 10299 | 188724 |
| Mbuti | Yamnaya | Malta_LN | Croatia_EN | 0.0008 | 0.00396 | 0.205 | 8813 | 8799 | 160666 |
| Mbuti | Yamnaya | Malta_LN | Croatia_MN | 0.0017 | 0.004417 | 0.385 | 9213 | 9181 | 166565 |
| Mbuti | Yamnaya | Malta_LN | Czech_MN | 0.0009 | 0.003883 | 0.244 | 10071 | 10051 | 184247 |
| Mbuti | Yamnaya | Malta_LN | Germany_EN | -0.0001 | 0.003237 | -0.023 | 10838 | 10839 | 197917 |
| Mbuti | Yamnaya | Malta_LN | Greece_EN | 0.0001 | 0.004519 | 0.016 | 9370 | 9369 | 173021 |
| Mbuti | Yamnaya | Malta_LN | Greece_LN | 0.0035 | 0.003337 | 1.04 | 10835 | 10760 | 197198 |
| Mbuti | Yamnaya | Malta_LN | Hungary_CA | 0.0093 | 0.003054 | 3.04 | 10851 | 10651 | 195221 |
| Mbuti | Yamnaya | Malta_LN | Hungary_EN | 0.0007 | 0.004621 | 0.142 | 8377 | 8366 | 152514 |
| Mbuti | Yamnaya | Malta_LN | Hungary_LN | 0.0031 | 0.004412 | 0.694 | 10125 | 10063 | 183793 |
| Mbuti | Yamnaya | Malta_LN | Hungary_MN | 0.0006 | 0.003481 | 0.173 | 10843 | 10830 | 197922 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Mbuti | Yamnaya | Malta_LN | Italy_CA | 0.003 | 0.003434 | 0.888 | 10809 | 10744 | 197822 |
| Mbuti | Yamnaya | Malta_LN | Italy_EN | 0.0014 | 0.003304 | 0.437 | 10760 | 10729 | 196690 |
| Mbuti | Yamnaya | Malta_LN | Sicily_MN | 0.013 | 0.003815 | 3.398 | 10328 | 10064 | 186696 |
| Mbuti | Yamnaya | Malta_LN | Anatolia_Neolithic | -0.0001 | 0.002978 | -0.031 | 10811 | 10813 | 197921 |

***Table 2.S1:*** *D* statistics results in the form of (Mbuti, Yamnaya/EHG, Malta_LN, Neolithic European farmers) investigating introgression of Steppe populations into Late Neolithic Maltese.

| Outgroup | Pop1 | Pop2 | Pop3 | D | stderr | Z-score | BABA | ABBA | nsnps |
|---|---|---|---|---|---|---|---|---|---|
| Mbuti | Iran_EN | Malta_LN | Austria_EN | 0.0044 | 0.003752 | 1.163 | 10326 | 10236 | 190758 |
| Mbuti | Iran_EN | Malta_LN | Croatia_EN | 0.0005 | 0.004545 | 0.12 | 8640 | 8631 | 161047 |
| Mbuti | Iran_EN | Malta_LN | Croatia_MN | 0.0018 | 0.00468 | 0.388 | 9096 | 9063 | 167438 |
| Mbuti | Iran_EN | Malta_LN | Czech_MN | -0.0025 | 0.004373 | -0.564 | 10004 | 10054 | 186805 |
| Mbuti | Iran_EN | Malta_LN | Germany_EN | -0.001 | 0.003691 | -0.278 | 10896 | 10918 | 202620 |
| Mbuti | Iran_EN | Malta_LN | Greece_EN | 0.0034 | 0.004896 | 0.69 | 9428 | 9365 | 176385 |
| Mbuti | Iran_EN | Malta_LN | Greece_LN | 0.0055 | 0.003645 | 1.517 | 10927 | 10807 | 201668 |
| Mbuti | Iran_EN | Malta_LN | Hungary_CA | 0.0008 | 0.00338 | 0.244 | 10759 | 10741 | 199113 |
| Mbuti | Iran_EN | Malta_LN | Hungary_EN | 0.0015 | 0.004764 | 0.311 | 8251 | 8227 | 153073 |
| Mbuti | Iran_EN | Malta_LN | Hungary_LN | -0.005 | 0.004888 | -1.021 | 10025 | 10126 | 186991 |
| Mbuti | Iran_EN | Malta_LN | Hungary_MN | -0.0018 | 0.00358 | -0.49 | 10886 | 10924 | 202626 |
| Mbuti | Iran_EN | Malta_LN | Italy_CA | -0.0013 | 0.00383 | -0.349 | 10837 | 10866 | 202509 |
| Mbuti | Iran_EN | Malta_LN | Italy_EN | 0.0013 | 0.003546 | 0.355 | 10832 | 10805 | 201384 |
| Mbuti | Iran_EN | Malta_LN | Sicily_MN | 0.0003 | 0.00425 | 0.06 | 10098 | 10093 | 188433 |
| Mbuti | Iran_EN | Malta_LN | Anatolia_Neolithic | 0.0013 | 0.003254 | 0.392 | 10911 | 10883 | 202625 |
| Mbuti | CHG | Malta_LN | Austria_EN | 0.0032 | 0.003917 | 0.805 | 10298 | 10233 | 188697 |
| Mbuti | CHG | Malta_LN | Croatia_EN | 0.002 | 0.004943 | 0.409 | 8692 | 8657 | 159202 |
| Mbuti | CHG | Malta_LN | Croatia_MN | 0.0042 | 0.005281 | 0.788 | 9125 | 9049 | 165589 |
| Mbuti | CHG | Malta_LN | Czech_MN | 0.0005 | 0.004678 | 0.101 | 10038 | 10028 | 184775 |
| Mbuti | CHG | Malta_LN | Germany_EN | 0.0007 | 0.00396 | 0.177 | 10930 | 10915 | 200486 |
| Mbuti | CHG | Malta_LN | Greece_EN | 0 | 0.005245 | -0.001 | 9383 | 9383 | 174513 |
| Mbuti | CHG | Malta_LN | Greece_LN | 0.0055 | 0.004034 | 1.354 | 10925 | 10806 | 199540 |
| Mbuti | CHG | Malta_LN | Hungary_CA | -0.0024 | 0.003693 | -0.658 | 10733 | 10785 | 196997 |
| Mbuti | CHG | Malta_LN | Hungary_EN | 0.0054 | 0.005723 | 0.949 | 8310 | 8220 | 151347 |
| Mbuti | CHG | Malta_LN | Hungary_LN | 0.0014 | 0.005237 | 0.274 | 10121 | 10092 | 184988 |
| Mbuti | CHG | Malta_LN | Hungary_MN | -0.0063 | 0.00388 | -1.631 | 10812 | 10950 | 200491 |
| Mbuti | CHG | Malta_LN | Italy_CA | 0.0019 | 0.004264 | 0.446 | 10876 | 10835 | 200371 |
| Mbuti | CHG | Malta_LN | Italy_EN | 0.002 | 0.003881 | 0.519 | 10826 | 10783 | 199255 |
| Mbuti | CHG | Malta_LN | Sicily_MN | 0.0044 | 0.004832 | 0.907 | 10150 | 10062 | 186371 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Mbuti | CHG | Malta_LN | Anatolia_Neolithic | 0.0032 | 0.003543 | 0.897 | 10932 | 10863 | 200490 |

***Table 2.S2:*** *D* statistics analysis in the form (Mbuti, Iran_EN/CHG, Malta_LN, Neolithic European farmers) looking at an excess of allele sharing between ancient Caucasus populations and Neolithic Maltese.

| Outgroup | Pop1 | Pop2 | Pop3 | D | stderr | Z-score | BABA | ABBA | nsnps |
|---|---|---|---|---|---|---|---|---|---|
| Chimp | Morocco_EN | Malta_LN | Austria_EN | 0.003 | 0.00628 | 0.484 | 6191 | 6154 | 101212 |
| Chimp | Morocco_EN | Malta_LN | Croatia_EN | -0.0052 | 0.007679 | -0.679 | 5145 | 5199 | 84902 |
| Chimp | Morocco_EN | Malta_LN | Croatia_MN | 0.009 | 0.007804 | 1.149 | 5495 | 5397 | 88976 |
| Chimp | Morocco_EN | Malta_LN | Czech_MN | -0.0041 | 0.00684 | -0.605 | 6186 | 6238 | 100929 |
| Chimp | Morocco_EN | Malta_LN | Germany_EN | -0.0028 | 0.00609 | -0.467 | 6559 | 6597 | 106841 |
| Chimp | Morocco_EN | Malta_LN | Greece_EN | -0.0027 | 0.008103 | -0.336 | 5708 | 5739 | 93867 |
| Chimp | Morocco_EN | Malta_LN | Greece_LN | 0.0001 | 0.006053 | 0.018 | 6547 | 6545 | 106399 |
| Chimp | Morocco_EN | Malta_LN | Hungary_CA | -0.0028 | 0.005845 | -0.476 | 6450 | 6486 | 105230 |
| Chimp | Morocco_EN | Malta_LN | Hungary_EN | -0.0062 | 0.007699 | -0.809 | 6044 | 6120 | 97864 |
| Chimp | Morocco_EN | Malta_LN | Hungary_LN | -0.0032 | 0.008023 | -0.401 | 6094 | 6133 | 98684 |
| Chimp | Morocco_EN | Malta_LN | Hungary_MN | -0.0085 | 0.006055 | -1.403 | 6507 | 6619 | 106845 |
| Chimp | Morocco_EN | Malta_LN | Italy_CA | 0.0047 | 0.00675 | 0.695 | 6549 | 6488 | 106782 |
| Chimp | Morocco_EN | Malta_LN | Italy_EN | -0.0092 | 0.005973 | -1.545 | 6442 | 6562 | 106241 |
| Chimp | Morocco_EN | Malta_LN | Sicily_MN | -0.0049 | 0.006723 | -0.73 | 6053 | 6112 | 99210 |
| Chimp | Morocco_EN | Malta_LN | Anatolia_Neolithic | -0.0023 | 0.005529 | -0.419 | 6528 | 6558 | 106841 |

***Table 2.S3****: D* statistics results in the form (Mbuti, Morocc_EN, Malta_LN, Neolithic European farmers) investigating the presence of ancient North African introgression in Neolithic Maltese.

| Outgroup | Pop1 | Pop2 | f3 | stderr | Z-score | nsnps |
|---|---|---|---|---|---|---|
| Mbuti | Malta_LN | Greece_EN | 0.291988 | 0.003887 | 75.119 | 134132 |
| Mbuti | Malta_LN | Italy_EN | 0.290841 | 0.003466 | 83.914 | 164149 |
| Mbuti | Malta_LN | Italy_CA | 0.289754 | 0.003578 | 80.994 | 159917 |
| Mbuti | Malta_LN | Sicily_MN | 0.289496 | 0.003606 | 80.276 | 146627 |
| Mbuti | Malta_LN | England_EN | 0.289191 | 0.003609 | 80.122 | 160163 |
| Mbuti | Malta_LN | Austria_EN | 0.289023 | 0.003465 | 83.414 | 154593 |
| Mbuti | Malta_LN | Sardinia_CA | 0.288699 | 0.003492 | 82.679 | 163986 |
| Mbuti | Malta_LN | Portugal_LN | 0.288545 | 0.003576 | 80.7 | 158105 |
| Mbuti | Malta_LN | Germany_EN | 0.288285 | 0.003547 | 81.281 | 163414 |
| Mbuti | Malta_LN | Hungary_EN | 0.288282 | 0.003676 | 78.425 | 116725 |
| Mbuti | Malta_LN | Scotland_MN | 0.288123 | 0.003508 | 82.144 | 165866 |
| Mbuti | Malta_LN | Croatia_EN | 0.288104 | 0.003705 | 77.752 | 125356 |
| Mbuti | Malta_LN | Spain_EN | 0.288023 | 0.003623 | 79.502 | 156904 |
| Mbuti | Malta_LN | Portugal_MN | 0.287827 | 0.003597 | 80.02 | 157583 |
| Mbuti | Malta_LN | Serbia_LN | 0.287822 | 0.003697 | 77.86 | 127691 |
| Mbuti | Malta_LN | Hungary_MN | 0.287695 | 0.003533 | 81.424 | 163571 |
| Mbuti | Malta_LN | Hungary_CA | 0.287603 | 0.003455 | 83.252 | 165226 |
| Mbuti | Malta_LN | Ireland_MN | 0.287544 | 0.003382 | 85.019 | 177670 |
| Mbuti | Malta_LN | Anatolia_Neolithic | 0.28752 | 0.003427 | 83.898 | 171217 |
| Mbuti | Malta_LN | Ireland_EN | 0.287478 | 0.003757 | 76.525 | 156773 |
| Mbuti | Malta_LN | Greece_LN | 0.287382 | 0.00346 | 83.059 | 160891 |
| Mbuti | Malta_LN | Spain_CA | 0.287134 | 0.003428 | 83.768 | 170304 |
| Mbuti | Malta_LN | France_MN | 0.287029 | 0.003438 | 83.479 | 162027 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Mbuti | Malta_LN | Spain_MN | 0.286929 | 0.003499 | 82.012 | 162155 |
| Mbuti | Malta_LN | CzechRepublic_MN | 0.286885 | 0.003597 | 79.759 | 150487 |
| Mbuti | Malta_LN | Portugal_CA | 0.285564 | 0.003556 | 80.302 | 157565 |
| Mbuti | Malta_LN | Bulgaria_CA | 0.285492 | 0.003644 | 78.356 | 119132 |
| Mbuti | Malta_LN | Bulgaria_EN | 0.285311 | 0.003665 | 77.846 | 116543 |
| Mbuti | Malta_LN | Hungary_LN | 0.285147 | 0.003663 | 77.851 | 142729 |
| Mbuti | Malta_LN | Romania_EN | 0.285127 | 0.003925 | 72.635 | 111168 |
| Mbuti | Malta_LN | Croatia_MN | 0.285097 | 0.003824 | 74.551 | 128596 |
| Mbuti | Malta_LN | Sardinia_MN | 0.284842 | 0.004064 | 70.088 | 87880 |
| Mbuti | Malta_LN | Scotland_LN | 0.284255 | 0.003517 | 80.826 | 139846 |
| Mbuti | Malta_LN | Sweden_MN | 0.283154 | 0.003464 | 81.736 | 161545 |
| Mbuti | Malta_LN | Scotland_EN | 0.282478 | 0.003543 | 79.724 | 144538 |
| Mbuti | Malta_LN | Serbia_EN | 0.279485 | 0.003802 | 73.503 | 124804 |
| Mbuti | Malta_LN | WHG | 0.273637 | 0.003587 | 76.29 | 162336 |
| Mbuti | Malta_LN | Yamnaya | 0.260051 | 0.003383 | 76.877 | 158793 |
| Mbuti | Malta_LN | EHG | 0.2588 | 0.003866 | 66.944 | 129625 |
| Mbuti | Malta_LN | CHG | 0.258264 | 0.00366 | 70.56 | 153508 |

***Table 2.S4****: f3*-statistics results in the form (Mbuti, Malta_LN, Neolithic European farmers) investigating the shared drift between the Maltese and other European Neolithic populations.

| Site | Group_label | Lower_BC | Upper_BC | NW-A-Neolithic | WHG | stderr_NW-A-Neolithic | stderr_WHG | nsnps | pvalue |
|---|---|---|---|---|---|---|---|---|---|
| Abony-Turjányos_dűlő | Central_EU | 2900 | 3900 | 0.841 | 0.159 | 0.021 | 0.021 | 104836 | 0.105009 |
| Alepotrypa_Cave | Greece | 3900 | 4900 | 0.95 | 0.05 | 0.025 | 0.025 | 101825 | 0.0235678 |
| Alto_de_la_Huesera | Basque | 1900 | 2900 | 0.751 | 0.249 | 0.026 | 0.026 | 97984 | 0.265037 |
| Anghelu_Ruju | Sardinia | 1900 | 2900 | 0.858 | 0.142 | 0.026 | 0.026 | 87562 | 0.856875 |
| Annagh | Ireland | 2900 | 3900 | 0.865 | 0.135 | 0.031 | 0.031 | 96853 | 0.948022 |
| Ansarve | Gotland | 2900 | 3900 | 0.754 | 0.246 | 0.025 | 0.025 | 113485 | 0.0668167 |
| Ashleypark | Ireland | 2900 | 3900 | 0.763 | 0.237 | 0.027 | 0.027 | 104530 | 0.16707 |
| Bataszek-Lajver | Central_EU | 4900 | 5900 | 0.969 | 0.031 | 0.025 | 0.025 | 95253 | 0.456843 |
| Budakalász-Luppa_csárda | Central_EU | 2900 | 3900 | 0.857 | 0.143 | 0.022 | 0.022 | 100508 | 0.561269 |
| Cabeço_da_Arruda | Iberia | 2900 | 3900 | 0.78 | 0.22 | 0.029 | 0.029 | 102843 | 0.13301 |
| Carrowkeel | Ireland | 1900 | 2900 | 0.823 | 0.177 | 0.023 | 0.023 | 110368 | 0.261716 |
| Carrowkeel | Ireland | 2900 | 3900 | 0.825 | 0.175 | 0.024 | 0.024 | 111433 | 0.910768 |
| Clos_de_Roque | South_France | 3900 | 4900 | 0.783 | 0.217 | 0.025 | 0.025 | 96130 | 0.732537 |
| Cova_Moura | Iberia | 1900 | 2900 | 0.787 | 0.213 | 0.028 | 0.028 | 110633 | 0.480496 |
| Distillery_Cave | Great_Britain | 2900 | 3900 | 0.712 | 0.288 | 0.026 | 0.026 | 93848 | 0.465808 |
| El_Portalón | Iberia | 2900 | 3900 | 0.775 | 0.225 | 0.021 | 0.021 | 112693 | 0.775002 |
| Fleury-sur-Orne | North_France | 3900 | 4900 | 0.835 | 0.165 | 0.025 | 0.025 | 94097 | 0.542038 |
| Fossato_di_Stretto_Partana | Sicily_Neolithic | 3900 | 4900 | 0.893 | 0.107 | 0.02 | 0.02 | 108160 | 0.429442 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Grotta_Continenza | Italy_Neolithic | 2900 | 3900 | 0.86 | 0.14 | 0.025 | 0.025 | 112160 | 0.966931 |
| Grotta_Continenza | Italy_Neolithic | 4900 | 5900 | 1 | 0 | 0.018 | 0.018 | 110790 | 0.0240964 |
| Gurgy_les_Noisats | North_France | 3900 | 4900 | 0.831 | 0.169 | 0.017 | 0.017 | 105624 | 0.59994 |
| Halberstadt-Sonntagsfeld | Central_EU | 4900 | 5900 | 0.974 | 0.026 | 0.023 | 0.023 | 106968 | 0.77356 |
| Holm_of_Papa | Scotland_island | 2900 | 3900 | 0.792 | 0.208 | 0.02 | 0.02 | 101887 | 0.446121 |
| Isbister | Scotland_island | 1900 | 2900 | 0.745 | 0.255 | 0.02 | 0.02 | 100222 | 0.764964 |
| Isbister | Scotland_island | 2900 | 3900 | 0.76 | 0.24 | 0.016 | 0.016 | 102285 | 0.649799 |
| Jentillarri | Basque | 2900 | 3900 | 0.719 | 0.281 | 0.025 | 0.025 | 94749 | 0.317461 |
| Ke_Stirce_Street | Central_EU | 3900 | 4900 | 0.909 | 0.091 | 0.024 | 0.024 | 96554 | 0.18725 |
| Les_Bréguières | S_France | 3900 | 4900 | 0.727 | 0.273 | 0.027 | 0.027 | 90115 | 0.628982 |
| Lorga_de_Dine | Iberia | 1900 | 2900 | 0.798 | 0.202 | 0.024 | 0.024 | 111776 | 0.107501 |
| Lugar_Canto | Iberia | 3900 | 4900 | 0.779 | 0.221 | 0.022 | 0.022 | 112399 | 0.99497 |
| Mandubi_Zelaia | Basque | 2900 | 3900 | 0.729 | 0.271 | 0.024 | 0.024 | 96051 | 0.0463776 |
| Newgrange | Ireland | 2900 | 3900 | 0.739 | 0.261 | 0.028 | 0.028 | 113741 | 0.707167 |
| Parknabinnia | Ireland | 2900 | 3900 | 0.776 | 0.224 | 0.014 | 0.014 | 107465 | 0.587605 |
| Polgár-Ferenci-hát | Central_EU | 4900 | 5900 | 0.937 | 0.063 | 0.034 | 0.034 | 114546 | 0.2898 |
| Poulnabrone | Ireland | 2900 | 3900 | 0.79 | 0.21 | 0.013 | 0.013 | 106976 | 0.276698 |
| Primrose_Grange | Ireland_ | 2900 | 3900 | 0.802 | 0.198 | 0.018 | 0.018 | 110463 | 0.157141 |
| Raschoille_Cave | Great_Britaini | 2900 | 3900 | 0.797 | 0.203 | 0.02 | 0.02 | 101226 | 0.400973 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ripabiance | Italy | 4900 | 5900 | 0.907 | 0.093 | 0.026 | 0.026 | 95815 | 0.93051 |
| Schletz | Central_EU | 4900 | 5900 | 0.929 | 0.071 | 0.015 | 0.015 | 107326 | 0.739192 |
| Serra_Cabriles | Sardinia | 1900 | 2900 | 0.829 | 0.171 | 0.02 | 0.02 | 99937 | 0.412174 |
| Sima_del_Ángel | Iberia | 1900 | 2900 | 0.803 | 0.197 | 0.016 | 0.016 | 103953 | 0.690415 |
| Spain_EN_Cardial | Iberia | 4900 | 5900 | 0.885 | 0.115 | 0.026 | 0.026 | 112104 | 0.00258626 |
| Stuttgart-Mühlhausen | Central_EU | 4900 | 5900 | 0.983 | 0.017 | 0.02 | 0.02 | 113364 | 0.558464 |
| Su_Crocefissu | Sardinia | 1900 | 2900 | 0.836 | 0.164 | 0.02 | 0.02 | 107729 | 0.403098 |
| Vojvodina_Hrtkovci_Gomolova | Balkans | 3900 | 4900 | 0.975 | 0.025 | 0.024 | 0.024 | 100588 | 0.184864 |
| Xagħra_Circle | Malta | 1900 | 2900 | 0.932 | 0.068 | 0.025 | 0.025 | 109883 | 0.0513541 |

***Table 2.S5****: qpAdm* results investigating the amount of WHG in Neolithic European populations. The reference group consists of the following populations: Mbuti.DG, Ust-Ishim, Mal'ta, Villabruna, GoyetQ116, Han.DG, Papuan.DG, Mixe.DG, Kartiana.DG, AHG, CHG, EHG. The source group consist of the following populations: North Western Anatolia Neolithic(NW-A-Neolithic), WHG

| Site | Group_label | Lower_BC | Upper_BC | NW-A-Neolithic | WHG | Iran_Neolithic | stderr_NW-A-Neolithi | stderr_WHG | stderr_Iran_Neolithic | nsnps_used | pvalue |
|---|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | c | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Alepotrypa_Cave | Greece | 3900 | 4900 | 0.855 | 0.055 | 0.089 | 0.054 | 0.025 | 0.045 | 97705 | 0.0604 |
| Grotta_Continenza | Italy | 4900 | 5900 | 0.882 | 0.006 | 0.112 | 0.038 | 0.018 | 0.033 | 105922 | 0.42 |

***Table 2.S6****: qpAdm* results for test in Table S4 with significant p-values ($p<0.05$). In this test the population Iran_Neolithic was also included.

# 3. Inbreeding and finescale population structure in ancient Europe

## 3.1 Introduction

### 3.1.1 Genotype imputation as a new frontier

Even though the methods of DNA extraction and sequencing have made great advancement in recent years, researchers often have to tackle the problem of missing genotype information when analysing ancient samples. To help with this issue few methods have been developed that use advanced statistical approaches to predict the state of missing genotypes. These genotype imputation methods take advantage of the linkage that exists between genomic loci. Linked regions or haplotypes, segregate among members of a population as genomic units with common ancestry. An obvious context where these methods were first applied was within families with known pedigree information. This was relatively straightforward because close family members who have a recent common ancestor also share long stretches of haplotypes that can be traced through generations (Y. Li *et al.*, 2009). The approach becomes more challenging when the genomes to impute are from individuals of unknown ancestry and we can not use the pedigree to keep track of the shared haplotypes. Several *in silico* methods have recently been developed to solve this issue using advanced statistical methods. These use known short haplotypes and recombination maps from a reference source to reconstruct the unknown genotype information in target unknown populations (Browning, 2006; LI and Y, 2006; Marchini *et al.*, 2007).

Given the incredible diversity of combinations of short haplotypes that are generated these methods usually have the drawback of long computational times to impute large datasets featuring millions of variants and hundreds of individuals. For example, the software Impute2 imputes around 25 thousands SNPs in 60

hours. One way to ameliorate this problem consists of reducing the number of haplotypes present in the reference dataset by selecting only the ones closest to the target genome. For example using the same set of 25 thousands SNPs it takes only 40 hours when the reference dataset is reduced from 2504 to 504 individuals (Shi *et al.*, 2018).

In addition to consideration of computational time, the selection of a good haplotype reference data set can also affect the quality of the imputed target. For example in most non-model organisms, the absence of a good reference for both haplotypes and recombination map information makes imputing genotypes particularly challenging.To this end a recently developed algorithm, LD-kNNi (Money *et al.*, 2015), tackled this problem by imputing missing genotypes without the use of a reference recombination map or phasing information. Similarly to under-represented species, rare variants are not well imputed even for well annotated organisms given their poor representation in the haplotypes of the reference datasets (Das, Abecasis and Browning, 2018). Finally even within the same species, populations that highly diverge between the reference and the target can give problems during the genotype imputation, lowering its quality and performance.

Despite all these challenges described, genotype imputation has been successfully applied not just in modern but also in ancient humans. For example Gamba and colleagues in 2014 (Gamba *et al.*, 2014) were the first to apply genotype imputation to a small group of ancient individuals. They demonstrated that an early hunter-gatherer individual showed less genetic diversity, as indicated by a higher fraction of the genome under run of homozygosity (ROH), compared to Neolithic farmers, some of whom came from the same archaeological context. Another recent work from the same group (Martiniano *et al.*, 2017) imputed a bigger set of ancient samples spanning from the Palaeolithic to the Anglo-Saxon periods allowing to investigate fine-population structure and phenotypic traits in ancient human populations. With the same purpose Antonio et al. (Antonio *et al.*, 2019) used imputation to investigate the fine-scale structure of ancient Italian genomes that lived between Mesolithic and Early modern time. Thanks to the high resolution provided by the haplotype information the authors of this work could

detect a range of ancestries present during the imperial time and probably following the expansion of the Roman empire through the Mediterranean coasts. More recently Cassidy and colleagues (Cassidy *et al.*, 2020), by applying the same imputation method, discovered a Neolithic sample from a first order incestuous lineage that belonged to an elite dynasty society. In the same year diploid imputed genotypes were used to investigate the expansion of Viking populations from Scandinavia to Europe (Margaryan *et al.*, 2020). Lastly, recent work expanded the range of samples imputed to include also SNP captured samples covering more than 300 individuals (Ariano *et al.*, 2022). In this work the authors investigated population structure, population size and inbreeding for populations ranging from Paleolithic to Neolithic periods. In particular, by using diploid imputed data for 258 Neolithic individuals this work highlighted the migration rate that shaped the genetic architecture of ancient Europe (Table 3.S1).

Although imputation represents a promising tool for increasing the resolution of many analyses it also requires a minimum amount of coverage in a sample in order to obtain a significant accuracy. This problem was partially resolved in a two-step pipeline applied to ultra-low coverage genomes ranging from 0.05 to 1X (Hui *et al.*, 2020). By applying this method the authors of this work achieved an accuracy of approximately 90% for a genome with a coverage as low as 0.05X and 97% for genome with a coverage of at least 0.5X.

### 3.1.2 Haplotype sharing methods

When two individuals share a recent common ancestor, chromosome chunks or shared haplotypes, termed IBD, are passed through generations (Browning, 2008; Browning and Browning, 2010). The size and quantity of these IBD segments can be used in genomics to investigate phenomena such as kinship, demography and phenotypic traits. For example, first cousins share approximately 25% in IBD, with an average length of segments of 25 centiMorgans cM (approximately 25 millions nucleotide bases) (Thompson, 2013). This example illustrates that, when haplotype information is available, it is possible to use the amount and length of

IBD chunks to estimate the probability of two individuals being related at a certain time in the past. After the detection of IBD segments from dense diploid genotype data other tools can be used to infer kinship or to estimate population size (Manichaikul *et al.*, 2010; Browning and Browning, 2015). For example, recently Amy Williams and colleagues developed a series of tools that help understanding how IBD segments pattern can be used to infer genealogies (https://hapi-dna.org). In another example Belbin and colleagues used IBD to discover a variant associated with short stature that was present in high frequency in Puerto Rican individuals due to a founding population event (Belbin *et al.*, 2017).

It was only in recent times, thanks to the use of imputation methods, that haplotype-based approaches such as IBD have been investigated in ancient humans. For example Margaryan and colleagues (Margaryan *et al.*, 2020) analysed the IBD segments of 298 imputed viking genomes to investigate their genetic structure.

### 3.1.3 Inbreeding analysis

When two IBD segments are identical and shared within the same individual they form a long stretch of genome defined as homozygous by descent (HBD). HBD segments are usually studied by detecting long regions of the genomes in homozygosity, also called ROH. Analogous to IBD, ROH segments can be used to infer the size of a population and primarily also to investigate inbreeding events occurring within the genealogy of an individual. As for IBD, the quantity and length of the ROH segments are important to understand the relatedness between the parents of an inbred organism. For example Yengo and colleagues have used the pattern of ROH segments to investigate inbreeding in approximately 450 thousands individuals from the UK Biobank dataset (Yengo, Wray and Visscher, 2019). The results of this work showed that approximately 1/3652 individuals of UK origin were extremely inbred with parents being at least 2nd degree relatives. As with the detection of IBD segments, HBD studies require diploid genomes to be called and thus it has proved difficult to apply to ancient individuals. Thanks to imputation methods recent works have reported surprising results about inbreeding

in the ancient times. For example Cassidy and colleagues (Cassidy *et al.*, 2020) discovered a high quantity of long stretch of ROH segments in an ancient Neolithic Irish individual, marking him as the offspring of a union between first degree relatives.

To date, imputation has been applied to ancient genomes sampled by shotgun methods, with a threshold of at least 0.4X coverage. However, the majority of ancient individuals published have been sequenced using targeted capture using RNA baits homologous with a subset (in later analyses 1.24 million) of variable positions (Haak *et al.*, 2015; Mathieson *et al.*, 2015). These are typically of low coverage and data is usually analysed using tools adapted to single allele calls (pseudohaploid) at available positions. A clear imperative is to investigate whether this large data source may be leveraged by imputation to give diploid genomes. With this purpose in mind other groups imputed the genotypes of a limited set SNP capture ancient samples to improve the estimates of polygenic traits (Marciniak *et al.*, 2021; Cox *et al.*, 2022).

For the project described in this chapter we imputed a large set of ancient SNP capture samples to investigate the fine-scale structure, population size, and inbreeding in ancient Eurasian populations.

A useful step towards understanding inbreeding in ancient populations came with a novel framework developed by Ringbauer and colleagues (Ringbauer, Novembre and Steinrücken, 2021). This software was developed to detect long stretches of homozygous genomes in ancient human individuals without requiring an imputation step and showed utility with SNP capture data. This method, called haproh, works by using a hidden markov model (HMM) with 2 states for the detection of ROH. This HMM was specifically designed to work on low coverage SNP capture data from shotgun genomes, downsampled to include the 1240K SNP positions only. This technique demonstrated a high accuracy in detecting ROH for samples with a low density of SNPs (at least 400K SNPs). However, one limitation that affects this process is that it clusters the ROH segments in bins of defined size, losing the possibility of analysing them on a continuous scale. Despite this limitation, the authors of this method successfully applied it to

investigate inbreeding in 1,785 ancient humans that lived in the last 45000 years. In addition to detecting ROH additional features have been implemented in this method such as detecting the effective population size of a group of individuals using their ROH profile.

## 3.2 Methods

### 3.2.1 Genotype Imputation

From samples that had been screened using an in-solution target capture method 231 published genomes for imputation with a reported coverage on target regions of at least 0.6X and 650K SNPs called from the 1240K panel were selected. To increase the number of samples from Neolithic Sardinia 4 samples with a coverage higher than 0.6X and at least 460K SNPs safely called from the 1240K panel were also included(Table 3.S1). Overall these samples lived in Europe and South-West Asia between early Neolithic and late Copper Age periods. All of the samples have been previously shown to be characterised by two major genetic components: one is the Early Anatolian farmer and the other is WHG ancestry. Before imputation a set of approximately 6.2 million SNPs were chosen to be called on the target dataset using the 1000 Genomes Project (1000 Genomes Project Consortium *et al.*, 2015) resource as reference, filtered for individuals of African origin (defined with the AFR label) and with a minor allele frequency of 5%. Variants were called using the tool UnifiedGenotyper in GATK v2.4 (McKenna *et al.*, 2010) program with parameters (--output_mode EMIT_ALL_SITES, --genotyping_mode GENOTYPE_GIVEN_ALLELES). The VCF files created were then split first by chromosome and then by windows of 1 Mb. Genotype imputation was performed on approximately 28 million variants using the tool Beagle v.4.1. (Browning and Browning, 2007, 2016) with a reference dataset of 1843 modern individuals of non-African origin from the 1000 Genomes project. The program was run in multi-thread mode taking advantage of the Irish Centre for High-End Computing (ICHEC) cluster. The genetic map used was taken from the Beagle website

(http://bochet.gcc.biostat.washington.edu/beagle/genetic_maps/). The imputed VCF files were filtered for SNPs only and genotype probability of 0.95 using bcftools v.1.3 (H. Li *et al.*, 2009) and PLINK v1.9 (--vcf-min-gp 0.95)(Chang *et al.*, 2015) obtaining 25.8 million variants.

After completion of imputation, four samples with high genotype missingness (>=0.1) were removed from subsequent analyses. Separately, thanks to a collaboration with Lara Cassidy and Shyam Gopalakrishnan from respectively Trinity College Dublin and the University of Copenhagen 120 WGS samples were selected with a coverage of at least 0.4X to impute using the Software Impute2 (Howie, Donnelly and Marchini, 2009; Howie, Marchini and Stephens, 2011) Table 3.S1). For these samples, and similarly to the SNP capture imputation, variants were called by our collaborators using the 1000 Genomes project (1000 Genomes Project Consortium *et al.*, 2015) as reference and the tool UnifiedGenotyper in GATK v.2.4(McKenna *et al.*, 2010) using the same parameters described in the previous paragraph. For this particular imputation analysis they used the whole 1000 Genomes project dataset as reference for the genotype imputation. Prior to imputation transition SNPs were excluded from this dataset resulting in calling of approximately 28 million. The VCF file was then split first by chromosome and then in windows containing 15000 markers. For each input file, the program Impute2 was called using the parameters ( -Ne 20000 -buffer 500 -allow_large_regions -k 400 -k_hap 2000). After imputation they filtered for genotype probability higher than 0.99 (GP > 0.99) resulting in 77.8 million SNPs.

Finally this combined resource was merged with 21 low coverage shotgun sequenced genomes which had been previously imputed (Antonio *et al.*, 2019). Between WGS and SNP capture samples a final resource of 357 unique imputed diploid genomes emerged (Table 3.S1).

After merging these resources they were tested for differences in genotype missingness between datasets. This considered a set of 12 million SNPs common across all three datasets. The missingness for each dataset was calculated and averaged across samples. Genotype missing percentages for the SNP capture and WGS imputed data were 12% and 13.5% respectively.

### 3.2.2 ROH and inbreeding analysis

To estimate the inbreeding coefficients of the imputed samples, a measure based on the proportion of the genome that is homozygous-by-descent (runs of homozygosity that are identical by descent), as employed in (Cassidy *et al.*, 2020), and labelled here as $F_{ROH}$was used. Separately, the hunter-gatherer and Neolithic farmer datasets were filtered for genotypes missingness and minor allele frequency using PLINK v1.9 (--geno 0.02, --maf 0.05, --indep 50 2 2) obtaining respectively 51,289 SNPs and 41,426 SNPs. Using this set of SNPs ROH segments were identified using PLINK v1.9. with the same parameters used in (Gazal *et al.*, 2014) (--homozyg-window-het 0 --homozyg-snp 50 --homozyg-kb 1 --homozyg-density 5000 --homozyg-gap 5000). Physical measures were converted to cM. The total length of the genome in ROH above this threshold divided by the length of the autosomal part was used to estimate the $F_{ROH}$ coefficients (McQuillan *et al.*, 2008). To assess the concordance between samples imputed from different sources $F_{ROH}$ estimates obtained from imputed SNP capture were compared with those calculated using imputed data from WGS data available for the same samples. The same set of SNPs were considered in both data types. For two samples that were whole genome screened and where the coverage was sufficiently high $F_{ROH}$ coefficients were also estimated using diploid genotype calls. For these two the same protocol was applied as described in the imputation accuracy paragraph. In brief, diploid genotypes with a depth below 10 or higher than 30 and a quality below 30 were excluded. As shown in Figure 3.2 there is no visible deviation of substance between the measures.

For the sample, Xaghra9, which has sufficient coverage, the software ROHan (Renaud *et al.*, 2019) was run to validate the inbreeding results. As suggested software we used the program bam2prof with different threshold values (--length 5, 10, 15, 20) to account for post-mortem deamination damage. We then ran the program ROHan using transversions only (--tvsonly).

### 3.2.3 Pedigree simulation

To better understand the degree of relatedness between the parents of inbred samples different pedigree scenarios were simulated using dummy genotypes. This started from the same dataset described in the previous paragraph with filtering for genotype missingness and minor allele frequency. This filtered resource was first split by chromosome and then phased using SHAPEIT v.2.r837(Delaneau, Coulonges and Zagury, 2008). After phasing I filtered for linkage disequilibrium with plink using (--indep 50 2 2) and selected a common set of SNPs. 21 Irish imputed individuals published in (Cassidy *et al.*, 2020) were selected from this dataset as founders to build the simulated pedigrees. This set of founders were not influenced by inbreeding, relatedness, population structure, or recent change in population size. This dataset was then used as input for the software ped-sim(Campbell *et al.*, 2015; Caballero *et al.*, 2019) with a refined genetic map taken from (Bhérer, Campbell and Auton, 2017). Three different inbreeding scenarios were tested:

- First degree: siblings and parent-offspring

- Second degree: uncle-niece/aunt-nephew and grandparent-grandchild

- Third degree: first cousins and great aunt-great nephew/great uncle-great niece

Each of these scenarios was simulated 400 times using random sampled founders. ROH segments were found using PLINK with the same parameters described in the previous section (--homozyg-window-het 0 --homozyg-snp 50 --homozyg-kb 1 --homozyg-density 5000 --homozyg-gap 5000) and inbreeding coefficients estimates were also obtained using the same pipeline for both simulated and real genomes.

### 3.2.4 IBD analysis

In this work the software IBDseq vr1206 (Browning and Browning, 2013) was used to identify segments of the genome inherited by recent common ancestors (identical by descent) in European Neolithic samples. Genotype missingness and

minor allele frequency filters were applied to the imputed dataset using the software PLINK v.1.9 (--geno 0.02, --maf 0.05). Related individuals with a relatedness estimated by the software KING v.2.2.6 (Manichaikul *et al.*, 2010) higher than 4th degree relatives were also removed from analyses, resulting in 258 unrelated samples. Filtered files in PLINK format were converted to VCF using the option (--vcf) in PLINK v1.9. and used as input to the program IBDSeq with parameters (errormax=0.005 and LOD >= 3; (Schroeder *et al.*, 2019)). IBD segments shorter than 2 cM were excluded following the advice of (Browning and Browning, 2013).

To test that no systematic bias was present between types of data, the results obtained from those samples were compared where it was possible to impute genome wide calls using both WGS and SNP capture data. This used a common set of SNPs for both data types that were pruned for genotype missingness and minor allele frequency, obtaining approximately 900 thousands markers per comparison. This set of common SNPs was then used to calculate the total amount of IBD that each sample type, WGS or SNP capture, shared with the rest of the Neolithic dataset.

### 3.2.5 Population size estimates

To estimate the effective population size the IBD information obtained from IBDSeq was used as an input for the software IBDNe v.23Apr20.ae9 (Browning and Browning, 2015). This software was run for 50 generations with default settings and only for groups that shared at least 90 IBD segments longer than 2cM. An estimate of population size for each group was calculated by taking the harmonic mean over 25 generations (from 5 to 30).

Separately the effective population size of the Maltese group was estimated using the software hapROH v0.3a4 (Ringbauer, Novembre and Steinrücken, 2021). First the outlying highly inbred sample Xaghra9 was excluded. For the remaining two imputed samples (Xaghra5 and Xaghra6), diploid genotypes were downsampled to “1240K" SNPs panel and ROH were called with plink similar to what is

described above (--homozyg-window-het 0 --homozyg-snp 50 --homozyg-kb 1 --homozyg-density 5000 --homozyg-gap 5000). For each of the two Maltese samples the ROH results were then used to estimate the effective population size using the function “MLE_ROH_Ne" from the hapROH package using the parameters (min_len=4, max_len=20, ne=10000, bin_range=[0.04, 0.5], nbins=1000, error_model=False).

### 3.2.6 Chromopainter/fineSTRUCTURE

To investigate fine-scale population structure in the imputed dataset the software fineSTRUCTURE v2 was used (Lawson *et al.*, 2012). The same set of unrelated samples used in the IBDseq analysis were used for this analysis. These ancient imputed samples were filtered for genotype missingness and minor allele frequency using the software PLINK v.1.9. with parameters (--geno 0 --maf 0.01). After filtering, approximately 220K SNPs were used to phase the genotypes using the software SHAPEIT v.2.r778 (Delaneau, Marchini and Zagury, 2011). For each chromosome separately Chromopainter was run first to estimate the “Ne” and “mu” parameters using 10 expectation maximisation iteration (-i 10). These parameters were then used to paint each individual against all the others (-a 0 0). Finally “Chromocombine” was used to merge the painting information from each chromosome and obtain the normalisation parameter “c”.

The estimated matrix of chunk counts obtained from Chromocombine was then used as input to the fineSTRUCTURE algorithm. This program was run using 1,000,000 burnin and sampling iterations with sampling every 1000 iterations for the MCMC. Following the method described in (Leslie *et al.*, 2015) the state with the highest posterior probability was extracted and I performed an additional 100,000 burn-in iterations using the maximum concordance method to obtain the final tree. The information about the optimal number of groups and the cluster assignment of each sample was taken from the file “.tree” generated by the program.

### 3.2.7 Estimated effective migration surface

To visualise how geographical barriers affected migration between populations the software EEMS was used (Petkova, Novembre and Stephens, 2016). The same set of non-related ancient samples used for the IBD analyses were used to generate a pairwise dissimilarity matrix using the bed2diffs v.2. program. EEMS was initially run using 500 demes with MCMC chains parameters of 100,000 burn-in and 200,000 sampling iterations. This run was repeated 10 times using different random seeds. The run with the highest likelihood was then selected for further refinement using the same number of demes and MCMC settings of 1000,000 burn-in and 2000,000 sampling iterations.

## 3.3 Results and Discussion

### 3.3.1 Accuracy of imputation and comparison of the two pipelines

Genotype imputation is a powerful technique used to increase the genetic information present from a sampled genome. This method requires a target imputed sample to be well represented by a reference dataset. Moreover, for imputing the whole genome, a target sample must possess uniformly good coverage. Because of these two reasons most of the published work on ancient genomes has been barely leveraged using this approach as these are based on SNP capture assays that target only specific loci within the genome. With this in mind, I decided to expand the application of genotype imputation by including an extensive set of SNP capture ancient individuals from the in literature. Thanks to a collaboration with other researchers within and outside Trinity College Dublin (Lara Cassidy and Shyam Gopalakrishnan) we also coupled this analysis with a set of WGS imputed ancient individuals. To make sure that the results obtained from both imputation datasets were consistent, the genotypes imputed from the

downsampled version of the Neolithic sample LBK were compared with its high coverage (15 X) version using both its WGS and SNP captured data alternates. From Figure 3.1 it is clear that genotypes called from both SNP capture and WGS sourced imputed genomes possess a high accuracy (>95%). As expected both transitions and heterozygous genotypes were imputed with less accuracy compared to respectively transversion and homozygous variants. Interestingly, for SNP capture source imputed samples the 1240K positions are imputed with higher accuracy, undoubtedly because of higher certainty during the genotype calling process.

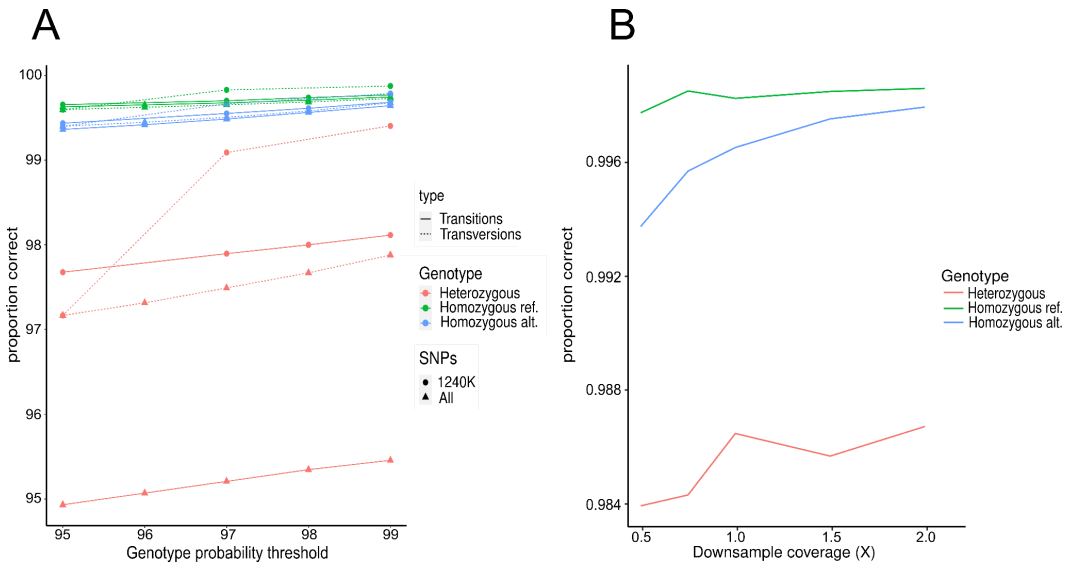


***Figure 3.1: Accuracy of imputed genotypes***. *A) Genotypes imputed from a SNP captured genome (LBK) compared against those called from its high coverage genome sequence. Heterozygous and transition SNPs show overall lower accuracy compared to homozygous and transition imputed SNPs. The SNPs corresponding to the 1240K positions were predicted overall with higher accuracy compared to other variants. B) Genotypes imputed from a downsampled WGS genome (LBK) also here compared against its high coverage genome sequence. Only transversions are considered in this analysis and heterozygous calls show overall less accuracy than homozygous ones. Overall the percentage of safely imputed genotypes remains high for each genotype class across different probability thresholds.*

To further assess whether results from both SNP capture and WGS were consistent the ROH and IBD calls for both data types were compared. For the ROH analysis inbreeding coefficients were compared for a set of 15 samples which were imputed from both WGS and SNP capture data. As shown in Figure 3.2A the regression line obtained from these results fall very close to the ideal regression line indicating a good concordance between the two types of imputed samples.

For the IBD analysis we compared the total amount of segments shared between the types of data (SNP capture or WGS) and the rest of the dataset for each of the nine Neolithic genomes imputed alternately from published WGS and SNP capture sequences. For each test the resulting points fall randomly around the perfect regression line indicating no particular bias between the type of data imputed. Moreover the correlation for the tests are all highly significant ($p < 10^{-15}$)indicating good agreement between alternately imputed data.

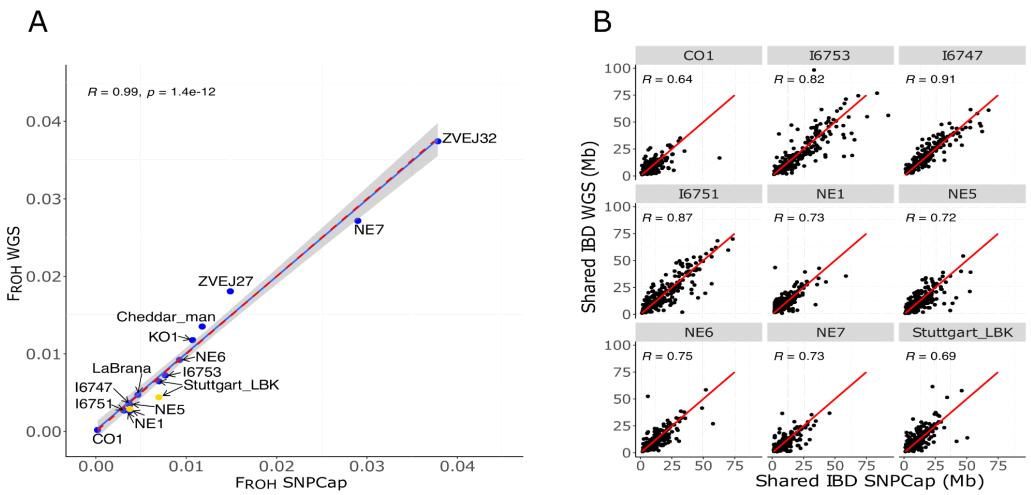


***Figure 3.2: Comparisons of IBD and $F_{ROH}$ estimates from imputed WGS and SNP captured data:*** *A) $F_{ROH}$ were compared between imputations of WGS and SNP captured data (respectively "$F_{ROH}$ WGS" and "$F_{ROH}$ SNPCap") where these are available from the same samples. These correlate with a P-value lower than 0.01 and the regression line (coloured blue) error margins overlap with the 1:1 plot (red line). Also two yellow coloured points denote where three genomes (NE1, Stuttgart_LBK) also had WGS $F_{ROH}$ estimates available from high coverage SNP calls - these are plotted Vs $F_{ROH}$ SNPCap. B) Plots of IBD sharing values involving each of nine Neolithic samples for which WGS and SNP capture - based estimates are available. For each, total estimates of the genome shared with all other Neolithic samples is plotted alternately using the two different data sets. The WGS and SNP-derived values correlate significantly (each at $p < 10^{-15}$) and vary around the 1:1 plotline, drawn in red.*

## 3.3.2 Inbreeding in the ancient Mesolithic and Neolithic Eurasia

Genome-wide diploid data allow haplotype-based assessments of population diversity – specifically, the distribution of shared ancestry within genomes, using ROH, and the distribution between individuals by identifying shared tracts that are IBD. ROH analysis shows outlying behaviour in the Maltese genomes. Xaghra9 has the second most extreme levels of long ROH (> 5cM) yet reported in prehistory; an assertion secured by its high genome coverage (Figure 3.4B) and a confirmatory analysis using a second analysis method (using ROHan (Renaud *et al.*, 2019)) which estimated 19.12% of the genome under ROH. This is only exceeded within an individual deposited in an Irish passage tomb (Newgrange10) who was the offspring of a first order consanguineous union(Cassidy *et al.*, 2020).

However, Xaghra9 has a ROH size spectrum which has less skew toward very long tracts of identity (>15 cM; Figure 3.3A).

To explore this signal, a range of consanguineous parentages were simulated and the number of ROH segments with the total fraction of the genome in these ROH ($F_{ROH}$) were plotted and compared with ancient individuals (Figure 3.3B). Unlike Newgrange10, Xaghra9 falls at the edge of the distribution seen for matings between 1st degree relatives and may result from a more complex combination of multiple inbreeding loops within his genealogy. However, this is similar to Israeli Chalcolithic sample I1178 (Harney *et al.*, 2018) ($F_{ROH}$ = 0.16; Data S1D) who was previously identified in a different analysis as a possible product of brother-sister or parent-offspring consanguinity(Ringbauer, Novembre and Steinrücken, 2021). Consequently, it is difficult to assert a precise scenario for the parentage of Xaghra9. Given the small size and relative isolation of Gozo island, it is possible that the inbreeding loops that gave rise to the Xaghra9 genome are the result of both recent genealogical inbreeding and a historically small ancestral population size. This interpretation is supported by the observation of less pronounced but relatively inflated levels of the fraction of the genome in ROH in the other two Maltese genomes (Xaghra5, Xaghra6; Figure 3.3A, 3.3B, Figure 3.4A), one of which predates Xaghra9 by ~400 years. The values for these two samples are more typical of those found in European hunter-gatherers, who maintained smaller population sizes than later farming populations (Figure 3.3B). To investigate further, levels of ROH within a range of 4 and 20 cM and a maximum likelihood framework (Ringbauer, Novembre and Steinrücken, 2021) (Fernandes *et al.*, 2021) were used to estimate effective population size, giving 515 (95% CI 397-633) individuals.

The effective population size for the Xagħra population was also calculated using the software IBDNe (Browning and Browning, 2015), which leverages patterns of IBD sharing between individuals. For comparison, this included other European Neolithic sites with more than 90 IBD segments shared between individuals in total. Xagħra, and to a lesser extent the remains from the Tomb of the Eagles on Isbister in the Orkney islands, show recent dips in population size, with the Late

Neolithic Maltese sample giving a 30 generation average of only 382 individuals (Figure 3.5A).

Thus these preserved Maltese samples show a genomic signature of an unusually small and restricted population, a signal which is distributed over a period of at least 400 years. Interestingly, the later two individuals (Xaghra5 and Xaghra9) derive from a turning point in Maltese prehistory *c.* 2450 BC, with a reducing density of radiocarbon dates (McLaughlin *et al.*, 2020) and marked worsening in diet and nutritional status (Richards *et al.*, 2001). Driving these changes seems to have been a long-term trend towards increasing aridity and thinning soils that began as early as 5500 BC (French *et al.*, 2020), implying the Late Neolithic population was less than the Early Neolithic carrying capacity estimate of two or three thousand individuals for Gozo island (67 sq Km)(French *et al.*, 2020, p. 258). This is only a small multiple of the calculated effective population size values, which are therefore not surprising. However, these estimates suggest isolation, with mating networks largely confined within the island's shores. Several strands of evidence suggest the sample is representative of the wider Neolithic community on Gozo. First, the age profile of Xagħra burials coincides closely with expectations of the mortality rates of a full early farming community, namely high infant and adolescent mortality and a relatively equal balance of adult males and females(Stoddart *et al.*, 2009). Second, the spatial analysis of the mortuary remains suggests a rich and elaborate treatment of the burials as one community (Malone and Stoddart, 2009; Malone *et al.*, 2018; Thompson *et al.*, 2020), and finally, the chosen samples are drawn from different parts of the site and span the entirety of its use.

Archaeological evidence for overseas communication with Malta in this period is mixed. Some products such as obsidian, types of chert and polished stone were definitely imported (Malone and Stoddart, 2009; Malone *et al.*, 2020). However these tend to be small, of high prestige value and have a finished state when they appear; suggesting they may not have been accompanied by a substantial volume of human traffic. Moreover, the means of cultivation of crops, raising of animals and construction were local in nature, consistent with a degree of insularity.

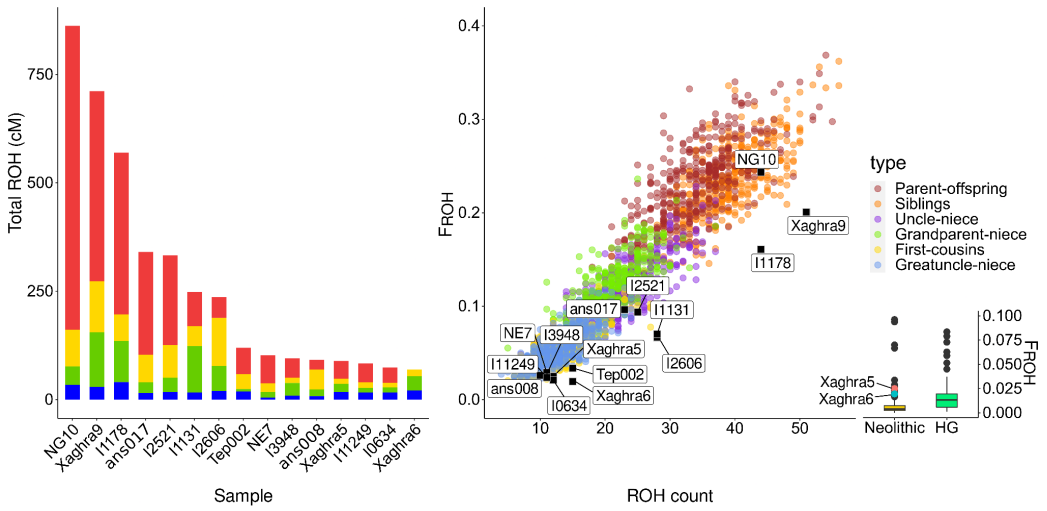


***Figure 3.3. ROH and inbreeding coefficient($F_{ROH}$) distributions among ancient Neolithic populations A:*** *Runs of homozygosity totals for Maltese samples are within the upper extreme in the Neolithic distribution. Xaghra9 particularly has a very high total, and includes long runs indicating familial inbreeding – however not as pronounced as Newgrange10 (NG10). B: Simulations of ROH spectra using specific genealogical scenarios (n= 400 for each) generate parameter distributions consistent with individuals from Gotland, Copper Age Israel and Newgrange, Ireland having resulted from recent familial inbreeding via simple pedigree loops. However, both the Xaghra9 and Israeli CopperAge (I1178) individuals have different spectra; higher contributions from short ROH indicate that they likely have multiple, complex inbreeding loops in their ancestry. The inset compares boxplots of ancient European hunter-gatherer (HG) and Neolithic $F_{ROH}$ values; the Xaghra5 and Xaghra6 genomes are more typical of the former despite having material culture of the latter.*

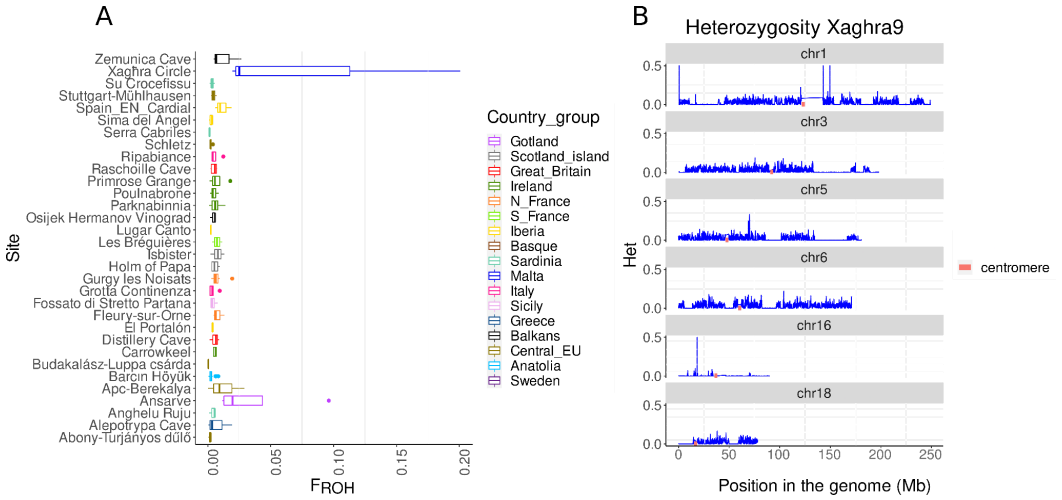


***Figure 3.4: Inbreeding $F_{ROH}$ coefficient estimates and chromosome heterozygosity plots for Xaghra9.***. *A)* $F_{ROH}$ *coefficients within sites with a minimum of 3 individuals. The Maltese possess the highest median inbreeding coefficient followed by values from Gotland. B) The average heterozygosity is plotted for Xaghra9 using a window size of 100 kb and illustrates examples of long genome tracts of homozygosity in selected chromosomes.*

### 3.3.3 Validating IBD results using Chromopainter and FineStructure analysis

Shared IBD is sensitive to recent common ancestry and, because it is a genealogical rather than frequency-based method (Thompson, 2013; Mooney *et al.*, 2018), it may be less skewed by factors such as the differences in levels of hunter-gatherer ancestry which are known among European Neolithic populations (Gamba *et al.*, 2014; Skoglund, Malmström, *et al.*, 2014; Haak *et al.*, 2015; Mathieson *et al.*, 2015; Cassidy *et al.*, 2016). Figure 3.5C shows a heatmap of the average IBD length (≥ 2cM) observed between and within European Neolithic archaeological sites with more than one imputed genome after filtering for related individuals. The highest within-site values are observed for samples from small islands, with Xagħra (Malta) producing the most extreme result, followed by Ansarve (Gotland), Holm of Papa (Orkney) and Isbister (Orkney), supporting restricted population histories for insular Neolithic societies. Figure 3.5B plots the averaged values for different geographical regions and reveals an additional trend

of higher within-group IBD sharing in the north and west of the continent, relative to the south and east.

This geographical difference also manifests in patterns of between-site sharing (Figure 3.5C), with three distinct regional clusters apparent. The Basque region, situated between the Atlantic Ocean and the western Pyrenees mountains, shows extremely inflated values between Later Neolithic sites, implying a degree of geographic isolation. Close genealogical ties are also seen across Britain and Ireland, consistent with a seaborne colonisation of the islands derived from a single or closely related founder populations. Finally, French sites cluster together, within which extreme sharing is observed between two Early Neolithic sites from Southern France, potentially reflective of the enclave colonisation process that characterised Neolithic expansion across the Mediterranean. To explore this signal further, three sites from the earliest horizon of the Spanish Neolithic (*c.* 5500-5000 BC), previously excluded given only a single sample was available from each, were considered together. Surprisingly, despite the large geographic distances between them, these three individuals show very high levels of sharing with one another and with the Mediterranean French sites, despite large differences in their hunter-gatherer ancestral contribution (Figure 2.5). This implies a population size restriction accompanied Neolithic migration into the western Mediterranean.

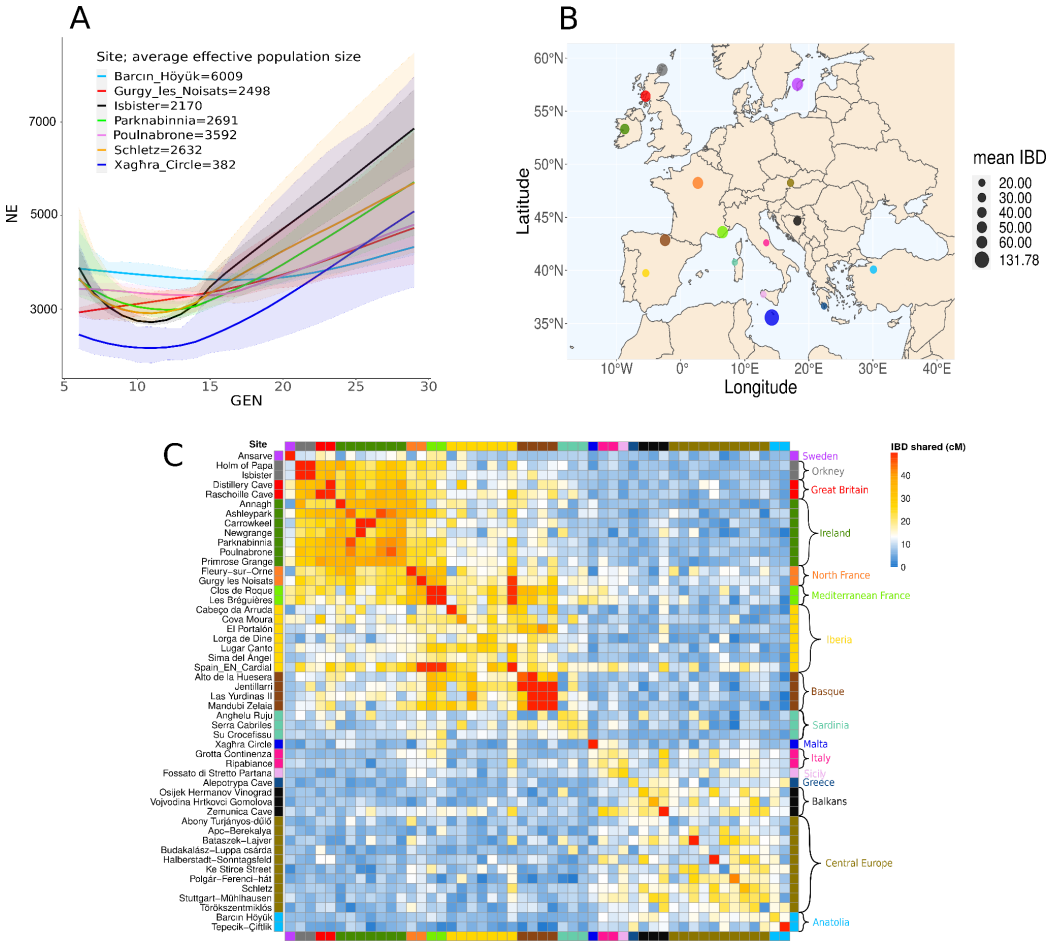


***Figure 3.5. IBD within and between sites***. *A: Population size estimated for site samples showing at least 90 IBD shared segments. The Xagħra Circle plot estimates a marked size reduction in recent ancestry and has the lowest 30 generation average effective population size of 382. B: Average IBD length in cM shared within groups defined in panel C. Malta, Gotland and Scottish islands display the highest within-site IBD average values suggesting ancestral population restriction. C: IBD sharing heatmap among those sites with two or more representatives. Note a British/Irish cluster in the top left. French individuals share some affinity with this but also cluster with Iberians and Sardinians in a large west Mediterranean group. Two island samples Xagħra (Malta) and Ansarve (Gotland) are relatively distinct and all other sites show a loose affinity in an East Mediterranean/Central European grouping.*

To explore the potential impact of maritime colonisation and continental topography on Neolithic genetic structure, a PCA was carried out on a matrix of

pairwise IBD sharing between individual imputed ancient individuals (Figure 3.7), as well as ChromoPainter and clustering using fineSTRUCTURE analysis (Lawson *et al.*, 2012) (Figure 3.6). In addition to these haplotype-based methods, an allele frequency-based approach was also applied(estimated effective migration surface, EEMS (Petkova, Novembre and Stephens, 2016)).

Results from each show a convergence on the existence of three clusters: first the Western Mediterranean including Iberian, French and Sardinian individuals; second the Eastern Mediterranean featuring Greek, Balkan and Anatolian individuals as well as Central Europeans; and third the British and Irish archipelago. These are visible as blocks in the IBD heatmap Figure 3.5C) and form three apices of variation in the PCA (Figure 3.7). They also form separate primary branches in a fineSTRUCTURE tree (Figure 3.6)(Lawson *et al.*, 2012). Intermediate samples are also intermediate in geography. For example in the PCA plot, which visibly mirrors geography (Figure 3.7), Northern French samples are placed close to Iberians but also stretch toward the British and Irish cluster. Also, the mid-Mediterranean samples from Sardinia, Malta, Sicily and Italy fall between the western and eastern poles. Neolithic populations migrated through Europe *via* two major routes, an overland transfer into central Europe and a maritime dissemination along the Mediterranean coast(Bocquet-Appel *et al.*, 2012). The most striking feature in our analyses are the contrasting outcomes of these two processes. Particularly, there is minimal distinction between central European individuals and their source populations in the Balkans and Anatolia, whereas the separation of western European individuals from those in the southeast forms the primary divide in the data.

This supports a model of agricultural expansion into central Europe from the Balkans that involved substantial numbers of migrants and strong backward communication during the dissemination of the LBK complex, with populations remaining relatively well connected throughout the Neolithic period.

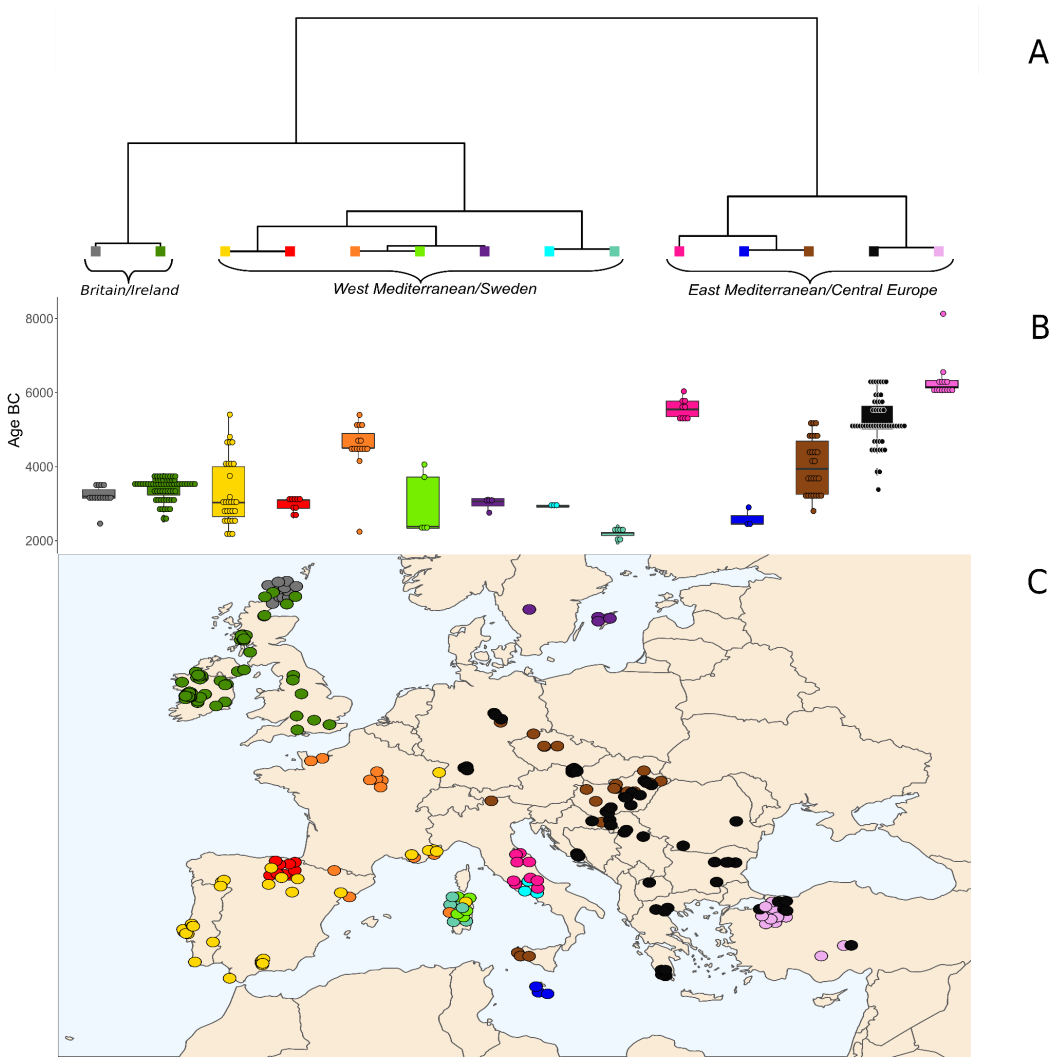


***Figure 3.6. Fine population structure analysis of European Neolithic populations.*** *A) fineSTRUCTURE tree of Neolithic European populations. From left to right three main branches define respectively Britain and Ireland, West Mediterranean and East Mediterranean as higher order population groupings. The Maltese samples emerge as a cluster and group with Italian and late Central European Neolithic groups. B) Boxplot indicating the age in years BC of each group defined by fineSTRUCTURE. Note the structuring of the East Mediterranean/Central Europe grouping by both age and geography. C) Location of samples coloured according to their groups defined by fineSTRUCTURE. A jitter of 0.6 was used to visualise points.*

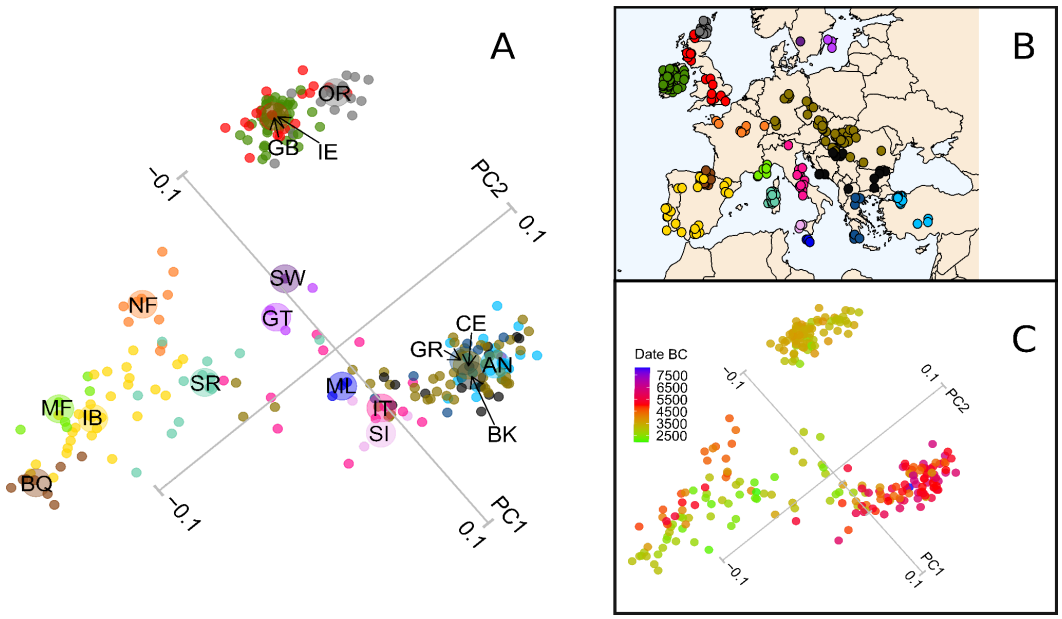


***Figure 3.7. Principal components analysis of shared IBD****. A: Principal components analysis of European Neolithic imputed ancient individuals based on total length of identity by descent segments. The variance explained by PC1 and PC2 are respectively 19% and 4.7%. Regional origins of samples are denoted by colour and two letter codes in the inset map and centroids for each group are denoted as larger circles in the plot. Three main clusters emerge: British/Irish, France/Iberia and Anatolia/Balkans/Central Europe. Island Mediterranean Maltese, Sardinian and Sicilian samples, along with Italian individuals, fall between the latter two groups, in approximate geographical sequence. Orcadian samples also distinguish from the broader British/Irish group, as do Basque sites within Iberia. AN: Anatolia, Balkans:BK, BQ: Basque, CE: Central Europe, GB: Great Britain, GR: Greece, GT: Gotland island, IB: Iberia, IE: Ireland, IT: Italy, MF: Mediterranean France, ML: Malta, NF: North France, OR: Orkney, SI: Sicily, SR: Sardinia, SW: Sweden mainland. B: Location of each sample coloured using the PCA as reference. C: Same principal component plot as figure A with samples coloured according to their estimated age in years BC.*

To explore further, EEMS was also estimated using a stepping stone model and a distance matrix computed from allele frequencies (Petkova, Novembre and Stephens, 2016); Figure 3.8 shows cold and hotspots of estimated migration rates within Neolithic Europe. The communication corridor between Anatolia, the Balkans and Central Europe is the most striking feature of this analysis and

contrasts strongly with east-west barriers both in the Mediterranean sea, the Alpine region and further north where the two Neolithic migratory streams are purported to meet (Rivollat *et al.*, 2020). In common with the other approaches, EEMS does not take account of temporal differences among samples, which would be expected to be a differentiating factor. For example the barrier between English samples and the continent might be less pronounced with the addition of more contemporaneous French genomes. However, one can assert that the major divisions are explained at least partially by geography. These correspond with those that emerge in the haplotype-informed fineSTRUCTURE analysis, where sample dates are also plotted (Figure 3.6). From this it is clear that genomes separate into different groups despite overlapping contemporaneity across the basal branches. Also there are considerable temporal differences within clusters, particularly among the samples in the Anatolian-Central European high communication corridor. The rapid Neolithic colonisation of the western Mediterranean from the east was associated with the Impressed Cardial complex, and likely took place through iterative coastal nucleations along the northern maritime littoral (Zilhão, 2014; Cunliffe, 2017). Models of this process based on archaeological data indicate that long-range voyaging is required to explain the speed of agricultural spread, which was significantly faster than that seen in Central Europe (Bocquet-Appel *et al.*, 2012; Isern *et al.*, 2017). The results accord with a limited capacity of sea craft used in this cabotage which likely restricted pioneer numbers and subsequent backward exchange. It can be inferred that the observed east-west genomic distinction derives at least partially from this foundational process as earlier individuals plot toward the extremes in the PCA graph with mid and late Neolithic individuals showing a more central tendency (Figure 3.7C). The sharp divide between eastern and western Europe echoes the analysis of French and neighbouring Neolithic genomes by Rivollat et al. (Rivollat *et al.*, 2020) who also identify that the two Neolithic streams differed in their degree of ancestral admixture with European hunter-gatherers. However, this difference in ancestry is less marked in comparisons with earlier western genomes; for example, those of the Iberian Early Neolithic.

British and Irish populations form a sister grouping to the Mediterranean Neolithic in the second fineSTRUCTURE branching (Figure 3.6) and visibly

show IBD affinity (Figure 5C), according with prior assertions that they primarily owe their origins to this southern migratory stream (Cassidy *et al.*, 2016, 2020; Olalde *et al.*, 2018; Sánchez-Quinto *et al.*, 2019). However, their maritime separation is mirrored by a degree of cluster distinction (Figure 3.5, Figure 3.6) and an estimated migration barrier (Figure 3.8). Interestingly, Irish and mainland British individuals do not separate from each other as clusters in any of these analyses, supporting shared elements of a rapid foundation process *ca* 3800 BC (Sheridan, 2010). This is an additional indication of the absence of significant batch effects as the British were imputed from SNP capture data and the Irish from shotgun sequenced libraries. However, fineSTRUCTURE confirms the emerging distinctiveness of (SNP-captured) Orcadian individuals, as well as that of Basque Late Neolithic sites (Figure 3.6), also captured in patterns of IBD-sharing (Figure 3.5C). An additional marker of separation is that Orkney islander ancient genomes have also recently been found to show unusual majority retention of male lineages across the Neolithic-Bronze Age transition (Dulias *et al.*, no date), a feature unique within Northern and Central Europe.

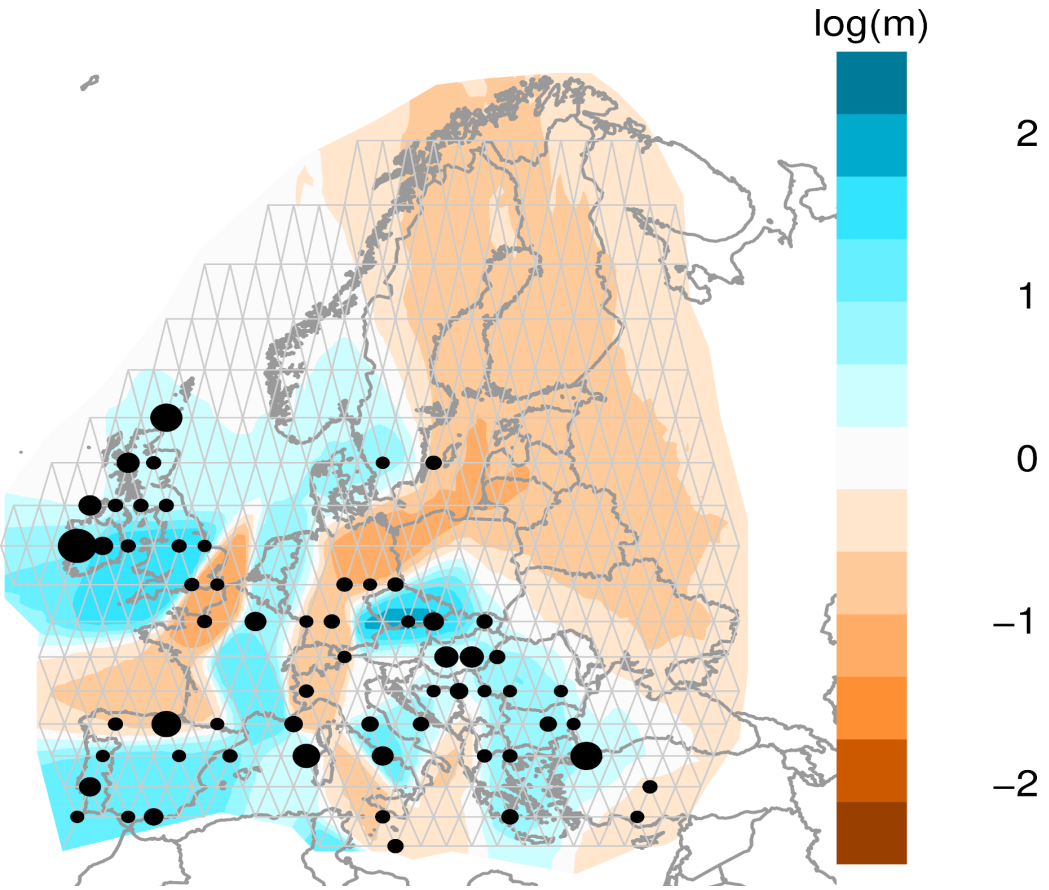


***Figure 3.8. Estimated Neolithic EEMS****(Petkova, Novembre and Stephens, 2016).* *Computed using a stepping stone model and imputed allele frequency data, migration rates are plotted as* $log_{10}$ *of the mean effective migration rate. Blue regions are surfaces over which genomic similarity is implied, orange denote barriers to genetic exchange. Dots represent the location of the samples in the constructed grid while their size indicates the number of samples. Apparent barriers separate Western and Eastern Europe and mainland Europe from Britain and Ireland.*

## 3.4 Conclusions

Since recently, ancient DNA studies were strongly limited in their findings by the low amount of genotype information that could be used. Although genotype imputation algorithms were already available, their application was mostly limited to whole genome sequenced samples from modern and ancient human

populations. In this work I have demonstrated how it is possible to impute with high accuracy the genotype of ancient human individuals with a coverage as low as 0.5X in both whole genome and SNP capture sequenced ancient samples. This achievement has allowed us to infer the demography and fine-scale genetic structure of a high number of ancient individuals covering most of the European regions. Basque, Orcadian and Irish distinctiveness emerged in pioneering studies of modern human genetic variation (Mourant, 1947; Cavalli-Sforza, Menozzi and Piazza, 1994; Bodmer, 2015; Leslie *et al.*, 2015) and genome scale investigation has compellingly recapitulated the geography of Europe in PCA, particularly its maritime features (Novembre *et al.*, 2008). It is striking that these same features emerge independently within data from an earlier genomic era in the same continent, speaking to the repeated shaping of genetic variation by the same physical topography, particularly its seascapes. One of the great debates of prehistory is the level of maritime connectivity during the course of millennia and how that connectivity interacted with marine technology and cultural response. Relationships among ancient European populations indicate that sea travel was one driver of genomic differentiation during the establishment of the Neolithic. On a wide scale, multiple analyses highlight the genetic separation between western Mediterranean sites and their source eastern Mediterranean populations. This resulted from coastal seaborne colonisation and contrasts sharply with the lack of differentiation associated with the overland establishment of Central European LBK populations from southeastern Europe and Anatolia. That maritime routes are a retardant rather than accelerant of genetic exchange is also clear from small islands. Orcadian, Gotland and Maltese genomes show signals of high ROH or within-site IBD suggesting limited populations. Particularly, effective population size estimates of only several hundred for the Late Neolithic Maltese Xagħra site suggest a population with mating networks no larger than the island of Gozo and are a powerful example of genomic insularity in prehistory.

# Supplementary Material

| ID | Country | Group_label | Site | Study | Data type | Imputed |
|---|---|---|---|---|---|---|
| RISE489 | Italy | Italy | Remedello di Sotto | (Allentoft *et al.*, 2015) | WGS | Current study |
| R10 | Italy | Italy | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R2 | Italy | Italy | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R3 | Italy | Italy | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R4 | Italy | Italy | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R5 | Italy | Italy | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R8 | Italy | Italy | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R9 | Italy | Italy | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R16 | Italy | Italy | Ripabiance | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R17 | Italy | Italy | Ripabiance | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |

| | | | | | | |
|---|---|---|---|---|---|---|
| R18 | Italy | Italy | Ripabiance | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R19 | Italy | Italy | Ripabiance | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R1014 | Italy | Italy | San Biagio | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R22 | Sardinia | Sardinia | Su Crocefissu | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R25 | Sardinia | Sardinia | Su Crocefissu | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R26 | Sardinia | Sardinia | Su Crocefissu | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R27 | Sardinia | Sardinia | Su Crocefissu | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R28 | Sardinia | Sardinia | Su Crocefissu | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R29 | Sardinia | Sardinia | Su Crocefissu | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R11 | Italy | Italy-HG | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R15 | Italy | Italy-HG | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |
| R7 | Italy | Italy-HG | Grotta Continenza | (Antonio *et al.*, 2019) | WGS | (Antonio *et al.*, 2019) |

| | | | | | | |
|---|---|---|---|---|---|---|
| I6757 | Great_Britain | Great_Britain | Attermire Scar | (Brace *et al.*, 2019) | SNP capture | Current study |
| I6760 | Great_Britain | Great_Britain | BurnGround | (Brace *et al.*, 2019) | SNP capture | Current study |
| I6747 | Great_Britain | Great_Britain | Carsington Pasture | (Brace *et al.*, 2019) | WGS | Current study |
| I6747 | Great_Britain | Great_Britain | Carsington Pasture | (Brace *et al.*, 2019) | SNP capture | Current study |
| I6753 | Great_Britain | Great_Britain | Coldrum | (Brace *et al.*, 2019) | WGS | Current study |
| I6753 | Great_Britain | Great_Britain | Coldrum | (Brace *et al.*, 2019) | SNP capture | Current study |
| I6766 | Scotland | Great_Britain | Sutherland_Embo | (Brace *et al.*, 2019) | SNP capture | Current study |
| Cheddar_man-I6767 | Great Britain | Great Britain-HG | Gough Cave | (Brace *et al.*, 2019) | WGS | Current study |
| Cheddar_man-I6767 | Great Britain | Great Britain-HG | Gough Cave | (Brace *et al.*, 2019) | SNP capture | Current study |
| BA64 | Ireland | Ireland | Ballynahatty | (Cassidy *et al.*, 2016) | WGS | Current study |
| ANN1 | Ireland | Ireland | Annagh | (Cassidy *et al.*, 2020) | WGS | Current study |
| ANN2 | Ireland | Ireland | Annagh | (Cassidy *et al.*, 2020) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| ARD2 | Ireland | Ireland | Ardcrony | (Cassidy *et al.*, 2020) | WGS | Current study |
| ASH1 | Ireland | Ireland | Ashleypark | (Cassidy *et al.*, 2020) | WGS | Current study |
| ASH3 | Ireland | Ireland | Ashleypark | (Cassidy *et al.*, 2020) | WGS | Current study |
| BG72 | Ireland | Ireland | Baunogenasraid | (Cassidy *et al.*, 2020) | WGS | Current study |
| CAK530 | Ireland | Ireland | Carrowkeel | (Cassidy *et al.*, 2020) | WGS | Current study |
| CAK531 | Ireland | Ireland | Carrowkeel | (Cassidy *et al.*, 2020) | WGS | Current study |
| CAK532 | Ireland | Ireland | Carrowkeel | (Cassidy *et al.*, 2020) | WGS | Current study |
| CAK533 | Ireland | Ireland | Carrowkeel | (Cassidy *et al.*, 2020) | WGS | Current study |
| CAK68 | Ireland | Ireland | Carrowkeel | (Cassidy *et al.*, 2020) | WGS | Current study |
| CH448 | Ireland | Ireland | Cohaw | (Cassidy *et al.*, 2020) | WGS | Current study |
| GNM1007 | Ireland | Ireland | Glennamong | (Cassidy *et al.*, 2020) | WGS | Current study |
| GNM1076 | Ireland | Ireland | Glennamong | (Cassidy *et al.*, 2020) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| JP14 | Ireland | Ireland | Jerpoint West | (Cassidy *et al.*, 2020) | WGS | Current study |
| MB6 | Ireland | Ireland | Millin Bay (Keentagh Td.) | (Cassidy *et al.*, 2020) | WGS | Current study |
| NG10 | Ireland | Ireland | Newgrange | (Cassidy *et al.*, 2020) | WGS | Current study |
| NGZ1 | Ireland | Ireland | Newgrange | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB1327 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB1794 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB186 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB2031 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB357 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB443 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB581 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB672 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| PB675 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB754 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PB768 | Ireland | Ireland | Parknabinnia | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN02 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN03 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN04 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN05 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN06 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN07 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN10_PN113 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN107 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN112 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| PN12 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN13 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| PN16 | Ireland | Ireland | Poulnabrone | (Cassidy *et al.*, 2020) | WGS | Current study |
| KGH6 | Ireland | Ireland-HG | Killuragh | (Cassidy *et al.*, 2020) | WGS | Current study |
| SRA62 | Ireland | Ireland-HG | Sramore | (Cassidy *et al.*, 2020) | WGS | Current study |
| I15941 | Sardinia | Sardinia | Anghelu Ruju | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I15942 | Sardinia | Sardinia | Anghelu Ruju | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I15946 | Sardinia | Sardinia | Anghelu Ruju | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I4062 | Sicily | Sicily | Fossato di Stretto Partana | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I4063 | Sicily | Sicily | Fossato di Stretto Partana | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I4064 | Sicily | Sicily | Fossato di Stretto Partana | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I4065 | Sicily | Sicily | Fossato di Stretto Partana | (Fernandes *et al.*, 2020) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I16165 | Sardinia | Sardinia | Sa Ucca de su Tintirriolu | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I14675 | Sardinia | Sardinia | Serra Cabriles | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I14676 | Sardinia | Sardinia | Serra Cabriles | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I14677 | Sardinia | Sardinia | Serra Cabriles | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| I14678 | Sardinia | Sardinia | Serra Cabriles | (Fernandes *et al.*, 2020) | SNP capture | Current study |
| Vestonice16 | Czech Republic | Czech-PA | Dolni Vestonice | (Fu *et al.*, 2016) | SNP capture | Current study |
| ElMiron | Spain | Spain-HG | El Miron | (Fu *et al.*, 2016) | SNP capture | Current study |
| Villabruna | Italy | Italy-PA | Villabruna | (Fu *et al.*, 2016) | SNP capture | Current study |
| CO1-I1497 | Hungary | Central_EU | Apc-Berekalya | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| NE6-I1496 | Hungary | Central_EU | Apc-Berekalya | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| NE7-I1495 | Hungary | Central_EU | Apc-Berekalya | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| CO1-I1497 | Hungary | Central_EU | Apc-Berekalya | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| NE6-I1496 | Hungary | Central_EU | Apc-Berekalya | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | WGS | Current study |
| NE7-I1495 | Hungary | Central_EU | Apc-Berekalya | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | WGS | Current study |
| I1498 | Hungary | Central_EU | Debrecen Tócópart Erdoalja | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1499 | Hungary | Central_EU | Garadna | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| NE5-I1500 | Hungary | Central_EU | Kompolt-Kigyoser | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| NE5-I1500 | Hungary | Central_EU | Kompolt-Kigyoser | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | WGS | Current study |
| I1505 | Hungary | Central_EU | Polgár-Ferenci-hát | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| NE1-I1506 | Hungary | Central_EU | Polgár-Ferenci-hát | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| NE1-I1506 | Hungary | Central_EU | Polgár-Ferenci-hát | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | WGS | Current study |
| KO1-I1507 | Hungary | Hungary-HG | Tiszaszőlős-Domaháza | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| KO1-I1507 | Hungary | Hungary-HG | Tiszaszőlős-Domaháza | (Gamba *et al.*, 2014);(Mathieson *et al.*, 2015) | WGS | Current study |
| Canes1_Meso | Spain | Spain-HG | Canes | (González-Fortes *et al.*, 2017) | WGS | Current study |
| Chan_Meso | Spain | Spain-HG | Chan do Lindeiro | (González-Fortes *et al.*, 2017) | WGS | Current study |
| OC1_Meso | Romania | Romania-HG | Ostrovul Corbului | (González-Fortes *et al.*, 2017) | WGS | Current study |
| SC1_Meso | Romania | Romania-HG | Schela Cladovei | (González-Fortes *et al.*, 2017) | WGS | Current study |
| SC2_Meso | Romania | Romania-HG | Schela Cladovei | (González-Fortes *et al.*, 2017) | WGS | Current study |
| LU339 | Portugal | Iberia | Lorga de Dine | (González-Fortes *et al.*, 2019) | WGS | Current study |
| LD1174 | Portugal | Iberia | Lorga de Dine | (González-Fortes *et al.*, 2019) | WGS | Current study |
| LD270 | Portugal | Iberia | Lorga de Dine | (González-Fortes *et al.*, 2019) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| atp002 | Spain | Iberia | El Portalón | (Günther *et al.*, 2015);(Valdiosera *et al.*, 2018) | WGS | Current study |
| atp016 | Spain | Iberia | El Portalón | (Günther *et al.*, 2015);(Valdiosera *et al.*, 2018) | WGS | Current study |
| atp12-1420 | Spain | Iberia | El Portalón | (Günther *et al.*, 2015);(Valdiosera *et al.*, 2018) | WGS | Current study |
| Hum2 | Norway | Norway-HG | Hummervikholmen | (Günther *et al.*, 2018) | WGS | Current study |
| Hum1 | Norway | Norway-HG | Hummervikholmen | (Günther *et al.*, 2018) | WGS | Current study |
| Steigen | Norway | Norway-HG | Steigen | (Günther *et al.*, 2018) | WGS | Current study |
| SBj | Sweden | Sweden-HG | Stora Bjers | (Günther *et al.*, 2018) | WGS | Current study |
| sf12 | Sweden | Sweden-HG | Stora Förvar | (Günther *et al.*, 2018) | WGS | Current study |
| sf9 | Sweden | Sweden-HG | Stora Förvar | (Günther *et al.*, 2018) | WGS | Current study |
| I1178 | Israel | Israel | Peki'in | (Harney *et al.*, 2018) | SNP capture | Current study |
| Bar31 | Turkey | Anatolia | Barcın Höyük | (Hofmanová *et al.*, 2016) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| Bar8 | Turkey | Anatolia | Barcın Höyük | (Hofmanová *et al.*, 2016) | WGS | Current study |
| Klei10 | Greece | Greece | Kleitos | (Hofmanová *et al.*, 2016) | WGS | Current study |
| Pal7 | Greece | Greece | Paliambela | (Hofmanová *et al.*, 2016) | WGS | Current study |
| Rev5 | Greece | Greece | Revenia | (Hofmanová *et al.*, 2016) | WGS | Current study |
| Bichon | Switzerland | Switzerland-PA | Grotte du Bichon | (Jones *et al.*, 2015) | WGS | Current study |
| ZVEJ25 | Latvia | Latvia-HG | Zvejnieki | (Jones *et al.*, 2017) | WGS | Current study |
| ZVEJ31 | Latvia | Latvia-HG | Zvejnieki | (Jones *et al.*, 2017) | WGS | Current study |
| I1819 | Ukraine | Ukraine-HG | Vasil'evka | (Jones *et al.*, 2017);(Mathieson *et al.*, 2018) | SNP capture | Current study |
| ZVEJ27-I4628 | Latvia | Latvia-HG | Zvejnieki | (Jones *et al.*, 2017);(Mathieson *et al.*, 2018) | WGS | Current study |
| ZVEJ32-I4632 | Latvia | Latvia-HG | Zvejnieki | (Jones *et al.*, 2017);(Mathieson *et al.*, 2018) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| ZVEJ27-I4628 | Latvia | Latvia-HG | Zvejnieki | (Jones *et al.*, 2017);(Mathieson *et al.*, 2018) | SNP capture | Current study |
| ZVEJ32-I4632 | Latvia | Latvia-HG | Zvejnieki | (Jones *et al.*, 2017);(Mathieson *et al.*, 2018) | SNP capture | Current study |
| Ötzi | Italy | Italy | Tisenjoch | (Keller *et al.*, 2012) | WGS | Current study |
| Bon002 | Turkey | Anatolia | Boncuklu | (Kılınç *et al.*, 2016) | WGS | Current study |
| Tep002 | Turkey | Anatolia | Tepecik-Çiftlik | (Kılınç *et al.*, 2016) | WGS | Current study |
| Tep004 | Turkey | Anatolia | Tepecik-Çiftlik | (Kılınç *et al.*, 2016) | WGS | Current study |
| Loschbour | Luxembourg | Luxembourg-HG | Echternach | (Lazaridis *et al.*, 2014) | WGS | Current study |
| Motala12 | Sweden | Sweden-HG | Kanaljorden | (Lazaridis *et al.*, 2014) | WGS | Current study |
| Stuttgart_LBK-I0018 | Germany | Central_EU | Stuttgart-Mühlhausen | (Lazaridis *et al.*, 2014);(Lipson *et al.*, 2017) | SNP capture | Current study |
| Stuttgart_LBK-I0018 | Germany | Central_EU | Stuttgart-Mühlhausen | (Lazaridis *et al.*, 2014);(Lipson *et al.*, 2017) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I0172 | Germany | Central_EU | Esperstedt | (Lazaridis *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0406 | Spain | Iberia | La Mina | (Lazaridis *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0412 | Spain | Iberia | La Mina | (Lazaridis *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0025 | Germany | Central_EU | Stuttgart-Mühlhausen | (Lazaridis *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0026 | Germany | Central_EU | Stuttgart-Mühlhausen | (Lazaridis *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0054 | Germany | Central_EU | Unterwiederstedt | (Lazaridis *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0012 | Sweden | Sweden-HG | Kanaljorden | (Lazaridis *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0014 | Sweden | Sweden-HG | Kanaljorden | (Lazaridis *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I2788 | Hungary | Central_EU | Abony-Turjányos dűlő | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2790 | Hungary | Central_EU | Abony-Turjányos dűlő | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2791 | Hungary | Central_EU | Abony-Turjányos dűlő | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2370 | Hungary | Central_EU | Alsonemedi | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I4189 | Hungary | Central_EU | Alsonyek | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2753 | Hungary | Central_EU | Balatonlelle Fels-Gamász | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I1904 | Hungary | Central_EU | Bataszek-Lajver | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2739 | Hungary | Central_EU | Bataszek-Lajver | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2366 | Hungary | Central_EU | Budakalász-Lupa csárda | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2367 | Hungary | Central_EU | Budakalász-Lupa csárda | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2369 | Hungary | Central_EU | Budakalász-Lupa csárda | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I5838 | Spain | Iberia | El Mirador | (Lipson *et al.*, 2017) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I2384 | Hungary | Central_EU | Hajdúnánás-Eszlári út | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I0449 | Hungary | Central_EU | Hódmezővásárhely-Gorzsa | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I1880 | Hungary | Central_EU | Lánycsók Gata-Csatola | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I1838 | Spain | Basque | Las Yurdinas II | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I3269 | Spain | Basque | Las Yurdinas II | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I1907 | Hungary | Central_EU | Proletár dűlő | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I1895 | Hungary | Central_EU | Szederkény-Kukorica-dűlő | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2793 | Hungary | Central_EU | Törökszentmiklós | (Lipson *et al.*, 2017) | SNP capture | Current study |
| I2794 | Hungary | Central_EU | Törökszentmiklós | (Lipson *et al.*, 2017) | SNP capture | Current study |
| CabecoArruda117B | Portugal | Iberia | Cabeço da Arruda | (Martiniano *et al.*, 2017) | WGS | Current study |
| CabecoArruda122A | Portugal | Iberia | Cabeço da Arruda | (Martiniano *et al.*, 2017) | WGS | Current study |
| CovaMoura364 | Portugal | Iberia | Cova Moura | (Martiniano *et al.*, 2017) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| CovaMoura9B | Portugal | Iberia | Cova Moura | (Martiniano *et al.*, 2017) | WGS | Current study |
| LugarCanto41 | Portugal | Iberia | Lugar Canto | (Martiniano *et al.*, 2017) | WGS | Current study |
| LugarCanto42 | Portugal | Iberia | Lugar Canto | (Martiniano *et al.*, 2017) | WGS | Current study |
| LugarCanto44 | Portugal | Iberia | Lugar Canto | (Martiniano *et al.*, 2017) | WGS | Current study |
| I0707 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0708 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0709 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0744 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0745 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0746 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I0854 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1096 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I1097 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1098 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1101 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1579 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1580 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1581 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1583 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1585 | Turkey | Anatolia | Barcın Höyük | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| I1300 | Spain | Iberia | El Mirador | (Mathieson *et al.*, 2015) | SNP capture | Current study |
| Karelia_HG-I00061 | Russia | EHG | Yuzhnyy Oleni Ostrov | (Mathieson *et al.*, 2015);(Fu *et al.*, 2016) | WGS | Current study |
| Karelia_HG-I00061 | Russia | EHG | Yuzhnyy Oleni Ostrov | (Mathieson *et al.*, 2015);(Fu *et al.*, 2016) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I0046 | Germany | Central_EU | Halberstadt-Sontagsfeld | (Mathieson *et al.*, 2015);(Lipson *et al.*, 2017) | SNP capture | Current study |
| I0100 | Germany | Central_EU | Halberstadt-Sontagsfeld | (Mathieson *et al.*, 2015);(Lipson *et al.*, 2017) | SNP capture | Current study |
| I3708 | Greece | Greece | Alepotrypa Cave | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I3709 | Greece | Greece | Alepotrypa Cave | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I3920 | Greece | Greece | Alepotrypa Cave | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5427 | Greece | Greece | Alepotrypa Cave | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I2533 | Romania | Central_EU | Carcea | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I2431 | Bulgaria | Balkans | Ivanovo | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5077 | Croatia | Balkans | Osijek Hermanov Vinograd | (Mathieson *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I5078 | Croatia | Balkans | Osijek Hermanov Vinograd | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5079 | Croatia | Balkans | Osijek Hermanov Vinograd | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I2532 | Romania | Central_EU | Radovanci | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4918 | Serbia | Central_EU | Saraorci-Jezava | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5069 | Austria | Central_EU | Schletz | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5070 | Austria | Central_EU | Schletz | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5204 | Austria | Central_EU | Schletz | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5205 | Austria | Central_EU | Schletz | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5206 | Austria | Central_EU | Schletz | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5207 | Austria | Central_EU | Schletz | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5208 | Austria | Central_EU | Schletz | (Mathieson *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I0676 | Macedonia | Balkans | Skopje Sopite Govrlevo | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I2424 | Bulgaria | Balkans | Smyadovo | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I2427 | Bulgaria | Balkans | Sushina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4089 | Romania | Central_EU | Urziceni | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I2521 | Bulgaria | Balkans | Veliko Tarnovo Dzhulyunitsa | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I0634 | Serbia | Balkans | Vojvodina Hrtkovci Gomolova | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I1131 | Serbia | Balkans | Vojvodina Hrtkovci Gomolova | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I0698 | Bulgaria | Balkans | Yabalkovo Dimitrovgrad Haskovo | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I3433 | Croatia | Balkans | Zemunica Cave | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I3947 | Croatia | Balkans | Zemunica Cave | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I3948 | Croatia | Balkans | Zemunica Cave | (Mathieson *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I4111 | Ukraine | Ukraine-HG | Dereivka I | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4914 | Serbia | HG-IronGates | Hajdučka Vodenica | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4915 | Serbia | HG-IronGates | Hajdučka Vodenica | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4916 | Serbia | HG-IronGates | Hajdučka Vodenica | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4917 | Serbia | HG-IronGates | Hajdučka Vodenica | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5402 | Serbia | HG-IronGates | Hajdučka Vodenica | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5407 | Serbia | HG-IronGates | Lepenski Vir | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4582 | Romania | HG-IronGates | Ostrovul Corbului | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5233 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5234 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5235 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5236 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I5237 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5238 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5239 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5240 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5242 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5244 | Serbia | HG-IronGates | Padina | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5411 | Romania | HG-IronGates | Schela Cladovei | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I5436 | Romania | HG-IronGates | Schela Cladovei | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I1734 | Ukraine | Ukraine-HG | Vasil'evka | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I1736 | Ukraine | Ukraine-HG | Vasil'evka | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I1763 | Ukraine | Ukraine-HG | Vasil'evka | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4873 | Serbia | HG-IronGates | Vlasac | (Mathieson *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I4874 | Serbia | HG-IronGates | Vlasac | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4875 | Serbia | HG-IronGates | Vlasac | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4876 | Serbia | HG-IronGates | Vlasac | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4877 | Serbia | HG-IronGates | Vlasac | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4878 | Serbia | HG-IronGates | Vlasac | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4880 | Serbia | HG-IronGates | Vlasac | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4881 | Serbia | HG-IronGates | Vlasac | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I1732 | Ukraine | Ukraine-HG | Vovnigi | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I1738 | Ukraine | Ukraine-HG | Vovnigi | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4432 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4434 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4435 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I4436 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4437 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4438 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4439 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4440 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4441 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4550 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4551 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4552 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4554 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4595 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4596 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I4627 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I4630 | Latvia | Latvia-HG | Zvejnieki | (Mathieson *et al.*, 2018) | SNP capture | Current study |
| I6677 | Czech Republic | Central_EU | Bilina | (Narasimhan *et al.*, 2019) | SNP capture | Current study |
| LaBrana1-I0585 | Spain | Spain-HG | La Braña-Arintero | (Olalde *et al.*, 2014);(Mathieson *et al.*, 2015) | SNP capture | Current study |
| LaBrana1-I0585 | Spain | Spain-HG | La Braña-Arintero | (Olalde *et al.*, 2014);(Mathieson *et al.*, 2015) | WGS | Current study |
| CB13 | Spain | Iberia/Spain_EN_Cardial | Cova Bonica | (Olalde *et al.*, 2015) | WGS | Current study |
| I2988 | Scotland | Great_Britain | Clachaig | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I4304 | France | S_France | Clos de Roque | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I4305 | France | S_France | Clos de Roque | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2659 | Scotland | Great_Britain | Distillery Cave | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2660 | Scotland | Great_Britain | Distillery Cave | (Olalde *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I2691 | Scotland | Great_Britain | Distillery Cave | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2606 | Great_Britain | Great_Britain | Eton Rowing Course | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I6759 | Great_Britain | Great_Britain | Giggleswick Scar | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2636 | Scotland | Scotland_island | Holm of Papa | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2637 | Scotland | Scotland_island | Holm of Papa | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2650 | Scotland | Scotland_island | Holm of Papa | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2651 | Scotland | Scotland_island | Holm of Papa | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2630 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2631 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2932 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2933 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2934 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I2935 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2977 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2978 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2979 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I3085 | Scotland | Scotland_island | Isbister | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I4893 | Czech Republic | Central_EU | Ke Stírce Street | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I4894 | Czech Republic | Central_EU | Ke Stírce Street | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2980 | Scotland | Scotland_island | Point of Cott | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I3041 | Scotland | Great_Britain | Raschoille Cave | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I3133 | Scotland | Great_Britain | Raschoille Cave | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I3134 | Scotland | Great_Britain | Raschoille Cave | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I3136 | Scotland | Great_Britain | Raschoille Cave | (Olalde *et al.*, 2018) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I3138 | Scotland | Great_Britain | Raschoille Cave | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I2635 | Great Britain | Scotland_island | Tulloch of Assery | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I7554 | Scotland | Scotland_island | Unstan Chamber | (Olalde *et al.*, 2018) | SNP capture | Current study |
| I6751 | Great_Britain | Great_Britain | Fussells Lodge | (Olalde *et al.*, 2018);(Brace *et al.*, 2019) | WGS | Current study |
| I6751 | Great_Britain | Great_Britain | Fussells Lodge | (Olalde *et al.*, 2018);(Brace *et al.*, 2019) | SNP capture | Current study |
| I1978 | Spain | Basque | Alto de la Huesera | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I8134 | Spain | Iberia | Campo de Hockey | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I4565 | Spain | Iberia | Galls Carboners | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I11248 | Spain | Basque | Jentillarri | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I11249 | Spain | Basque | Jentillarri | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I1846 | Spain | Basque | Las Yurdinas II | (Olalde *et al.*, 2019) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| I7604 | Spain | Basque | Mandubi Zelaia | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I7606 | Spain | Basque | Mandubi Zelaia | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I5076 | Portugal | Iberia | Monte Canelas | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I5429 | Portugal | Iberia | Perdigões | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I8197 | Spain | Iberia | Sima del Ángel | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I8198 | Spain | Iberia | Sima del Ángel | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I8199 | Spain | Iberia | Sima del Ángel | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I8364 | Spain | Iberia | Sima del Ángel | (Olalde *et al.*, 2019) | SNP capture | Current study |
| I8365 | Spain | Iberia | Sima del Ángel | (Olalde *et al.*, 2019) | SNP capture | Current study |
| FLR003 | France | N_France | Fleury-sur-Orne | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| FLR004 | France | N_France | Fleury-sur-Orne | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| FLR007 | France | N_France | Fleury-sur-Orne | (Rivollat *et al.*, 2020) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| GRG003 | France | N_France | Gurgy les Noisats | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| GRG008 | France | N_France | Gurgy les Noisats | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| GRG015 | France | N_France | Gurgy les Noisats | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| GRG019 | France | N_France | Gurgy les Noisats | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| GRG022 | France | N_France | Gurgy les Noisats | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| GRG025 | France | N_France | Gurgy les Noisats | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| GRG027 | France | N_France | Gurgy les Noisats | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| GRG032 | France | N_France | Gurgy les Noisats | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| GRG043 | France | N_France | Gurgy les Noisats | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| LBR001 | France | S_France | Les Bréguières | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| LBR002 | France | S_France | Les Bréguières | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| LBR003 | France | S_France | Les Bréguières | (Rivollat *et al.*, 2020) | SNP capture | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| OBN008 | France | N_France | Obernai | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| OBN009 | France | N_France | Obernai | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| PEN003 | France | S_France | Pendimoun | (Rivollat *et al.*, 2020) | SNP capture | Current study |
| ans008 | Gotland | Gotland | Ansarve | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |
| ans014 | Gotland | Gotland | Ansarve | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |
| ans017 | Gotland | Gotland | Ansarve | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |
| ans016 | Gotland | Gotland | Ansarve | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |
| bal004 | Scotland | Great_Britain | Balintore | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |
| kol006 | Czech Republic | Central_EU | Kolin | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |
| Primrose13 | Ireland | Ireland | Primrose Grange | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |
| Primrose16 | Ireland | Ireland | Primrose Grange | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |
| Primrose2 | Ireland | Ireland | Primrose Grange | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| Primrose9 | Ireland | Ireland | Primrose Grange | (Sánchez-Quinto *et al.*, 2019) | WGS | Current study |
| 611 | Great_Britain | Great_Britain | Trumpington Meadows | (Scheib *et al.*, 2019) | WGS | Current study |
| 613 | Great_Britain | Great_Britain | Trumpington Meadows | (Scheib *et al.*, 2019) | WGS | Current study |
| Kostenki14 | Russia | Russia-PA | Kostenki | (Seguin-Orlando *et al.*, 2014) | WGS | Current study |
| Sunghir3 | Russia | Russia-PA | Sunghir | (Sikora *et al.*, 2017) | WGS | Current study |
| Gokhem2 | Sweden | Sweden | Gökhem Parish | (Skoglund, Malmström, *et al.*, 2014) | WGS | Current study |
| Ajv58 | Sweden | Sweden-HG | Gotland | (Skoglund, Malmström, *et al.*, 2014) | WGS | Current study |
| Xaghra5 | Malta | Malta | Xagħra Circle | (Ariano *et al.*, 2022) | WGS | Current study |
| Xaghra6 | Malta | Malta | Xaghra Circle | (Ariano *et al.*, 2022) | WGS | Current study |
| Xaghra9 | Malta | Malta | Xaghra Circle | (Ariano *et al.*, 2022) | WGS | Current study |

| | | | | | | |
|---|---|---|---|---|---|---|
| por002 | Spain | Iberia | El Portalón | (Valdiosera *et al.*, 2018) | WGS | Current study |
| mur | Spain | Iberia/Spain_EN_Cardial | Murciélagos de Zuheros | (Valdiosera *et al.*, 2018) | WGS | Current study |

***Table 3.S1:*** *Description of the samples imputed for this project.*

# 4 Phenotypic traits in ancient populations

## 4.1 Introduction

### 4.1.1 Osteological studies of human stature

The study of body shape in ancient humans has been carried out mostly using osteological methods. By making use of available osteological data, several works have demonstrated that it is possible to estimate the height of an individual using skeletal elements as a proxy (Raxter, Auerbach and Ruff, 2006; Vercellotti *et al.*, 2009; Ruff *et al.*, 2012; Ruff, 2018). Depending on the type of bone used, two main methods have been developed for stature determination. The first one uses the sum of the parts of a fully articulated skeleton and was developed by Fully in 1973 (Fully, 1956). The main drawback with this method is that it is only very occasionally that complete inhumations are discovered in archaeology and it is more common that the remains of an individual are fragmented and dispersed. Motivated by this limitation, alternate approaches were developed to allow the estimation of stature using only measurement of a selected long bone such as a femur, humerus or tibia. These methods work first by estimating the parameters of a regression equation using a training dataset of known stature. Once the parameters are known they can be applied on a test dataset of individuals. Given that these methods are based on regression equations it is important that training and test datasets match in terms of the type of data analysed. For this reason many different regression formulas have been created depending on factors such as the sex or the place of origin of an individual (Ruff, 2018). While these formulae show good accuracy and have been extensively applied to the analysis of ancient human populations, they are also limited by the state of the sample recovered. For example, if the bone measured is not perfectly preserved or if it shows the presence of pathologies that might have affected the stature of an individual (such as osteoporosis) this can introduce extra error or bias.

### 4.1.2 GWAS to study phenotypes

A productive approach in recent years in genetics has been the search using genome wide association study (GWAS) for a set of common variants that together contribute to the expression of a particular phenotype. GWAS studies typically start by considering a large cohort of individuals, not related or divided by genetic structure, in which part of the population is phenotyped for the trait under investigation. Once this large group is divided or quantitatively measured based on this trait, statistical analyses are then performed to understand which SNPs contribute. Each of these variants is then associated with the trait of interest through a significance value (p-value). An odds ratio value (or beta score in some cases) is also estimated for each SNP to indicate how much this variant affects the trait of interest. When both these are known they can be used to estimate the SNP heritability of a trait, defined as the variance of the phenotype explained by the sets of variants discovered. For example, a recent study found that 25% of the variance in the stature of an individual can be explained by 3290 specific genotype markers (Yengo *et al.*, 2018). However, despite the promise that GWAS studies hold, they also have been criticised for the presence of confounding factors. One of the most common is the presence of uncorrected genetic structure that can lead to wrongly assuming some variants to be associated with a trait ((Sul, Martin and Eskin, 2018)). For example, a recent work highlighted the problem of replicating some experiments when passing from one GWAS dataset (GIANT) to another (UKBioBank). The reason was the presence of hidden genetic structure within the first dataset that misled the results of some analyses (Sohail *et al.*, 2019). Another common issue that affects the application of GWAS is the transferability of the information to target populations of different ancestries. This phenomenon is further exacerbated by the fact that most GWAS studies have been carried out using mainly individuals of European ancestry, thus limiting their applicability for target populations outside this area (Sirugo, Williams and Tishkoff, 2019). One interesting solution applied in recent work consisted of weighting each variant

present in a target individual by its amount of ancestry shared with the discovery GWAS dataset (Marnetto *et al.*, 2020).

### 4.1.3 The application of GWAS information to ancient humans

While GWAS-informed trait prediction has been extensively applied to present day groups, this can not be said for ancient populations. Indeed it was only in 2015 that common traits were initially thus investigated in ancient individuals. In this a group of researchers from Harvard investigated the trend in, and selection for, stature across different periods. By doing so they discovered that populations with a high level of Steppe ancestry also showed evidence of increased estimated genetic height. The opposite was true for those with high Neolithic farmer ancestry. However, when this analysis was extended to body mass index there was no significant difference between time periods (Mathieson *et al.*, 2015). Similar results were reported two years later using a larger set of individuals (Martiniano *et al.*, 2017) and subsequently a similar work investigated the stature of more than 1000 ancient individuals of Western Eurasian origin (using a limit of 100° E longitude)(Cox *et al.*, 2019). This last analysis was of particular interest as for the first time it coupled in individuals both ancient DNA and osteological information. To do so the authors first collected height information from a published database (Ruff, 2018) for individuals of provenance up to 38 thousands years BP. Then they grouped these individuals by time periods and compared their average stature with the average PRS score obtained using the UK Biobank summary statistics data. Using this analysis their results showed that the polygenic scores obtained using the GWAS data mirrored their average osteological stature. In Cox et al. (Cox *et al.*, 2022)the authors used 182 individuals ranging from 33,000 to 850 years BP to assess the correlation between GWAS results and stature. They estimated that the GWAS-based analysis explained 6.3% of the variance observed femur length. This work also tested the influence of different factors such as longitude, diet, temperature and precipitation to the height of ancient populations. Only longitude (R-squared = 0.033, SD = 0.008, p-value = 0.011) yielded a significant correlation.

## 4.2 Methods

### 4.2.1 Osteological data collection

The osteological stature data collected for this chapter includes 169 ancient humans ranging from the Palaeolithic to the Mediaeval periods (Table 4.S1). Of these 80 have not been published in any previous ancient DNA analysis. For these, where stature information was not available from the literature femur length was used to estimate height using the regression formulas indicated developed by Christopher Ruff (Ruff *et al.*, 2012):

***For female:*** $2.72 \times Femur\_length + 42.85$

***For male:*** $2.69 \times Femur\_length + 43.56$

Where available from literature N15 stable isotope information was incorporated to inform on the effect of protein intake on the stature of 63 individuals (Table 4.S1).

### 4.2.2 DNA data collection

Diploid genetic information was taken using the same dataset described in the previous chapter (Table 3.S1). With an additional 76 WGS and SNP capture genomes with genotypes imputed following the same approach described in paragraph 3.2.1. In addition, 60 WGS imputed genomes were made available by collaborators and are detailed in Table 4.S1 (Breslin E. Unpublished; Jackson I. Unpublished and (Margaryan *et al.*, 2020)).

### 4.2.6 Local ancestry inference

For each of the 169 samples, at each SNP position, we obtained local ancestry information using the software ELAI(Guan, 2014). This software uses a two layer hidden Markov model to infer local haplotypes from different source populations

into an admixed one. This method has many advantages that make it suitable for use in ancient imputed genomes. First it does not require phasing or a recombination map to detect local haplotype segments. This is particularly important in cases of ancient genomes where phasing can be inaccurate, especially for individuals that fall outside the modern human genetic variation(Günther and Nettelblad, 2019). Moreover the program allows missingness in its input data unlike other local ancestry programs such as HAPMIX(Price *et al.*, 2009). This feature is a particular advantage with ancient samples, especially for SNP capture data, where missingness represents a problem even after imputation. The first step of applying ELAI consists in defining the source populations used as a model to infer the haplotype in the target ones. In this case I have used two populations as sources, one of European and the other of African ancestry. As a European reference I have retained 105 individuals, labelled as EUR from the 1000 Genome project. To describe the African ancestry I have chosen as reference 108 genomes from the Yoruba population (labelled as YRI) published in the 1000 Genomes project. Both dataset were separately filtered for minor allele frequencies of 1% using the software PLINK v. 1.9. Using the whole summary statistics dataset from Yengo et al. I run the program ELAI using the parameters: (-mg 1500 -C 2 -c 10 -s 30).

### 4.2.3 Estimation of polygenic score

The summary statistic data used to calculate polygenic risk score (PRS) focused on 3290 SNPs which were most significantly associated with height. These SNPs were identified as independently and significantly associated with the height through a conditional and joint analysis (COJO) performed by Yengo et al. on a merged resource of approximately 700,000 individuals from the GIANT and the UK Biobank data sets (Yengo *et al.*, 2018). To calculate PRS the software PLINK version 1.9 was used with the option –score and default settings. Following the example of (Cox *et al.*, 2019), using the default option of PLINK, the allele count information for missing genotypes was replaced using the allele frequencies during the PRS calculation. To make sure that missingness was not a source of

directional bias, genotype missingness was plotted versus the calculated PRS (Figure 4.1 and Figure 4.2).

As a replicate test, PRS calculation was repeated using the summary statistical information provided by the Neal (http://www.nealelab.is) using the UK Biobank dataset (http://www.nealelab.is/uk-biobank). A clumping/thresholding approach was applied using the software PLINK. The parameters used to filter were similar to those described in Cox et al. (Cox *et al.*, 2019): a fixed linkage disequilibrium (--clump-kb 250) and variable p-values: $10^{-3}$ and, $10^{-4}$. After clumping we obtained respectively for these two p-value thresholds approximately 77 thousands and 43 thousands SNPs. These SNPs were used to calculate the PRS score as described for the Yengo-COJO analysis and tested whether missingness was affecting results (Figure 4.2).

In an independent analysis the amount of European ancestry estimated using the program ELAI was used to calculate a weighted PRS score. For each individual, each SNP annotated in the summary statistics data, was weighted using its amount of European ancestry reported by ELAI. To calculate the PRS score I have used the following formula:

$$\textbf{\textit{ELAI PRS:}}\ PRS = \frac{\sum_{i}^{N} S_i \times L_{ij} \times G_{ij}}{N}$$

Here $S_i$ is the beta score of the allele *i* present in the summary data; $G_{ij}$ is the number of alleles *i* present in the individual *j*. $L_{ij}$ is the European local ancestry proportion estimated by ELAI for SNP *i* in individual *j*. *N* is the total number of SNPs considered during the calculation of the PRS. The program that implemented this formula was created by me and will soon be available on github.

### 4.2.4 Regression analyses

The osteological stature of each sample was regressed as a function of the PRS alone and for male and females separately. In cases where both sex were considered together, we applied a correction to the male statures. This correction value was calculated by taking the difference between mean stature values for male and females. Once obtained, this value was subtracted from each of the males' heights (Figure 4.3A; Figure 4.4A , 4.4C; Figure 4.5A). Where isotope information was available I also built a regression model using height as a dependent variable and the PRS, sex and nitrogen (N15) as independent variables (*Height ~ PRS + sex + N15*).. For each model, the correlation values were obtained using the Pearson formula through the function *cor* in R. A p-value from correlation was also obtained using the function lm in R. To investigate change in stature and PRS through time we first grouped the individuals by time period (Table 4.S1). The stature for male individuals was adjusted following the method previously described and for both stature and PRS we compared different groups using a pairwise Wilcoxon test and adjusted the p-values using Hommel's method (Hommel, 1988).

### 4.2.5 BMI analysis

To investigate the distribution of body mass index (BMI) genetic values across European Neolithic populations PRS was calculated for the 247 Neolithic individuals described in chapter 2 using the summary statistics calculated by the Neale Lab (http://www.nealelab.is/uk-biobank) using the UK BioBank resource. Prior to obtaining PRS information individuals with more than 30% of BMI SNPs missing were filtered and Missing genotypes were excluded from this analysis. SNPs in this dataset were filtered using a clumping/threshold approach through the software PLINK 1.9 with parameters (--clump-p1 0.01 --clump-kb 1000 --clump-r2 0.1). After filtering approximately 12 thousands SNPs remained that were used to compute the PRS in 247 ancient samples using the --score option in PLINK (Figure 4.8).

## 4.3 Results

### 4.3.1 Predicting height using GWAS data

PRS for height were calculated for 169 individuals using three different reference summary statistics datasets. The first focuses on 3290 SNPs described in Yengo et al. (2018) (Table 4.S1); The other two use of 77413 and 42767 SNPS obtained after filtering the summary statistics from Neal (http://www.nealelab.is/uk-biobank) using a clumping threshold approach with p-values respectively of $p = 10^{-4}$ and $p = 10^{-3}$. It was first tested whether genotype missingness biassed the PRS scores. When using the summary statistics from Yengo et al. a non-significant correlation of $R = -0.14$ ($p = 0.08$) associated with the correlation between PRS and missingness was observed (Figure 4.1). Repeating the same analysis, using the Neale lab datasets filtered for p-values of $10^{-3}$ and, $10^{-4}$, non-significant correlations between genotype missingness and PRS ($R < 0.1; p > 0.05$) also emerged (Figure 4.2).

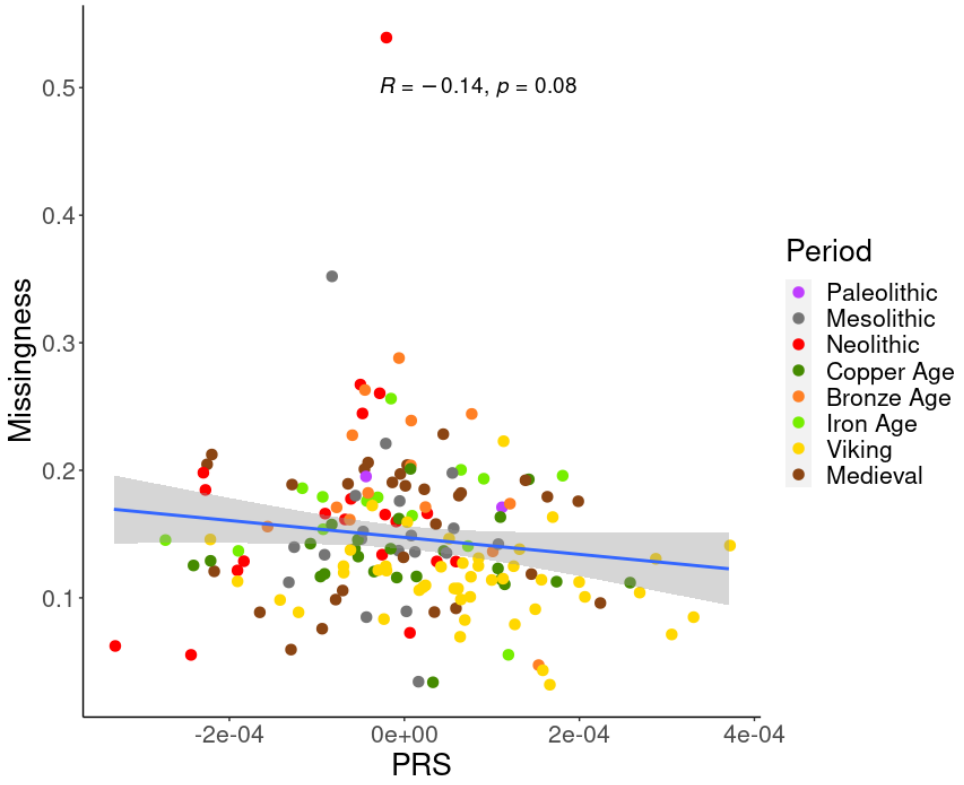


***Figure 4.1*****:** *Plot of genotype missingness versus polygenic risk score (PRS) obtained using 3290 SNPs from Yengo et al. 2018. There is no significant evidence of missingness biassing PRS values.*

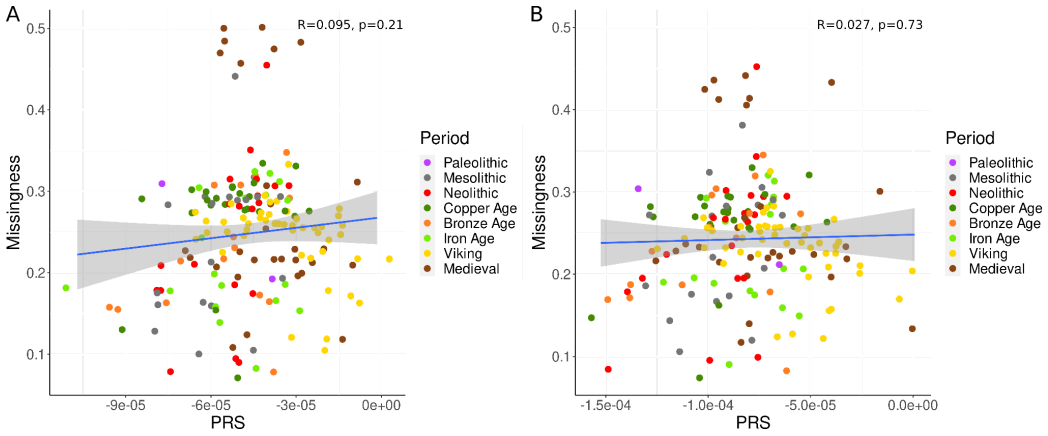


***Figure 4.2:*** *Missingness versus PRS score calculated using UK Biobank dataset: A) Analysis done using a p-value threshold during clumping of $10^{-4}$. B) Same as A using a p-value threshold for clumping of $10^{-3}$.*

When investigating the correlation between PRS and stature, Cox and colleagues reported that the PRS score alone explained 5% of height in 153 ancient individuals ($R^2 = 0.052$)(Cox *et al.*, 2022). Similarly, using the UK Biobank as a reference to calculate the PRS score, Marciniak and colleagues estimated that PRS explained 4% of observed stature ($R^2 = 0.043$) among 160 archaeological samples. Using the Yengo-COJO dataset as reference current results markedly improved on this with height and PRS scores showing a significant correlation (R=0.45) with a $R^2$ of 0.19 (Figure 4.3A; Table 4.1). Correlation coefficients and $R^2$ values were respectively R=0.46 and $R^2$=0.22 for males and R=0.38 and $R^2$=0.12 for females when considered separately (Figure 4.3B; Table 4.1). Using the summary statistics from the Neale lab as reference correlation and variance explained values were slightly lower than those obtained using the Yengo-COJO dataset (Figure 4.4; Table 4.1). When comparing these results with the ones obtained by using the same summary statistics on modern European populations the variance explained values are, as expected, slightly lower than the ones reported here. Using the UK Biobank height dataset on modern European populations the variance explained percentage was of 49%. Using the Yengo 3290 SNPs instead explains 24% of the variance in height in modern European individuals.

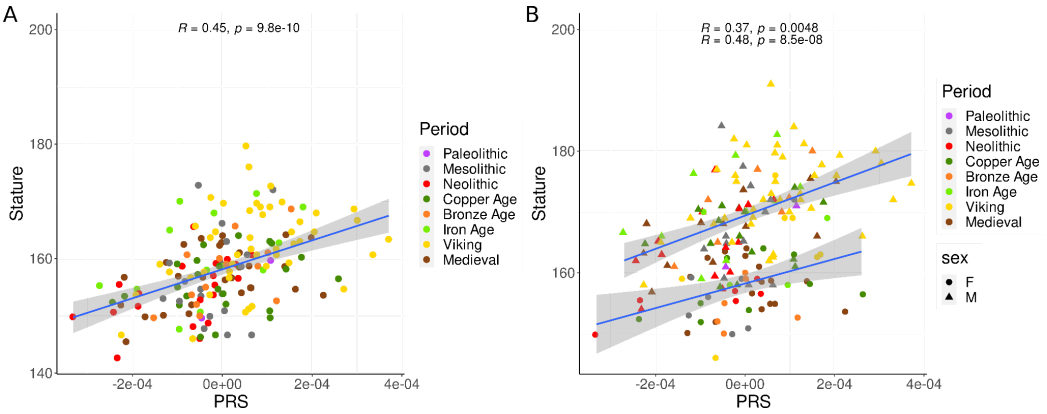


***Figure 4.3:*** *A) Plot of osteological stature versus PRS score obtained using 3290 SNPs from Yengo et al. 2018. Osteological stature for males were corrected by subtracting their value from the difference between male and females average height B) Same data with individuals divided by sex. In both panels the correlation between the variables is significant.*

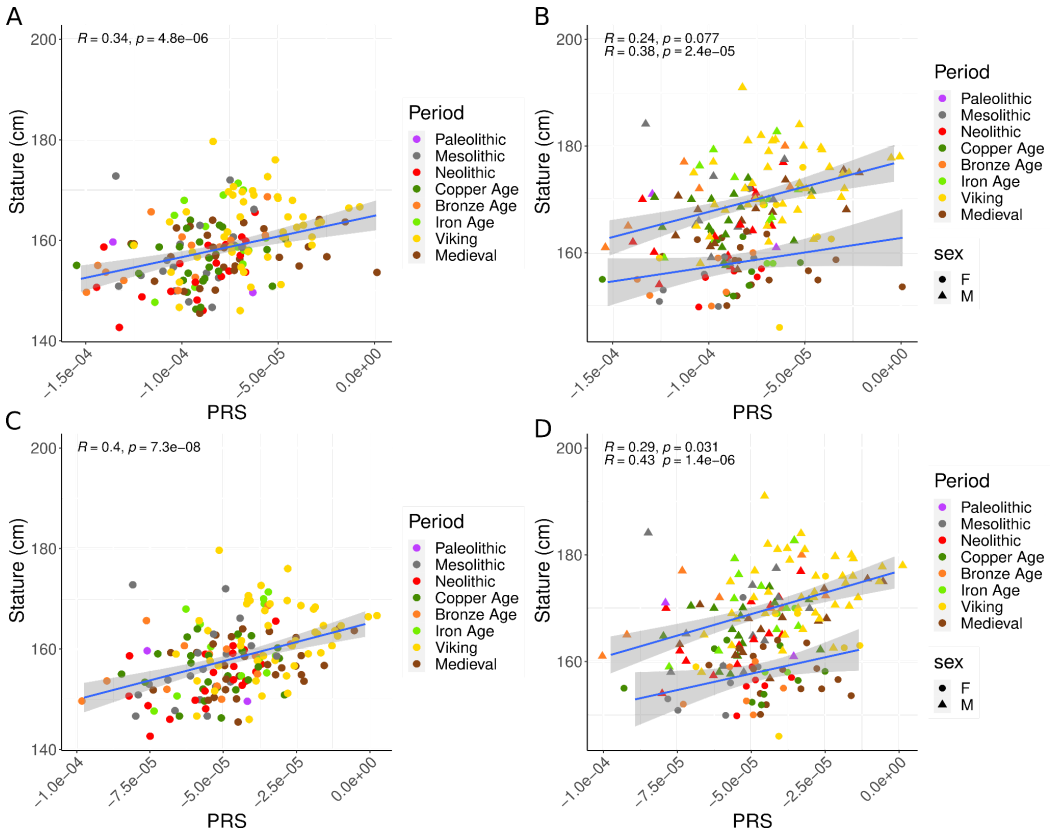


***Figure 4.4: Stature plotted versus PRS score calculated using the summary statistics dataset obtained from the Neale lab:*** *A) Analysis using a p-value threshold of $10^{-4}$ when clumping the summary statistics data with approximately 43 thousands SNPs used to obtain the PRS scores. Sex for male individuals was adjusted by applying a correction value (see Methods) B) Same p-value threshold applied in A with stature estimates considered separately for males and females. C) Analysis using a p-value threshold of $10^{-3}$ applied on while clumping the summary statistics data with approximately 77 thousands SNPs used for the analysis. Sex for males was adjusted as in panel A. D) Same method as applied in panel C,considering male and females separately.*

## 4.3.2 Using local ancestry to predict traits

The application of summary statistics data to predict particular phenotypes is usually limited to populations that are well represented in the GWAS reference dataset. When the GWAS and target dataset have distinct ancestries the power of

PRS can be greatly reduced (Martin *et al.*, 2019). To correct for this I applied a local ancestry method on 169 ancient individuals using the software ELAI and a reference dataset of modern European and African individuals. The amount of European ancestry calculated by the software for each SNP in the Yengo-COJO summary statistics data was multiplied by the beta score of the SNP. The sum of all these adjusted scores were used to build a new PRS score. As we can see in Figure 4.5 and Table 4.1 the variance explained and correlations between PRS and stature, respectively of 0.18 and 0.43, are slightly lower when weighting the polygenic risk score using the local ancestry information. This is true also when considering males and females independently.

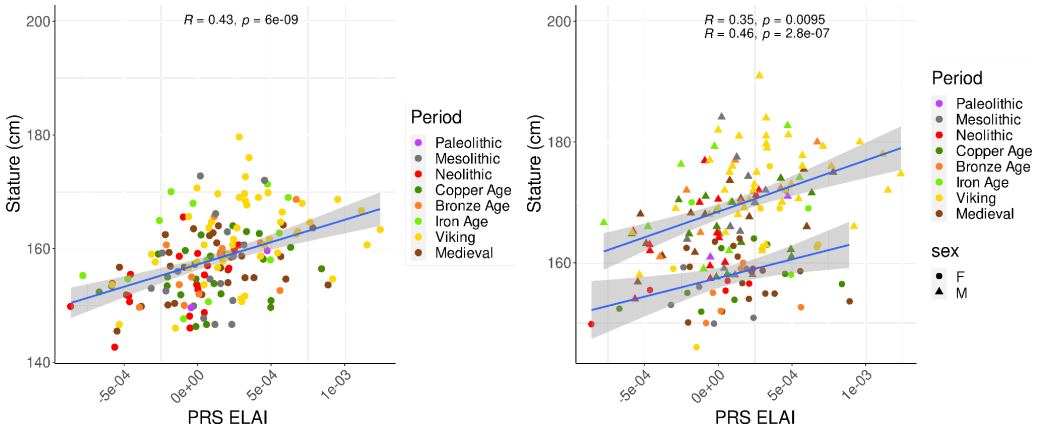


***Figure 4.5:*** *A) Plot of osteological stature versus PRS score weighted by local ancestry using 3290 SNPs from Yengo et al. 2018. Osteological statures for males were corrected by subtracting their value from the difference between male and females average height B) Same data with individuals divided by sex. In both panels the correlations and p-values associated with it are indicated.*

### 4.3.3 Change in osteological and genetic stature through time

Studies of temporal trends in height have been made using both archaeological and genetic data. In a study published in 2018 Christopher Ruff performed a comprehensive set of analyses of heights from 2,179 individuals spanning 30,000 years (Ruff, 2018). In this he estimated that the largest change in stature happened between the Upper and Late Palaeolithic periods. The second highest

change was a drop in stature between mediaeval and early modern populations which was followed by an increase between early modern and the 20th century. Later studies focused on the genetics of ancient populations agreed with these results, inferring a significant shift in height PRS scores between Upper and Late Palaeolithic periods and from Neolithic towards recent times (Cox *et al.*, 2019, 2022; Marciniak *et al.*, 2021). In the current analysis when using the PRS alone, a significant increase in the score between both Mesolithic or Neolithic and Viking groups is clear, with p-values respectively of 0.02 and 0.0003 (Figure 4.6B). When considering stature alone Viking individuals are significantly taller than populations that lived during and Mesolithic, Neolithic, Bronze Age and Mediaeval times with overall p-values lower than 0.05 (Figure 4.6A). Interestingly in this case we also see a significant drop in stature between Viking and Mediaeval cohorts (p-value = $10^{-5}$) that is not observed in the PRS analysis (Figure 4.6B).

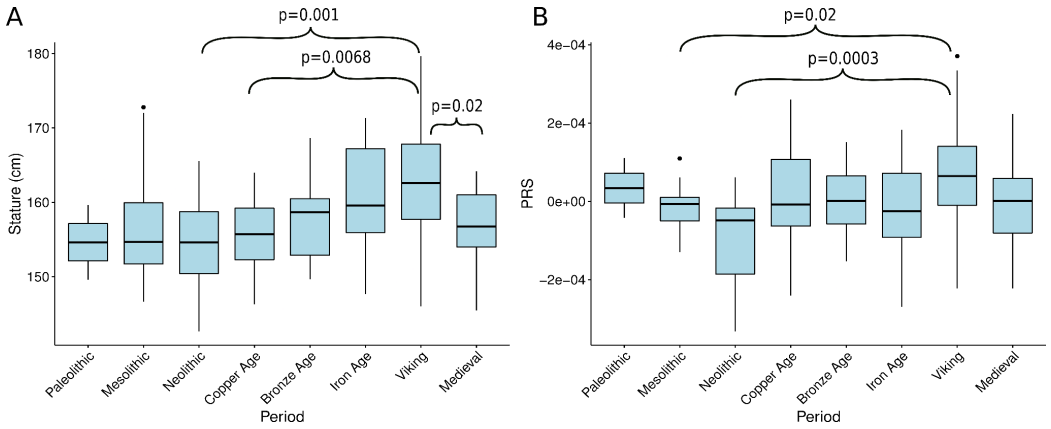


***Figure 4.6: Temporal trends in stature:*** *A) Change in stature across periods with significant shifts highlighted with brackets (p-value < 0.05). Stature for males was corrected using the difference in the average height between males and females; B) Comparisons of polygenic scores (PRS) across periods.*

## 4.3.4 Influence of diet on stature

Like many other traits, the stature of an individual is influenced by multiple factors such as diet and disease. With the introduction of agriculture, some changes in morphology of teeth and body occurred in Neolithic populations. For

example, an overall increase in caries rate is observed, especially in South Asia with the domestication of rice. The health and stature of North American populations decreased in conjunction with the domestication of maize (Richards 2002). To test how the amount of protein intake affected the stature of ancient populations across time we analysed the relation between N15 isotope and stature residuals not explained by genetics (estimated by PRS) and sex. In order to do so the residuals for the stature were obtained using the following regression formula: Stature~PRS + sex. Although the correlation between N15 and residuals is positive (R= 0.23), the p-value associated with this relationship was not significant (Figure 4.7, Table 4.1).

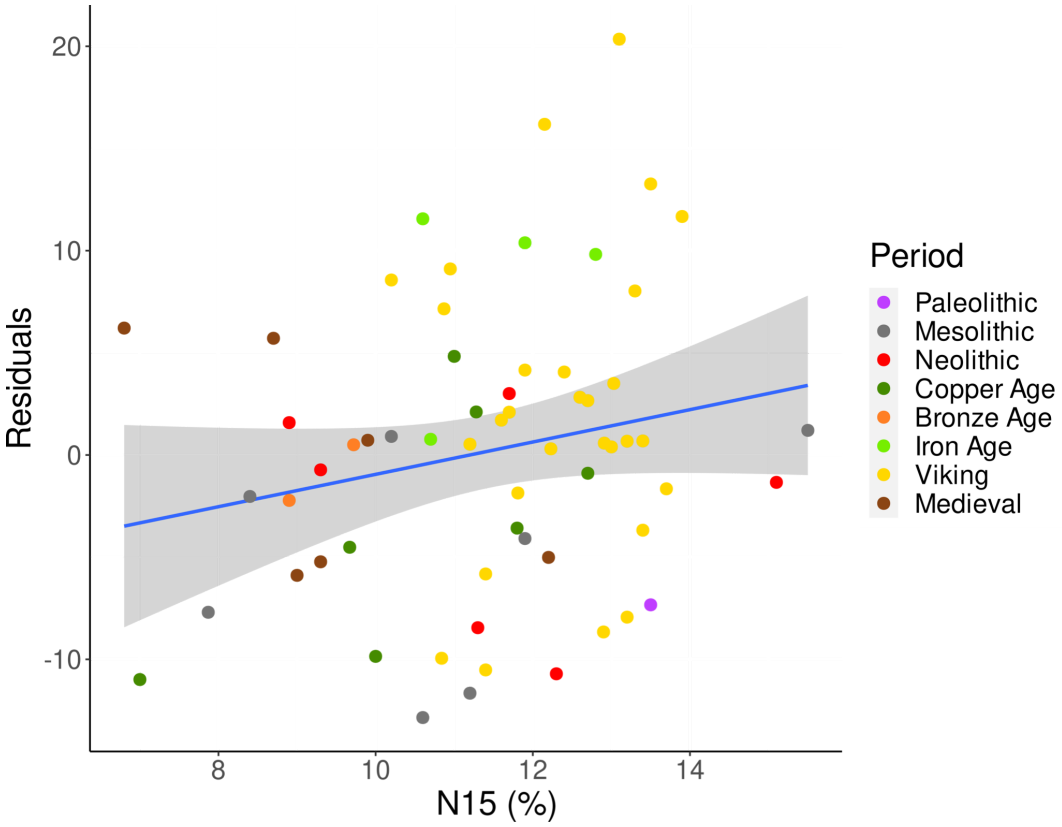


***Figure 4.7:*** *Isotope analysis. A) Nitrogen 15 isotope plotted versus stature residuals. The residuals on the Y-axis are obtained from the regression using Stature as the dependent variable and PRS and sex as independent variables.*

| *Dependent variable* | *Independent variables* | *Source* | *Clumping p-value* | *R-squared* | *p-value* |
|---|---|---|---|---|---|
| *Stature(corrected)* | *PRS* | *Yengo-COJO* | // | 0.19 | $9.8\times10^{-10}$ |
| *Stature(corrected)* | *PRS_ELAI* | | | 0.18 | $6\times10^{-10}$ |
| *Stature Male* | *PRS* | | | 0.22 | $8.5\times10^{-8}$ |
| *Stature Male* | *PRS_ELAI* | | | 0.18 | $2.8\times10^{-7}$ |
| *Stature Female* | *PRS* | | | 0.12 | $4\times10^{-2}$ |
| *Stature Female* | *PRS_ELAI* | | | 0.10 | $9\times10^{-2}$ |
| *Residuals* | *N15* | | | 0.03 | 0.07 |
| *Residuals ELAI* | *N15* | | | 0.02 | 0.09 |
| *Stature(corrected)* | *PRS* | *UK Biobank (Neale lab)* | $10^{-4}$ | 0.11 | $4.8\times10^{-6}$ |
| *Stature Male* | *PRS* | | | 0.14 | $2.4\times10^{-5}$ |
| *Stature Female* | *PRS* | | | 0.04 | 0.07 |
| *Residuals* | *N15* | | | 0.03 | 0.08 |
| *Stature(corrected)* | *PRS* | *UK Biobank (Neale lab)* | $10^{-3}$ | 0.15 | $7.3\times10^{-8}$ |
| *Stature Male* | *PRS* | | | 0.18 | $1.4\times10^{-6}$ |
| *Stature Female* | *PRS* | | | 0.06 | 0.03 |
| *Residuals* | *N15* | | | 0.03 | 0.09 |

**Table 4.1:** *Description of the regression analysis results using different variables*

### 4.3.5 Body mass index analysis of Neolithic populations

The carving and circulation of apparently obese human figurines was a marked feature of the late Maltese Neolithic (Malone and Stoddart, 2016), perhaps mirroring an unusual genetic predisposition within a restricted gene pool. Accordingly, we performed a polygenic risk score analysis on body mass index using the summary statistics from the UK Biobank dataset but found that the three Maltese Neolithic individuals sampled do not give atypical risk values compared with other Neolithic individuals (Figure 4.8).

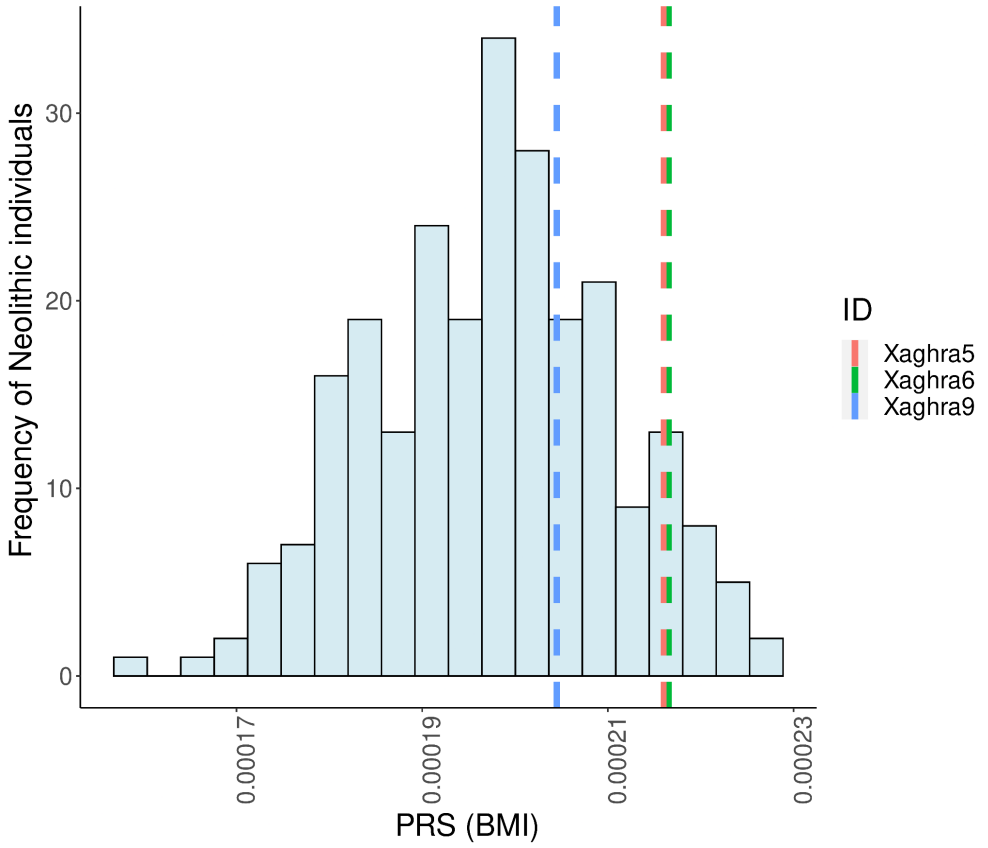


***Figure 4.8: Body mass index in ancient Maltese**. Polygenic risk score (PRS) information calculated from the UK Biobank dataset was used to estimate body mass index (BMI) of the ancient Maltese together with 247 European Neolithic samples. None of the former showed extreme values.*

## 4.4 Discussion

In this work genetic data from GWAS analyses are used to predict the stature of ancient individuals. By making use of both published and unpublished data it demonstrates that it is possible to achieve a good correlation between both these measures. This has been previously shown but the current analysis achieved the highest explanation of phenotypic variance for ancient data, probably due to the use of a focused set of SNPs which are highly correlated with height (Yengo *et al.*, 2018). These high quality predictions helped unveil a temporal pattern or variation in expected genetic stature that is compared with the osteological one. From these results the Viking group is noted as being both physically and genetically taller than people from the Neolithic period. This was not true for Copper Age, Mesolithic and Mediaeval populations where significant differences with the

Viking population observed for expected height were not matched in measured statures. One of the reasons for this could be that other factors besides genetics, such as living conditions, introduce variation that leave the tests underpowered to detect significance. Analyses using isotopic data did not show any detectable significant influence of protein intake in the stature of ancient individuals - although a trend is visible. Using a different set of SNPs it was also addressed whether the Maltese were particularly affected by obesity compared to other European Neolithic populations. This test was motivated by the fact that Malta was particularly rich in obese figurines during the Temple Period, possibly indicating status within the society. However our results did not show three Maltese genomes as possessing a particularly high predicted body mass index compared to other contemporary populations

# Supplementary Material

| ID | Sex | Country | Site | Period | Date BP | Stature (cm) | ðC | ðN | Stature Reference | DNA Reference | Isotope Reference | PRS Yengo-Cojo |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3DT16 | M | Great Britain | Driffield Terrace | Iron Age | 1750 | 171.8 | -19.4 | 10.7 | (Cox *et al.*, 2022) | (Martiniano *et al.*, 2016) | (Cox *et al.*, 2022) | 6.83E-05 |
| 3DT26 | M | Great Britain | Driffield Terrace | Iron Age | 1750 | 179.3 | -19.3 | 11.9 | (Cox *et al.*, 2022) | (Martiniano *et al.*, 2016) | (Cox *et al.*, 2022) | -1.70E-05 |
| 6DT22 | M | Great Britain | Driffield Terrace | Iron Age | 1750 | 176.3 | -19 | 12.8 | (Cox *et al.*, 2022) | (Martiniano *et al.*, 2016) | (Cox *et al.*, 2022) | -0.000115207 |
| 6DT3 | M | Great Britain | Driffield Terrace | Iron Age | 1750 | 182.7 | // | // | (Cox *et al.*, 2022) | (Martiniano *et al.*, 2016) | // | 7.28E-05 |
| AH1 | F | Iran | Tepe Abdul Hosein | Neolithic | 9925 | 159 | // | // | (Broushaki *et al.*, 2016) | (Broushaki *et al.*, 2016) | // | 3.03E-05 |
| ANN1 | M | Ireland | Annagh | Neolithic | 5405 | 170 | // | // | (Cahill and Sikora, 2011) | (Jones *et al.*, 2017; Martiniano *et al.*, 2017; Cassidy *et al.*, 2020; | // | -9.15E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Margaryan *et al.*, 2020) | | |
| Ajvide58 | M | Sweden | Ajvide | Neolithic | 4700 | 154 | // | // | (Marciniak *et al.*, 2021) | (Skoglund, Malmström, *et al.*, 2014) | // | -0.00022972 |
| Ardu2 | M | Estonia | Harju | Bronze Age | 4652 | 177 | // | // | (Kriiska, Lavento and Peets, 2005) | (Saag *et al.*, 2017) | // | 2.46E-05 |
| BOT2016 | F | Kazakhstan | Botai | Copper Age | 5448 | 164 | -19.3 | 11 | (de Barros Damgaard *et al.*, 2018) | (de Barros Damgaard *et al.*, 2018) | (de Barros Damgaard *et al.*, 2018) | 3.71E-05 |
| Bichon | M | Switzerland | Grotte du Bichon | Mesolithic | 13698 | 164.8 | // | // | (Cox *et al.*, 2022) | (Mathieson *et al.*, 2018) | // | -4.26E-05 |
| Canes | F | Spain | Los Canes | Mesolithic | 7089 | 149.91 | -20 | 7.87 | (Cox *et al.*, 2022) | (González-Fortes *et al.*, 2017) | (González-Fortes *et al.*, 2017) | -2.60E-05 |
| Chan_Meso | F | Spain | Chan do Lindeiro | Mesolithic | 9109 | 153 | -20.5 | 8.4 | (Marciniak *et al.*, 2021) | (González-Fortes *et al.*, 2017) | (González-Fortes *et al.*, 2017) | -0.00013007 |
| Cheddar man | M | England | Gough's Cave | Mesolithic | 8000 | 166 | // | // | (Marciniak *et al.*, 2021) | (Brace *et al.*, 2019) | // | -4.59E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CrKo9 | M | Croatia | Koprivno | Medieval | -1450 | 173.64 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | -7.20E-06 |
| CrPr10 | M | Croatia | Privlaka | Medieval | -850 | 168.07 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | -0.000217032 |
| CrPr14 | F | Croatia | Privlaka | Medieval | -850 | 151.91 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | -6.34E-05 |
| CrPr19 | M | Croatia | Privlaka | Medieval | -850 | 170.63 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | -9.72E-06 |
| CrPr193 | M | Croatia | Privlaka | Medieval | -850 | 163.2 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | -0.000222832 |
| CrPr35 | M | Croatia | Privlaka | Medieval | -850 | 165.99 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | 0.000163954 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CrPr45 | F | Croatia | Privlaka | Medieval | -850 | 154.88 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | 6.48E-05 |
| CrPr8 | F | Croatia | Privlaka | Medieval | -850 | 154.88 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | 5.94E-05 |
| CrPr96 | M | Croatia | Privlaka | Medieval | -850 | 167.61 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | 3.70E-06 |
| CrRu16 | F | Croatia | Rudina | Medieval | -1500 | 162.78 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | 1.17E-06 |
| CrSt62 | F | Croatia | Stenjevec | Medieval | -1200 | 159.32 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | -0.000124997 |
| CrSt73 | M | Croatia | Stenjevec | Medieval | -1200 | 168.07 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | -4.53E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DA247 | M | Russia | Shamanka | Neolithic | 7689 | 161.986 | -16.8 | 15.1 | (Cox *et al.*, 2022) | (de Barros Damgaard *et al.*, 2018) | (Cox *et al.*, 2022) | -0.000242606 |
| DA252 | F | Russia | Shamanka | Neolithic | 7317 | 149.815 | // | // | (Cox *et al.*, 2022) | (de Barros Damgaard *et al.*, 2018) | (Cox *et al.*, 2022) | -0.00033197 |
| DA342 | F | Russia | Ust'Ida | Neolithic | 5617 | 155.464 | -18.7 | 11.7 | (Cox *et al.*, 2022) | (de Barros Damgaard *et al.*, 2018) | (Cox *et al.*, 2022) | -0.000233817 |
| DA343 | M | Russia | Ust'Ida | Neolithic | 4819 | 157.362 | -18.7 | 12.3 | (Cox *et al.*, 2022) | (de Barros Damgaard *et al.*, 2018) | (Cox *et al.*, 2022) | -5.12E-05 |
| DA361 | M | Russia | Ust'Ida | Neolithic | 4130 | 160.082 | -19.3 | 11.3 | (Cox *et al.*, 2022) | (de Barros Damgaard *et al.*, 2018) | (Cox *et al.*, 2022) | -3.24E-05 |
| Vestonice16 | M | Czech Republic | Dolni Vestonice | Paleolithic | 29800 | 171 | // | // | (Marciniak *et al.*, 2021) | (Fu *et al.*, 2016) | // | 0.000110259 |
| Dzielnica243 | M | Poland | Dzielnica | Bronze Age | 4113 | 172.34 | // | // | (Marciniak *et al.*, 2021) | (Olalde *et al.*, 2018) | // | 0.000102568 |
| ElMiron | F | Spain | El Miron | Mesolithic | 18775 | 159 | -18.2 | 10.2 | (Marciniak *et al.*, 2021) | (Fu *et al.*, 2016) | (D'Anglade and Gorosquieta, 2017) | -6.50E-06 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GB1_Eneo | F | Romania | Gura Baciului | Copper Age | 3465 | 155 | -20 | 12.7 | (Marciniak *et al.*, 2021) | (González-Fortes *et al.*, 2017) | (González-Fortes *et al.*, 2017) | -9.52E-05 |
| GEN16a | F | Hungary | Alsónémedi | Copper Age | 5122 | 151.88 | // | // | (Marciniak *et al.*, 2021) | (Haak *et al.*, 2015; Lipson *et al.*, 2017; Rivollat *et al.*, 2020) | | -9.85E-05 |
| GEN59 | M | Hungary | Felső Ürge-hegyi dűlő | Bronze Age | 4254 | 170 | -19.71 | 9.72 | (Marciniak *et al.*, 2021) | (Olalde *et al.*, 2018) | (Olalde *et al.*, 2018) | 6.56E-06 |
| HAJE7a | M | Hungary | Hajdúnánás-Eszlári út | Neolithic | 7114 | 159.45 | // | // | (Marciniak *et al.*, 2021) | (Haak *et al.*, 2015; Lipson *et al.*, 2017; Rivollat *et al.*, 2020) | | -6.30E-05 |
| Hum1 | F | Norway | Hummervikholmen | Mesolithic | 9364 | 156 | // | // | (Sellevold and Skar, 1994) | (Günther *et al.*, 2018) | (Günther *et al.*, 2018) | -8.12E-05 |
| Hung849 | M | Hungary | Mezöcsát-Hörcsögös | Bronze Age | 5100 | 172 | // | // | (Marciniak *et al.*, 2021) | (Olalde *et al.*, 2018) | | -7.84E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I0059 | F | Germany | Benzingerode | Neolithic | 4179 | 157 | -20.3 | 9.3 | (Cox *et al.*, 2022) | (Mathieson *et al.*, 2015) | (Cox *et al.*, 2022) | -2.12E-05 |
| I0156 | M | Great Britain | Hinxton | Iron Age | 1981 | 159 | // | // | (Cox *et al.*, 2022) | (Schiffels *et al.*, 2016) | // | -9.26E-05 |
| I0157 | F | Great Britain | Hinxton | Medieval | 1227 | 158.6 | // | // | (Cox *et al.*, 2022) | (Schiffels *et al.*, 2016) | // | 0.000142287 |
| I0159 | F | Great Britain | Hinxton | Medieval | 1234 | 153.6 | // | // | (Cox *et al.*, 2022) | (Schiffels *et al.*, 2016) | // | 0.000223255 |
| I0160 | M | Great Britain | Hinxton | Iron Age | 1973 | 174.1 | // | // | (Cox *et al.*, 2022) | (Schiffels *et al.*, 2016) | // | 0.000121635 |
| I0161 | F | Great Britain | Hinxton | Medieval | 1157 | 163.5 | // | // | (Cox *et al.*, 2022) | (Schiffels *et al.*, 2016) | // | -4.35E-05 |
| I0698 | M | Bulgaria | Yabalkovo | Neolithic | 7900 | 170.5 | // | // | (Cox *et al.*, 2022) | (Mathieson *et al.*, 2018) | // | -1.08E-05 |
| I0777 | F | Great Britain | Oakington | Medieval | 1465 | 161 | // | // | (Cox *et al.*, 2022) | (Schiffels *et al.*, 2016) | // | 3.51E-05 |
| I0789 | F | England | Linton | Iron Age | 2163 | 158 | // | // | (Cox *et al.*, 2022) | (Schiffels *et al.*, 2016) | // | -3.39E-05 |
| I13219 | F | Pakistan | Swat Valley, Butkara II | Iron Age | 2850 | 169 | // | // | (Cox *et al.*, 2022) | (Narasimhan *et al.*, 2019) | // | 0.000182974 |
| I1784 | M | Turkmenistan | Gonur | Bronze Age | 4060 | 170 | // | // | (Dubova and Rykushina, 2007) | (Narasimhan *et al.*, 2019) | // | 7.85E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I1799 | F | Pakistan | Udegram | Iron Age | 3003 | 170 | // | // | (Vidale, Micheli and Olivieri, 2016) | (Narasimhan *et al.*, 2019) | // | -9.13E-05 |
| I1947 | M | Iran | Ganj Dareh | Neolithic | 9992 | 172 | // | // | http://repository.edition-topoi.org/collection/LIVE/ | (Narasimhan *et al.*, 2019) | // | 6.16E-05 |
| I1985 | M | Pakistan | Udegram | Iron Age | 2939 | 166 | // | // | (Vidale, Micheli and Olivieri, 2016) | (Narasimhan *et al.*, 2019) | // | -0.000187421 |
| I2369 | M | Hungary | Budakalász-Luppa csárda | Neolithic | 5186 | 176.9 | // | // | (Cox *et al.*, 2022) | (Haak *et al.*, 2015; Lipson *et al.*, 2017; Rivollat *et al.*, 2020) | // | -6.41E-05 |
| I2380 | M | Hungary | Mezőkövesd-Mocsolyás | Neolithic | 7350 | 164.98 | // | // | (Marciniak *et al.*, 2021) | (Haak *et al.*, 2015; Lipson *et al.*, 2017; | // | -1.97E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Rivollat *et al.*, 2020) | | |
| I2788 | M | Hungary | Abony, Turjányos -dűlő | Copper Age | 5685 | 170.4 | // | // | (Cox *et al.*, 2022) | (Haak *et al.*, 2015; Lipson *et al.*, 2017; Rivollat *et al.*, 2020) | // | -0.000104667 |
| I2790 | F | Hungary | Abony, Turjányos -dűlő | Copper Age | 5625 | 162.4 | // | // | (Cox *et al.*, 2022) | (Haak *et al.*, 2015; Lipson *et al.*, 2017; Rivollat *et al.*, 2020) | // | -3.75E-05 |
| I2791 | M | Hungary | Abony, Turjányos -dűlő | Copper Age | 5523 | 164.8 | // | // | (Cox *et al.*, 2022) | (Haak *et al.*, 2015; Lipson *et al.*, 2017; Rivollat *et al.*, 2020) | // | -8.11E-05 |
| I2925 | F | Iran | Tepe Hissar | Copper Age | 4715 | 158.154 | -19.7 | 11.8 | (Cox *et al.*, 2022) | (Narasimhan *et al.*, 2019) | (Cox *et al.*, 2022) | 0.000140406 |
| I4075 | F | Netherlands | De Tuithoorn | Copper Age | 3950 | 163 | -21.08 | 11.28 | (Cox *et al.*, 2022) | (Olalde *et al.*, 2018) | (Cox *et al.*, 2022) | 0.000106599 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | , Oostwoud , Noord-Holland | | | | | | | | | |
| I4241 | M | Iran | Hajji Firuz | Copper Age | 7904 | 157.634 | // | // | (Cox *et al.*, 2022) | (Narasimhan *et al.*, 2019) | // | -5.27E-05 |
| I4582 | F | Romania | Ostrovul Corbului | Mesolithic | 8618 | 172 | // | // | (Cox *et al.*, 2022) | (Mathieson *et al.*, 2018) | // | 0.000109479 |
| I5232 | M | Serbia | Padina | Mesolithic | 7901 | 163.89 | // | // | (Cox *et al.*, 2022) | (Mathieson *et al.*, 2018) | // | -9.45E-05 |
| I5233 | F | Serbia | Padina | Mesolithic | 8001 | 159.23 | // | // | (Cox *et al.*, 2022) | (Mathieson *et al.*, 2018) | // | -0.000121262 |
| I6579 | F | Poland | Iwiny | Copper Age | 4155 | 153 | // | // | (Marciniak *et al.*, 2021) | (Olalde *et al.*, 2018) | // | -2.08E-05 |
| I5260 | F | Hungary | Enese elkerülő, Kóny, Proletár-dülö, M85, Site 2 | Copper Age | 6183 | 152.37 | // | // | (Marciniak *et al.*, 2021) | (Haak *et al.*, 2015; Lipson *et al.*, 2017; Rivollat *et al.*, 2020) | // | -0.000240595 |
| I5514 | M | Czech Republic | Jinonice | Copper Age | 4155 | 162.2 | -18.9 | 10 | (Cox *et al.*, 2022) | (Olalde *et al.*, 2018) | (Cox *et al.*, 2022)1 | 0.000109622 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I6194 | M | Pakistan | Udegram | Iron Age | 2950 | 162 | // | // | (Vidale, Micheli and Olivieri, 2016) | (Narasimhan *et al.*, 2019) | // | -4.56E-05 |
| I6197 | M | Pakistan | Udegram | Iron Age | 2950 | 170 | // | // | (Vidale, Micheli and Olivieri, 2016) | (Narasimhan *et al.*, 2019) | // | 7.84E-06 |
| I6680 | M | England | Northumberland | Copper Age | 3671 | 161 | // | // | (Waddington and Bonsall, 2016) | (Olalde *et al.*, 2018) | // | 0.000110429 |
| I6775 | M | England | Wick Barrow | Copper Age | 4150 | 174 | // | // | (St. George Gray, 1908) | (Olalde *et al.*, 2018) | // | 0.000114341 |
| I6901 | F | Pakistan | Udegram | Iron Age | 2950 | 158 | // | // | (Vidale, Micheli and Olivieri, 2016) | (Narasimhan *et al.*, 2019) | // | 9.27E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I7201 | F | Czech Republic | Jinonice | Bronze Age | 4250 | 152.62 | // | // | (Marciniak *et al.*, 2021) | (Olalde *et al.*, 2018) | // | 0.000120804 |
| I7210 | M | Czech Republic | Radovesice | Copper Age | 4071 | 163.38 | // | // | (Cox *et al.*, 2022) | (Olalde *et al.*, 2018) | // | 1.06E-05 |
| I7275 | M | Czech Republic | Brandýsek | Copper Age | 4200 | 166 | // | // | (Cox *et al.*, 2022) | (Olalde *et al.*, 2018) | // | -1.12E-05 |
| I7278 | M | Czech Republic | Brandýsek | Copper Age | 4333 | 170 | // | // | (Cox *et al.*, 2022) | (Olalde *et al.*, 2018) | // | -4.99E-06 |
| I7282 | M | Czech Republic | Radovesice | Copper Age | 4200 | 168.56 | // | // | (Cox *et al.*, 2022) | (Olalde *et al.*, 2018) | // | -5.74E-05 |
| I7286 | M | Czech Republic | Radovesice | Copper Age | 4221 | 171.57 | // | // | (Cox *et al.*, 2022) | (Olalde *et al.*, 2018) | // | 0.00017301 |
| I8220 | M | Pakistan | Aligrama | Iron Age | 2566 | 166.61 | // | // | (Cox *et al.*, 2022) | (Narasimhan *et al.*, 2019) | // | -0.000270038 |
| Kostenki14 | M | Russia | Markina Gora | Paleolithic | 38052 | 160.95 | -18.2 | 13.5 | (Marciniak *et al.*, 2021) | (Seguin-Orlando *et al.*, 2014) | (Kotlyakov, Velichko and Vasil'ev, 2017) | -4.26E-05 |
| LBK2155 | F | Germany | Viesenhaeuser Hof, Stuttgart- | Neolithic | 7125 | 156.54 | // | // | (Marciniak *et al.*, 2021) | (Haak *et al.*, 2015) | // | 3.43E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Muehlhausen | | | | | | | | | |
| LGCS1a | M | Hungary | Gata-Csatola | Neolithic | 7600 | 164.14 | // | // | (Marciniak *et al.*, 2021) | (Haak *et al.*, 2015; Lipson *et al.*, 2017; Rivollat *et al.*, 2020) | // | -4.61E-05 |
| LaBrana1 | M | Spain | La Brana | Mesolithic | 6980 | 158 | -19.3 | 10.6 | (Marciniak *et al.*, 2021) | (Mathieson *et al.*, 2015) | // | 6.10E-05 |
| Loschbour | M | Luxembourg | Echternach | Mesolithic | 8050 | 165.2 | -20.3 | 11.9 | (Cox *et al.*, 2022) | (Lazaridis *et al.*, 2014) | (Drucker *et al.*, 2018) | -1.99E-06 |
| SC1_Meso | M | Romania | Schela Cladovei | Mesolithic | 8814 | 184.13 | −18.5 | 15 | (Cox *et al.*, 2022) | (González-Fortes *et al.*, 2017) | (González-Fortes *et al.*, 2017) | -5.36E-05 |
| SC2_Meso | M | Romania | Schela Cladovei | Mesolithic | 8825 | 159.09 | −19.1 | 14.7 | (Cox *et al.*, 2022) | (González-Fortes *et al.*, 2017) | (González-Fortes *et al.*, 2017) | -4.22E-05 |
| OC1_Meso | M | Romania | Ostrovul Corbului | Mesolithic | 8277 | 172 | -18.7 | 15.5 | (Marciniak *et al.*, 2021) | (González-Fortes *et al.*, 2017) | (González-Fortes *et al.*, 2017) | 5.88E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I5077 | M | Croatia | Osijek | Neolithic | 7001 | 163 | // | // | (Marciniak *et al.*, 2021) | (Mathieson *et al.*, 2018) | // | -0.000183923 |
| I5078 | M | Croatia | Osijek | Neolithic | 6571 | 165.21 | // | // | (Marciniak *et al.*, 2021) | (Mathieson *et al.*, 2018) | // | -0.000190332 |
| I5236 | M | Serbia | Padina | Mesolithic | 10038 | 177.49 | // | // | (Marciniak *et al.*, 2021) | (Mathieson *et al.*, 2018) | // | -2.15E-06 |
| R104 | M | Italy | Crypta Balbi | Medieval | 1319 | 161.714 | -20 | 9 | (Cox *et al.*, 2022) | (Antonio *et al.*, 2019) | (Cox *et al.*, 2022) | -7.00E-05 |
| R105 | M | Italy | Crypta Balbi | Medieval | 1450 | 168.242 | // | // | (Cox *et al.*, 2022) | (Antonio *et al.*, 2019) | // | -9.36E-05 |
| R106 | F | Italy | Crypta Balbi | Medieval | 1450 | 158.692 | // | // | (Cox *et al.*, 2022) | (Antonio *et al.*, 2019) | // | 5.87E-05 |
| R108 | M | Italy | Crypta Balbi | Medieval | 1450 | 161.17 | // | // | (Cox *et al.*, 2022) | (Antonio *et al.*, 2019) | // | -0.000169099 |
| R123 | M | Italy | Casale del Dolce | Medieval | 1780 | 156.818 | // | // | (Cox *et al.*, 2022) | (Antonio *et al.*, 2019) | // | -0.000213109 |
| R126 | F | Italy | Casale del Dolce | Medieval | 1650 | 162.458 | -20.5 | 6.8 | (Cox *et al.*, 2022) | (Antonio *et al.*, 2019) | (Cox *et al.*, 2022) | -8.09E-05 |
| R7 | M | Italy | Grotta Continenza | Mesolithic | 10706 | 158 | -17.5 | 11.2 | (Ruff, 2018) | (Antonio *et al.*, 2019) | (Antonio *et al.*, 2019) | 1.29E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R80 | F | Italy | Viale Rossini | Medieval | 1840 | 150.084 | -19.5 | 12.2 | (Cox *et al.*, 2022) | (Antonio *et al.*, 2019) | (Cox *et al.*, 2022) | -0.000127401 |
| RDVS67 | M | Czech Republic | Radovesice | Copper Age | 4289 | 164.76 | // | // | (Marciniak *et al.*, 2021) | (Olalde *et al.*, 2018) | | -0.000220473 |
| RISE98 | M | Sweden | Lilla Bedinge | Neolithic | 4082 | 171.2 | -20.87 | 8.9 | (Cox *et al.*, 2022) | (Mathieson *et al.*, 2018) | (Cox *et al.*, 2022) | 1.13E-05 |
| Rathlin1 | M | Ireland | Rathlin Island | Bronze Age | 3897 | 180 | // | // | (Marciniak *et al.*, 2021) | (Gamba *et al.*, 2014; Skoglund, Malmström, *et al.*, 2014; Haak *et al.*, 2015; Mathieson *et al.*, 2015; Cassidy *et al.*, 2016; Seguin-Orlando *et al.*, 2021) | | 0.000151107 |
| SZ19 | F | Hungary | Szólád | Medieval | 1456 | 154 | -18.2 | 9.3 | (Alt *et al.*, 2014) | (Amorim *et al.*, 2018) | (Alt *et al.*, 2014) | 3.93E-05 |
| SZ22 | M | Hungary | Szólád | Medieval | 1442 | 175 | -19.4 | 9.9 | (Alt *et al.*, 2014) | (Amorim *et al.*, 2018) | (Alt *et al.*, 2014) | 0.00019936 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SZ9 | F | Hungary | Szólád | Medieval | 1442 | 164 | -18.6 | 8.7 | (Alt *et al.*, 2014) | (Amorim *et al.*, 2018) | (Alt *et al.*, 2014) | 1.45E-06 |
| CrKo9 | F | Slovenia | Kranj | Medieval | -1450 | 156.61 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | 1.83E-05 |
| SlKr380 | M | Slovenia | Kranj | Medieval | -850 | 175.5 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | 0.000142586 |
| SlKr386 | F | Slovenia | Kranj | Medieval | -850 | 158.5 | // | // | Jackson et al. Unpublished | Jackson et al. Unpublished | // | 4.16E-05 |
| I2424 | F | Bulgaria | Smyadovo | Copper Age | 6288 | 153.85 | -19.65 | 9.67 | Marciniak et al. 2021 | (Mathieson *et al.*, 2018) | (Mathieson *et al.*, 2018) | 4.52E-06 |
| Sope_RISE00 | F | Estonia | Sope | Bronze Age | 4442 | 155 | -21.3 | 8.9 | (Rasmussen *et al.*, 2015) | (Saag *et al.*, 2017) | (Rasmussen *et al.*, 2015) | -4.16E-05 |
| TV31134 | F | Portugal | Torre Velha | Bronze Age | 1600 | 150 | // | // | Fidalgo 2014 | Breslin et al. Unpublished | // | -4.59E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TV32032extra | M | Portugal | Torre Velha | Bronze Age | 3550 | 165 | // | // | (Fidalgo, 2014) | (Martiniano *et al.*, 2016) | // | -6.46E-05 |
| TV32033 | M | Portugal | Torre Velha | Bronze Age | 1600 | 161 | // | // | (Fidalgo, 2014) | Breslin et al. Unpublished | // | -0.000153414 |
| TV32069 | F | Portugal | Torre Velha | Bronze Age | 1600 | 152 | // | // | (Fidalgo, 2014) | Breslin et al. Unpublished | // | -6.17E-05 |
| TV32203 | F | Portugal | Torre Velha | Bronze Age | 1600 | 160 | // | // | (Fidalgo, 2014) | Breslin et al. Unpublished | // | -3.86E-06 |
| TV32241 | F | Portugal | Torre Velha | Bronze Age | 1600 | 159 | // | // | (Fidalgo, 2014) | Breslin et al. Unpublished | // | 5.34E-06 |
| Ötzi | M | Italy | Oetz valley | Copper Age | 3200 | 158 | -21.2 | 7 | (Marciniak *et al.*, 2021) | (Keller *et al.*, 2012) | (Macko *et al.*, 1999) | -1.41E-05 |
| Urzi48 | M | Romania | Urziceni | Copper Age | 5636 | 170.05 | // | // | (Marciniak *et al.*, 2021) | (Mathieson *et al.*, 2018) | // | 4.29E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| V229 | M | Netherlands | De Tuithoorn, Oostwoud, Noord-Holland | Copper Age | 3966 | 171 | // | // | (Marciniak *et al.*, 2021) | (Olalde *et al.*, 2018) | // | -5.62E-05 |
| VK118 | F | Norway | Trondheim | Viking | 750 | 162.58 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 0.000162631 |
| VK150 | M | England | St John's College Oxford | Viking | 1010 | 179 | // | // | (Falys, no date) | (Margaryan *et al.*, 2020) | // | 8.66E-05 |
| VK151 | M | England | St John's College Oxford | Viking | 1010 | 184 | // | // | (Falys, no date) | (Margaryan *et al.*, 2020) | // | 0.000110759 |
| VK165 | M | England | St John's College Oxford | Viking | 1010 | 170 | -19.69 | 11.81 | (Falys, no date) | (Margaryan *et al.*, 2020) | // | 0.000101762 |
| VK170 | M | Isle of Man | Balladoole | Viking | 1050 | 176 | // | // | (Ratican, 2020) | (Margaryan *et al.*, 2020) | // | 0.000199812 |
| VK173 | M | England | St John's College Oxford | Viking | 1010 | 168 | -19.78 | 12.23 | (Falys, no date) | (Margaryan *et al.*, 2020) | // | -6.62E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VK174 | M | England | St John's College Oxford | Viking | 1010 | 180 | -19.96 | 13.03 | (Falys, no date) | (Margaryan *et al.*, 2020) | // | 0.000288954 |
| VK176 | M | England | St John's College Oxford | Viking | 1010 | 173 | -19.86 | 12.91 | (Falys, no date) | (Margaryan *et al.*, 2020) | // | 0.000124339 |
| VK178 | M | England | St John's College Oxford | Viking | 1010 | 177 | -19.53 | 10.95 | (Falys, no date) | (Margaryan *et al.*, 2020) | // | -5.82E-05 |
| VK204 | M | Scotland | Newark Deerness | Viking | 1000 | 175 | // | // | http://www.orkneyjar.com/archaeology/dhl/papers/tm/index.html | (Margaryan *et al.*, 2020) | // | -2.32E-05 |
| VK205 | M | Scotland | Newark Bay | Viking | 1501 | 162.53 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 5.89E-05 |
| VK332 | M | Sweden | Oland | Viking | 1085 | 172 | -19.3 | 13.2 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 8.02E-05 |
| VK333 | M | Sweden | Oland | Viking | 1052 | 172 | -19.3 | 11.6 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 3.86E-05 |
| VK334 | F | Sweden | Oland | Viking | 893 | 169 | -19.4 | 13.3 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 0.000109768 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VK335 | F | Sweden | Oland | Viking | 1000 | 169 | -19.1 | 10.2 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 8.81E-05 |
| VK336 | M | Sweden | Oland | Viking | 1091 | 163 | -19.1 | 13.2 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 6.44E-05 |
| VK337 | M | Sweden | Oland | Viking | 1069 | 169 | -19.3 | 12.7 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | -0.000120921 |
| VK342 | M | Sweden | Oland | Viking | 1000 | 165 | -19.8 | 13 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | -0.000191032 |
| VK343 | M | Sweden | Oland | Viking | 1000 | 168 | -20.3 | 13.7 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 1.28E-05 |
| VK344 | M | Sweden | Oland | Viking | 1000 | 191 | -20.5 | 13.1 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 5.29E-05 |
| VK345 | M | Sweden | Oland | Viking | 1000 | 158 | -19 | 11.4 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | -0.000222555 |
| VK346 | M | Sweden | Oland | Viking | 1000 | 173 | -19.2 | 11.7 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 6.32E-05 |
| VK348 | M | Sweden | Oland | Viking | 1000 | 182 | -19.1 | 13.5 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | -2.43E-05 |
| VK350 | F | Sweden | Oland | Viking | 1134 | 159 | -19.6 | 11.9 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | -0.000137512 |
| VK352 | M | Sweden | Oland | Viking | 1000 | 176 | -19.9 | 12.6 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 0.00015486 |
| VK354 | M | Sweden | Oland | Viking | 939 | 169 | -19.7 | 13.4 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | -4.14E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | |
| VK355 | M | Sweden | Oland | Viking | 1097 | 165 | -19.4 | 13.4 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | -2.65E-05 |
| VK357 | M | Sweden | Oland | Viking | 887 | 162 | -19.5 | 12.9 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 5.34E-05 |
| VK442 | F | Sweden | Oland | Viking | 1097 | 146 | -19.1 | 11.4 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | -7.00E-05 |
| VK443 | M | Sweden | Oland | Viking | 1000 | 175 | -19.2 | 12.4 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | 6.47E-05 |
| VK444 | M | Sweden | Oland | Viking | 1097 | 181 | -18.6 | 13.9 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | (Wilhelmson, 2017) | -3.82E-07 |
| VK481 | M | Estonia | Saaremaa | Viking | 1200 | 170.6 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 0.000144795 |
| VK487 | M | Estonia | Saaremaa | Viking | 1200 | 172 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 0.000334063 |
| VK490 | M | Estonia | Saaremaa | Viking | 1227 | 174 | // | // | (Douglas Price *et al.*, 2016) | (Margaryan *et al.*, 2020) | // | 0.000124086 |
| VK492 | M | Estonia | Saaremaa | Viking | 1200 | 169.1 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 1.93E-05 |
| VK495 | M | Estonia | Saaremaa | Viking | 1200 | 170.4 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 2.35E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VK496 | M | Estonia | Saaremaa | Viking | 1200 | 179 | // | // | (Douglas Price *et al.*, 2016) | (Margaryan *et al.*, 2020) | // | 0.000301988 |
| VK498 | M | Estonia | Saaremaa | Viking | 1200 | 182 | // | // | (Douglas Price *et al.*, 2016) | (Margaryan *et al.*, 2020) | // | 7.39E-05 |
| VK504 | M | Estonia | Saaremaa | Viking | 1200 | 181.2 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 0.0001358 |
| VK505 | M | Estonia | Saaremaa | Viking | 1200 | 174.7 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 0.000370818 |
| VK506 | M | Estonia | Saaremaa | Viking | 1237 | 179.3 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 0.000209148 |
| VK522 | F | Sweden | Oland | Viking | 1550 | 163 | -20.3 | 11.2 | (Wilhelmson, 2017) | (Margaryan *et al.*, 2020) | // | 0.000170278 |
| VK550 | M | Estonia | Saaremaa | Viking | 1200 | 179.7 | // | // | (Cox *et al.*, 2022) | (Margaryan *et al.*, 2020) | // | 6.18E-05 |
| VIK_84005.SG | M | Sweden | Sigtuna, cemetery 1 | Viking | 900 | 177.762 | // | // | (Cox *et al.*, 2022) | (Krzewińska *et al.*, 2018) | // | 0.000172945 |
| ZVEJ21 | M | Latvia | Zvejnieki | Mesolithic | 7074 | 174.31 | // | // | (Marciniak *et al.*, 2021) | (Mathieson *et al.*, 2018) | // | 7.65E-06 |
| ZVEJ30 | M | Latvia | Zvejnieki | Mesolithic | 9218 | 169.95 | // | // | (Marciniak *et al.*, 2021) | (Mathieson *et al.*, 2018) | // | 4.55E-05 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ZVEJ32 | F | Latvia | Zvejnieki | Mesolithic | 9218 | 150.86 | // | // | (Marciniak *et al.*, 2021) | (Mathieson *et al.*, 2018) | // | 7.18E-06 |
| Zerniki1 | F | Poland | Żerniki Wielkie | Copper Age | 4138 | 156.46 | // | // | (Marciniak *et al.*, 2021) | (Olalde *et al.*, 2018) | // | 0.000260244 |
| grt035 | M | Sweden | Sigtuna, cemetery 1 | Viking | -1000 | 166 | -21.82 | 10.84 | (Cahill and Sikora, 2011) | (Krzewińska *et al.*, 2018) | (Kjellström *et al.*, 2009) | 0.000266699 |
| stg021 | F | Sweden | Sigtuna, cemetery 1 | Viking | -900 | 176 | -20.31 | 12.15 | (Kjellström, 2005) | (Krzewińska *et al.*, 2018) | (Kjellström *et al.*, 2009) | 6.31E-05 |
| VIK_84001 | M | Sweden | Sigtuna, cemetery 1 | Viking | 964 | 176 | -21.05 | 10.87 | (Cox *et al.*, 2022) | (Krzewińska *et al.*, 2018) | (Cox *et al.*, 2022) | -1.98E-05 |

***Table 4.S1:*** *Description of the samples used for the stature analysis.*

# 5 Conclusion

This work aimed to research the potential of applying the latest wet lab and in silico lab methods for studying ancient genomics samples. First I have summarised the current achievement and challenge of analysing DNA from ancient specimens. Then I have described how some of these challenges can be addressed by means of novel wet-lab and in-silico laboratory methods. For example, in the first chapter I have described how we managed to extract and analyse DNA from ancient genomic samples coming from Malta, one of the most southern and hotter countries in southern Europe. This analysis allowed me to report interesting findings about the genetic structure, admixture and kinship of the ancient Maltese population. For example I have shown how ancient Neolithic Maltese resembled other South Eastern European populations that lived in the same period. Moreover, the low quantity of WHG present in late Neolithic Maltese gave me a first hint of a genetic isolation that this population might have experienced. Once I proved the possibility of extracting DNA from ancient specimens in Malta I believe that increasing the amount of samples sequenced from this place between Early Neolithic and Bronze Age period would help to highlight the shift in ancestry that this population experienced.

To expand the results from the first chapter, in chapter two I have then applied a genotype imputation method to increase the amount of data available from my Maltese samples. As expected all the Maltese showed clear signals of restricted population size that shrunk just a few generations before since the samples lived in agreement with archaeological findings. To contextualise these results I then extended the imputation analysis to other European samples that lived in the same period. With this regard the first important finding I have obtained was to show the possibility of accurately imputing low coverage SNP capture genotypes from ancient specimens.

Using this information I have then highlighted the fine genetic structure of ancient Neolithic European populations unveiling two important findings. First I have shown that the genetic structure of ancient Neolithic populations resemble their respective modern ones and was mostly shaped by maritime communication. Secondly, I have discovered that, similarly to the Maltese, other islands experienced similar recent restricted population size, probably due to both their size and distance from the mainland coasts. These findings surprisingly challenged previous archaeological works that suggested the sea was an accelerator of migration rather than a barrier. I believe to further prove my point it would be important to expand my analysis to other Neolithic populations from islands not yet considered such as the balearic islands and other non-European archipelagos.

In the last chapter of my thesis I have considered using results from the genotype imputation analysis to investigate the polygenic traits in ancient European populations that lived between the Mesolithic and Mediaeval times. Thanks to coupling these results with published archaeological data my results showed a good agreement between the genetic and osteological predicted stature. Once shown the good quality I have then used this model to unveil trends in stature between different periods and cultures. This analysis highlighted both Viking and Mediaeval populations to show unusual and opposite trends in their stature compared to other parts of Europe. Moreover the regression model that I have built allowed me to detect a trend, although not significant, between the protein intake of a person and his final stature. Overall, the analyses shown in this chapter highlighted the possibility of predicting with good accuracy polygenic traits in ancient European populations. However the scarcity of archaeological measures, such as the stature and isotope information still pose a challenge in understanding the influence of the environment in the phenotype of an individual. With this regard I believe more effort should be put in coupling DNA with anthropometric and isotope analysis in order to have a clear picture of the physical status of a population.

# 5. References


1000 Genomes Project Consortium *et al.* (2015) 'A global reference for human genetic variation', *Nature*, 526(7571), pp. 68–74.

Allentoft, M.E. *et al.* (2015) 'Population genomics of Bronze Age Eurasia', *Nature*, 522(7555), pp. 167–172.

Alt, K.W. *et al.* (2014) 'Lombards on the Move – An Integrative Study of the Migration Period Cemetery at Szólád, Hungary', *PLoS ONE*, p. e110793. Available at: https://doi.org/10.1371/journal.pone.0110793.

Amorim, C.E.G. *et al.* (2018) 'Understanding 6th-century barbarian social organization and migration through paleogenomics', *Nature communications*, 9(1), p. 3547.

Antonio, M.L. *et al.* (2019) 'Ancient Rome: A genetic crossroads of Europe and the Mediterranean', *Science*, 366(6466), pp. 708–714.

Ariano, B. *et al.* (2022) 'Ancient Maltese genomes and the genetic geography of Neolithic Europe', *Current biology: CB*, 32(12), pp. 2668–2680.e6.

Avila-Arcos, M.C. *et al.* (2011) 'Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA', *Scientific reports*, 1, p. 74.

de Barros Damgaard, P. *et al.* (2018) 'The first horse herders and the impact of early Bronze Age steppe expansions into Asia', *Science* [Preprint]. Available at: https://doi.org/10.1126/science.aar7711.

Belbin, G.M. *et al.* (2017) 'Genetic identification of a common collagen disease in puerto ricans via identity-by-descent mapping in a health system', *eLife*, 6. Available at: https://doi.org/10.7554/eLife.25060.

Beyer, R.M. *et al.* (2021) 'Climatic windows for human migration out of Africa in the past 300,000 years', *Nature communications*, 12(1), p. 4889.

Bhérer, C., Campbell, C.L. and Auton, A. (2017) 'Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales', *Nature Communications*. Available at: https://doi.org/10.1038/ncomms14994.

Bocquet-Appel, J.-P. *et al.* (2012) 'Understanding the rates of expansion of the farming system in Europe', *Journal of archaeological science*, 39(2), pp. 531–546.

Bodmer, W. (2015) 'Genetic Characterization of Human Populations: From ABO to a Genetic Map of the British People', *Genetics*, 199(2), pp. 267–279.

Boessenkool, S. *et al.* (2017) 'Combining bleach and mild predigestion improves ancient DNA recovery from bones', *Molecular ecology resources*, 17(4), pp. 742–751.

Brace, S. *et al.* (2019) 'Ancient genomes indicate population replacement in Early Neolithic Britain', *Nature ecology & evolution* [Preprint]. Available at: https://doi.org/10.1038/s41559-019-0871-9.

Bradbury, I.R. *et al.* (2011) 'Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, Gadus morhua', *Molecular ecology resources*, 11 Suppl 1, pp. 218–225.

Briggs, A.W. *et al.* (2007) 'Patterns of damage in genomic DNA sequences from a Neandertal', *Proceedings of the National Academy of Sciences of the United States of America*, 104(37), pp. 14616–14621.

Broushaki, F. *et al.* (2016) 'Early Neolithic genomes from the eastern Fertile Crescent', *Science*, 353(6298), pp. 499–503.

Browning, B.L. and Browning, S.R. (2013) 'Detecting identity by descent and estimating genotype error rates in sequence data', *American journal of human genetics*, 93(5), pp. 840–851.

Browning, B.L. and Browning, S.R. (2016) 'Genotype Imputation with Millions of Reference Samples', *American journal of human genetics*, 98(1), pp. 116–126.

Browning, S.R. (2006) 'Multilocus association mapping using variable-length Markov chains', *American journal of human genetics*, 78(6), pp. 903–913.

Browning, S.R. (2008) 'Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes', *Genetics*, 178(4), pp. 2123–2132.

Browning, S.R. and Browning, B.L. (2007) 'Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering', *American journal of human genetics*, 81(5), pp. 1084–1097.

Browning, S.R. and Browning, B.L. (2010) 'High-resolution detection of identity by descent in unrelated individuals', *American journal of human genetics*, 86(4), pp. 526–539.

Browning, S.R. and Browning, B.L. (2015) 'Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent', *American journal of human genetics*, 97(3), pp. 404–418.

Burger, J. *et al.* (2020) 'Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 3,000 Years', *Current biology: CB*, 30(21), pp. 4307–4315.e13.

Caballero, M. *et al.* (2019) 'Crossover interference and sex-specific genetic maps shape identical by descent sharing in close relatives', *PLoS genetics*, 15(12), p.

e1007979.

Cahill, M. and Sikora, M. (2011) *Breaking Ground, Finding Graves: Reports on the Excavations of Burials by the National Museum of Ireland, 1927-2006*. Wordwell.

Campbell, C.L. *et al.* (2015) 'Escape from crossover interference increases with maternal age', *Nature communications*, 6, p. 6260.

Cann, R.L. (1988) 'DNA and Human Origins', *Annual review of anthropology*, 17(1), pp. 127–143.

Cassidy, L.M. *et al.* (2016) 'Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome', *Proceedings of the National Academy of Sciences of the United States of America*, 113(2), pp. 368–373.

Cassidy, L.M. *et al.* (2020) 'A dynastic elite in monumental Neolithic society', *Nature*, 582(7812), pp. 384–388.

Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1994) *The history and geography of human genes*. Princeton university press.

Chang, C.C. *et al.* (2015) 'Second-generation PLINK: rising to the challenge of larger and richer datasets', *GigaScience*, 4, p. 7.

Cox, S.L. *et al.* (2019) 'Genetic contributions to variation in human stature in prehistoric Europe', *Proceedings of the National Academy of Sciences of the United States of America*, 116(43), pp. 21484–21492.

Cox, S.L. *et al.* (2022) 'Predicting skeletal stature using ancient DNA', *American Journal of Biological Anthropology*, 177(1), pp. 162–174.

Cramp, L.J.E. *et al.* (2014) 'Neolithic dairy farming at the extreme of agriculture in northern Europe', *Proceedings. Biological sciences / The Royal Society*, 281(1791), p. 20140819.

Cunliffe, B. (2018) *The Ancient Celts, Second Edition*. Oxford University Press.

Cunliffe, S.B. (2017) *On the Ocean: The Mediterranean and the Atlantic from prehistory to AD 1500*. Oxford University Press.

D'Anglade, A.G. and Gorosquieta, A.V. (2017) 'Caracterización isotópica de Elba, la mujer mesolítica de Chan do Lindeiro (Pedrafita, Lugo, Península Ibérica)', *Cadernos do Laboratorio Xeolóxico de Laxe. Revista de Xeoloxía Galega e do Hercínico Peninsular*, pp. 89–109. Available at: https://doi.org/10.17979/cadlaxe.2017.39.0.3549.

Das, S., Abecasis, G.R. and Browning, B.L. (2018) 'Genotype Imputation from Large Reference Panels', *Annual review of genomics and human genetics*, 19, pp. 73–96.

Davis, D.L. (2001) *Navigation in the ancient eastern Mediterranean*. Texas A&M

University. Available at: https://oaktrust.library.tamu.edu/handle/1969.1/ETD-TAMU-2001-THESIS-D384 (Accessed: 2 March 2022).

Delaneau, O., Coulonges, C. and Zagury, J.-F. (2008) 'Shape-IT: new rapid and accurate algorithm for haplotype inference', *BMC Bioinformatics*, p. 540. Available at: https://doi.org/10.1186/1471-2105-9-540.

Delaneau, O., Marchini, J. and Zagury, J.-F. (2011) 'A linear complexity phasing method for thousands of genomes', *Nature methods*, 9(2), pp. 179–181.

Douglas Price, T. *et al.* (2016) 'Isotopic provenancing of the Salme ship burials in Pre-Viking Age Estonia', *Antiquity*, 90(352), pp. 1022–1037.

Drucker, D.G. *et al.* (2018) 'Aquatic resources in human diet in the Late Mesolithic in Northern France and Luxembourg: insights from carbon, nitrogen and sulphur isotope ratios', *Archaeological and anthropological sciences*, 10(2), pp. 351–368.

Dubova, N.A. and Rykushina, G.V. (2007) 'New data on anthropology of the necropolis of Gonur-depe', in *Necropolis of Gonur*, pp. 296–329.

Dulias, K. *et al.* (no date) 'Ancient DNA at the edge of the world: Continental immigration and the persistence of Neolithic male lineages in Bronze Age Orkney', *Proceedings of the National Academy of Sciences, USA* [Preprint].

Ehler, E. *et al.* (2019) 'AmtDB: a database of ancient human mitochondrial genomes', *Nucleic acids research*, 47(D1), pp. D29–D32.

Falys, C. (no date) 'St john's College'. Oxford (St Brice's day massacre.

Fernandes, D.M. *et al.* (2020) 'The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean', *Nature Ecology & Evolution*, 4(3), pp. 334–345.

Fernandes, D.M. *et al.* (2021) 'A genetic history of the pre-contact Caribbean', *Nature*, 590(7844), pp. 103–110.

Fidalgo, D.F.F. (2014) *Contextos funerários e estudo antropológico dos restos ósseos humanos dos hipogeus de Torre Velha 3 (São Salvador, Serpa): Uma aproximação ao estudo das comunidades humanas do Bronze do Sudoeste*. Available at: https://eg.uc.pt/handle/10316/27862.

Fregel, R. *et al.* (2018) 'Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe', *Proceedings of the National Academy of Sciences of the United States of America*, 115(26), pp. 6774–6779.

French, C. *et al.* (2020) *Temple landscapes: Fragility, change and resilience of Holocene environments in the Maltese Islands*. Cambridge: McDonald Institute for Archaeological Research, Cambridge.

Fully, G. (1956) ‘Une nouvelle méthode de détermination de la taille’, *Annales de medecine legale, criminologie, police scientifique et toxicologie*, 36, pp. 266–273.

Fu, Q. *et al.* (2016) ‘The genetic history of Ice Age Europe’, *Nature*, 534(7606), pp. 200–205.

Gaastra, J.S., de Vareilles, A. and Vander Linden, M. (2022) ‘Bones and Seeds: An Integrated Approach to Understanding the Spread of Farming across the Western Balkans’, *Environmental Archaeology*, 27(1), pp. 44–60.

Gamba, C. *et al.* (2014) ‘Genome flux and stasis in a five millennium transect of European prehistory’, *Nature communications*, 5, p. 5257.

Gazal, S. *et al.* (2014) ‘Inbreeding coefficient estimation with dense SNP data: comparison of strategies and application to HapMap III’, *Human heredity*, 77(1-4), pp. 49–62.

George, V.T. and Elston, R.C. (1987) ‘Testing the association between polymorphic markers and quantitative traits in pedigrees’, *Genetic epidemiology*, 4(3), pp. 193–201.

Gilbert, M.T.P. *et al.* (2004) ‘Ancient mitochondrial DNA from hair’, *Current biology: CB*, 14(12), pp. R463–4.

González-Fortes, G. *et al.* (2017) ‘Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin’, *Current biology: CB*, 27(12), pp. 1801–1810.e10.

González-Fortes, G. *et al.* (2019) ‘A western route of prehistoric human migration from Africa into the Iberian Peninsula’, *Proceedings. Biological sciences / The Royal Society*, 286(1895), p. 20182288.

Green, R.E. *et al.* (2010) ‘A draft sequence of the Neandertal genome’, *Science*, 328(5979), pp. 710–722.

Guan, Y. (2014) ‘Detecting Structure of Haplotypes and Local Ancestry’, *Genetics*, 196(3), pp. 625–642.

Guilaine, J. (2017) ‘The Neolithic Transition: From the Eastern to the Western Mediterranean’, in O. García-Puchol and D.C. Salazar-García (eds) *Times of Neolithic Transition along the Western Mediterranean*. Cham: Springer International Publishing, pp. 15–31.

Guilaine, J. (2018) ‘A personal view of the neolithisation of the Western Mediterranean’, *Quaternary international: the journal of the International Union for Quaternary Research*, 470, pp. 211–225.

Günther, T. *et al.* (2015) ‘Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques’, *Proceedings of the National Academy of Sciences of the United States of America*, 112(38), pp. 11917–11922.

Günther, T. *et al.* (2018) ‘Population genomics of Mesolithic Scandinavia:

Investigating early postglacial migration routes and high-latitude adaptation', *PLoS biology*, 16(1), p. e2003703.

Günther, T. and Nettelblad, C. (2019) 'The presence and impact of reference bias on population genomic studies of prehistoric human populations', *PLoS genetics*, 15(7), p. e1008302.

Haak, W. *et al.* (2015) 'Massive migration from the steppe was a source for Indo-European languages in Europe', *Nature*, 522(7555), pp. 207–211.

Hagelberg, E., Sykes, B. and Hedges, R. (1989) 'Ancient bone DNA amplified', *Nature*, 342(6249), p. 485.

Harney, É. *et al.* (2018) 'Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation', *Nature communications*, 9(1), p. 3336.

Henderson, D.A. *et al.* (2014) 'Regional variations in the European Neolithic dispersal: the role of the coastlines', *Antiquity*, 88(342), pp. 1291–1302.

Henn, B.M. *et al.* (2012) 'Genomic ancestry of North Africans supports back-to-Africa migrations', *PLoS genetics*, 8(1), p. e1002397.

Higuchi, R. *et al.* (1984) 'DNA sequences from the quagga, an extinct member of the horse family', *Nature*, 312(5991), pp. 282–284.

Hodder, I. (2010) *Religion in the Emergence of Civilization: Çatalhöyük as a Case Study*. Cambridge University Press.

Hofmanová, Z. *et al.* (2016) 'Early farmers from across Europe directly descended from Neolithic Aegeans', *Proceedings of the National Academy of Sciences of the United States of America*, 113(25), pp. 6886–6891.

Hommel, G. (1988) 'A stagewise rejective multiple test procedure based on a modified Bonferroni test', *Biometrika*, 75(2), pp. 383–386.

Howie, B., Marchini, J. and Stephens, M. (2011) 'Genotype imputation with thousands of genomes', *G3* , 1(6), pp. 457–470.

Howie, B.N., Donnelly, P. and Marchini, J. (2009) 'A flexible and accurate genotype imputation method for the next generation of genome-wide association studies', *PLoS genetics*, 5(6), p. e1000529.

Hui, R. *et al.* (2020) 'Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes', *Scientific reports*, 10(1), p. 18542.

International HapMap Consortium (2003) 'The International HapMap Project', *Nature*, 426(6968), pp. 789–796.

Isern, N. *et al.* (2017) 'Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean', *Proceedings of the National Academy of Sciences of the United States of America*, 114(5), pp. 897–902.

Jessica Smyth and Richard P. Evershed (2015) ‘The molecules of meals: new insight into Neolithic foodways’, *Proceedings of the Royal Irish Academy: Archaeology, Culture, History, Literature*, 115C, pp. 27–46.

Johnson, R.J., Lanaspa, M.A. and Fox, J.W. (2021) ‘Upper Paleolithic Figurines Showing Women with Obesity may Represent Survival Symbols of Climatic Change’, *Obesity* , 29(1), pp. 11–15.

Jones, E.R. *et al.* (2015) ‘Upper Palaeolithic genomes reveal deep roots of modern Eurasians’, *Nature communications*, 6, p. 8912.

Jones, E.R. *et al.* (2017) ‘The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers’, *Current biology: CB*, 27(4), pp. 576–582.

Jónsson, H. *et al.* (2013) ‘mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters’, *Bioinformatics* , 29(13), pp. 1682–1684.

Jørkov, M.L.S., Jørgensen, L. and Lynnerup, N. (2010) ‘Uniform diet in a diverse society. Revealing new dietary evidence of the Danish Roman Iron Age based on stable isotope analysis’, *American journal of physical anthropology*, 143(4), pp. 523–533.

Keavney, B. *et al.* (1998) ‘Measured haplotype analysis of the angiotensin-I converting enzyme gene’, *Human molecular genetics*, 7(11), pp. 1745–1751.

Keller, A. *et al.* (2012) ‘New insights into the Tyrolean Iceman’s origin and phenotype as inferred by whole-genome sequencing’, *Nature communications*, 3, p. 698.

Kılınç, G.M. *et al.* (2016) ‘The Demographic Development of the First Farmers in Anatolia’, *Current biology: CB*, 26(19), pp. 2659–2666.

Kjellström, A. (2005) *The Urban Farmer : Osteoarchaeological Analysis of Skeletons from Medieval Sigtuna Interpreted in a Socioeconomic Perspective*. Department of Archaeology and Classical Studies, Stockholm University. Available at: https://www.diva-portal.org/smash/record.jsf?pid=diva2:195951 (Accessed: 16 March 2022).

Kjellström, A. *et al.* (2009) ‘Dietary patterns and social structures in medieval Sigtuna, Sweden, as reflected in stable isotope values in human skeletal remains’, *Journal of archaeological science*, 36(12), pp. 2689–2699.

Kotlyakov, V.M., Velichko, A.A. and Vasil’ev, S.A. (2017) *Human Colonization of the Arctic: The Interaction Between Early Migration and the Paleoenvironment*. Academic Press.

Krause, J. *et al.* (2010) ‘A complete mtDNA genome of an early modern human from Kostenki, Russia’, *Current biology: CB*, 20(3), pp. 231–236.

Kriiska, A., Lavento, M. and Peets, J. (2005) ‘New ams dates of the neolithic and bronze age ceramics in Estonia: Preliminary results and interpretations’, *Estonian*

*Journal of Archaeology*, 9(1), p. 3.

Krzewińska, M. *et al.* (2018) 'Genomic and Strontium Isotope Variation Reveal Immigration Patterns in a Viking Age Town', *Current biology: CB*, 28(17), pp. 2730–2738.e10.

Lachance, J. and Tishkoff, S.A. (2013) 'SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it', *BioEssays: news and reviews in molecular, cellular and developmental biology*, 35(9), pp. 780–786.

Lambeck, K. *et al.* (2011) 'Sea level and shoreline reconstructions for the Red Sea: isostatic and tectonic considerations and implications for hominin migration out of Africa', *Quaternary science reviews*, 30(25), pp. 3542–3574.

Latham, K.J. (2013) 'Human Health and the Neolithic Revolution: an Overview of Impacts of the Agricultural Transition on Oral Health, Epidemiology, and the Human Body'. Available at: https://digitalcommons.unl.edu/nebanthro/187/ (Accessed: 2 March 2022).

Lawson, D.J. *et al.* (2012) 'Inference of population structure using dense haplotype data', *PLoS genetics*, 8(1), p. e1002453.

Lazaridis, I. *et al.* (2014) 'Ancient human genomes suggest three ancestral populations for present-day Europeans', *Nature*, 513(7518), pp. 409–413.

Lazaridis, I. *et al.* (2016) 'Genomic insights into the origin of farming in the ancient Near East', *Nature*, 536(7617), pp. 419–424.

Lazaridis, I. (2018) 'The evolutionary history of human populations in Europe', *Current opinion in genetics & development*, 53, pp. 21–27.

Leslie, S. *et al.* (2015) 'The fine-scale genetic structure of the British population', *Nature*, 519(7543), pp. 309–314.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics* , 25(16), pp. 2078–2079.

Li, H. (2011) 'A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data', *Bioinformatics* , 27(21), pp. 2987–2993.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics* , 25(14), pp. 1754–1760.

Lindahl, T. (1993) 'Instability and decay of the primary structure of DNA', *Nature*, 362(6422), pp. 709–715.

Lipatov, M. *et al.* (2015) 'Maximum Likelihood Estimation of Biological Relatedness from Low Coverage Sequencing Data', *bioRxiv*. Available at: https://doi.org/10.1101/023374.

Lipson, M. *et al.* (2017) 'Parallel palaeogenomic transects reveal complex genetic

history of early European farmers', *Nature*, 551(7680), pp. 368–372.

LI and Y (2006) 'Mach 1.0 : rapid haplotype reconstruction and missing genotype inference', *American journal of human genetics*, 79, p. 2290.

Li, Y. *et al.* (2009) 'Genotype imputation', *Annual review of genomics and human genetics*, 10, pp. 387–406.

López, S., van Dorp, L. and Hellenthal, G. (2015) 'Human Dispersal Out of Africa: A Lasting Debate', *Evolutionary bioinformatics online*, 11(Suppl 2), pp. 57–68.

MacHugh, D.E. *et al.* (2000) 'The extraction and analysis of ancient DNA from bone and teeth: a survey of current methodologies', *Ancient biomolecules*, 3(2), pp. 81–103.

Macko, S.A. *et al.* (1999) 'The Ice Man's diet as reflected by the stable nitrogen and carbon isotopic composition of his hair', *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 13(3), pp. 559–562.

Mallick, S. *et al.* (2016) 'The Simons Genome Diversity Project: 300 genomes from 142 diverse populations', *Nature*, 538(7624), pp. 201–206.

Malone, C. *et al.* (1995) 'Mortuary Ritual of 4th Millennium bc Malta: the Zebbug Period Chambered Tomb from the Brochtorff Circle at Xaghra (Gozo)', *Proceedings of the Prehistoric Society*, 61, pp. 303–345.

Malone, C. *et al.* (2018) 'The social implications of death in prehistoric Malta'. Apollo - University of Cambridge Repository. Available at: https://doi.org/10.17863/CAM.15866.

Malone, C. *et al.* (2020) *Temple places: Excavating cultural sustainability in prehistoric Malta*. [object Object]: Apollo - University of Cambridge Repository.

Malone, C., Cutajar, N. and McLaughlin, T.R. (2019) 'Island questions: the chronology of the Brochtorff Circle at Xagħra, Gozo, and its significance for the Neolithic sequence on Malta', *Archaeological and anthropological sciences* [Preprint]. Available at: https://link.springer.com/article/10.1007/s12520-019-00790-y.

Malone, C. and Stoddart, S. (2009) *Mortuary Customs in Prehistoric Malta: Excavations at the Brochtorff Circle at Xagħra (1987-94)*. McDonald Institute for Archeological Research.

Malone, C. and Stoddart, S. (2016) 'Figurines of Malta', in *The Oxford Handbook of Prehistoric Figurines*.

Manen, C. *et al.* (2019) 'The Neolithic Transition in the Western Mediterranean: a Complex and Non-Linear Diffusion Process—The Radiocarbon Record Revisited', *Radiocarbon*, 61(2), pp. 531–571.

Manichaikul, A. *et al.* (2010) 'Robust relationship inference in genome-wide association studies', *Bioinformatics* , 26(22), pp. 2867–2873.

Marchini, J. *et al.* (2007) 'A new multipoint method for genome-wide association studies by imputation of genotypes', *Nature genetics*, 39(7), pp. 906–913.

Marciniak, S. *et al.* (2021) 'An integrative skeletal and paleogenomic analysis of prehistoric stature variation suggests relatively reduced health for early European farmers', *bioRxiv*. Available at: https://doi.org/10.1101/2021.03.31.437881.

Margaryan, A. *et al.* (2020) 'Population genomics of the Viking world', *Nature*, 585(7825), pp. 390–396.

Marnetto, D. *et al.* (2020) 'Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals', *Nature communications*, 11(1), p. 1628.

Martin, A.R. *et al.* (2019) 'Clinical use of current polygenic risk scores may exacerbate health disparities', *Nature genetics*, 51(4), pp. 584–591.

Martiniano, R. *et al.* (2016) 'Genomic signals of migration and continuity in Britain before the Anglo-Saxons', *Nature communications*, 7, p. 10326.

Martiniano, R. *et al.* (2017) 'The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods', *PLoS genetics*, 13(7), p. e1006852.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), pp. 10–12.

Mathieson, I. *et al.* (2015) 'Genome-wide patterns of selection in 230 ancient Eurasians', *Nature*, 528(7583), pp. 499–503.

Mathieson, I. *et al.* (2018) 'The genomic history of southeastern Europe', *Nature*, 555(7695), pp. 197–203.

McKenna, A. *et al.* (2010) 'The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data', *Genome research*, 20(9), pp. 1297–1303.

McLaughlin, R. *et al.* (2020) 'Dating Maltese prehistory', in. McDonald Institute for Archaeological Research.

McLaughlin, T.R., Stoddart, S. and Malone, C. (2018) 'Island risks and the resilience of a prehistoric civilization', *World archaeology*, pp. 1–14.

McQuillan, R. *et al.* (2008) 'Runs of homozygosity in European populations', *American journal of human genetics*, 83(3), pp. 359–372.

Mellaart, J. (1967) 'Çatal Hüyük: a neolithic town in Anatolia'. Available at: https://ixtheo.de/Record/1146419236 (Accessed: 2 March 2022).

Meller, C. *et al.* (2009) ‘Prevalence of oral pathologic findings in an ancient pre-Columbian archeological site in the Atacama Desert’, *Oral diseases*, 15(4), pp. 287–294.

Menozzi, P., Piazza, A. and Cavalli-Sforza, L. (1978) ‘Synthetic maps of human gene frequencies in Europeans’, *Science*, 201(4358), pp. 786–792.

Meyer, M. and Kircher, M. (2010) ‘Illumina sequencing library preparation for highly multiplexed target capture and sequencing’, *Cold Spring Harbor protocols*, 2010(6), p. db.prot5448.

Money, D. *et al.* (2015) ‘LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms’, *G3* , 5(11), pp. 2383–2390.

Mooney, J.A. *et al.* (2018) ‘Understanding the Hidden Complexity of Latin American Population Isolates’, *American journal of human genetics*, 103(5), pp. 707–726.

Mourant, A.E. (1947) ‘The blood groups of the Basques’, *Nature*, 160(4067), p. 505.

Narasimhan, V.M. *et al.* (2019) ‘The formation of human populations in South and Central Asia’, *Science*, 365(6457). Available at: https://doi.org/10.1126/science.aat7487.

Nasab, H.V. and Kazzazi, M. (2011) ‘Metric Analysis of a Female Figurine from Tepe Sarab’, *Iran*, 49(1), pp. 1–10.

Novembre, J. *et al.* (2008) ‘Genes mirror geography within Europe’, *Nature*, 456(7218), pp. 98–101.

Nurk, S. *et al.* (2021) ‘The complete sequence of a human genome’, *bioRxiv*. Available at: https://doi.org/10.1101/2021.05.26.445798.

Okonechnikov, K., Conesa, A. and García-Alcalde, F. (2016) ‘Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data’, *Bioinformatics* , 32(2), pp. 292–294.

Olalde, I. *et al.* (2014) ‘Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European’, *Nature*, 507(7491), pp. 225–228.

Olalde, I. *et al.* (2015) ‘A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures’, *Molecular biology and evolution*, 32(12), pp. 3132–3142.

Olalde, I. *et al.* (2018) ‘The Beaker phenomenon and the genomic transformation of northwest Europe’, *Nature*, 555(7695), pp. 190–196.

Olalde, I. *et al.* (2019) ‘The genomic history of the Iberian Peninsula over the past 8000 years’, *Science*, 363(6432), pp. 1230–1234.

Orlando, L. *et al.* (2021) ‘Ancient DNA analysis’, *Nature Reviews Methods*

*Primers*, 1(1), pp. 1–26.

Pääbo, S. (1985) 'Molecular cloning of Ancient Egyptian mummy DNA', *Nature*, 314(6012), pp. 644–645.

Pääbo, S. (1989) 'Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification', *Proceedings of the National Academy of Sciences of the United States of America*, 86(6), pp. 1939–1943.

Patterson, N. *et al.* (2012) 'Ancient admixture in human history', *Genetics*, 192(3), pp. 1065–1093.

Patterson, N., Price, A.L. and Reich, D. (2006) 'Population structure and eigenanalysis', *PLoS genetics*, 2(12), p. e190.

Petkova, D., Novembre, J. and Stephens, M. (2016) 'Visualizing spatial population structure with estimated effective migration surfaces', *Nature genetics*, 48(1), pp. 94–100.

Posth, C. *et al.* (2016) 'Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe', *Current biology: CB*, 26(6), pp. 827–833.

Price, A.L. *et al.* (2006) 'Principal components analysis corrects for stratification in genome-wide association studies', *Nature genetics*, 38(8), pp. 904–909.

Price, A.L. *et al.* (2009) 'Sensitive detection of chromosomal segments of distinct ancestry in admixed populations', *PLoS genetics*, 5(6), p. e1000519.

Raghavan, M. *et al.* (2014) 'Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans', *Nature*, 505(7481), pp. 87–91.

Rainbird, P. (2007) *The Archaeology of Islands*. Cambridge University Press.

Ralf, A. *et al.* (2018) 'Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data', *Molecular biology and evolution*, 35(5), pp. 1291–1294.

Rasmussen, S. *et al.* (2015) 'Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago', *Cell*, pp. 571–582. Available at: https://doi.org/10.1016/j.cell.2015.10.009.

Ratican, C. (2020) 'The other body: Persons in Viking Age multiple burials in Scandinavia and the Western Diaspora'. Apollo - University of Cambridge Repository. Available at: https://doi.org/10.17863/CAM.54718.

Raxter, M.H., Auerbach, B.M. and Ruff, C.B. (2006) 'Revision of the Fully technique for estimating statures', *American journal of physical anthropology*, 130(3), pp. 374–384.

Renaud, G. *et al.* (2019) 'Joint Estimates of Heterozygosity and Runs of Homozygosity for Modern and Ancient Samples', *Genetics*, 212(3), pp. 587–614.

Richards, M.P. *et al.* (2001) 'Neolithic diet at the Brochtorff Circle, Malta', *European Journal of Archaeology*, 4(2), pp. 253–262.

Richards, M.P., Schulting, R.J. and Hedges, R.E.M. (2003) 'Archaeology: sharp shift in diet at onset of Neolithic', *Nature*, 425(6956), p. 366.

Ringbauer, H., Novembre, J. and Steinrücken, M. (2021) 'Parental relatedness through time revealed by runs of homozygosity in ancient DNA', *Nature communications*, 12(1), p. 5425.

Rivollat, M. *et al.* (2020) 'Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers', *Science advances*, 6(22), p. eaaz5344.

Ruff, C.B. *et al.* (2012) 'Stature and body mass estimation from skeletal remains in the European Holocene', *American journal of physical anthropology*, 148(4), pp. 601–617.

Ruff, C.B. (2018) *Skeletal Variation and Adaptation in Europeans: Upper Paleolithic to the Twentieth Century*. John Wiley & Sons.

Saag, L. *et al.* (2017) 'Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe', *Current biology: CB*, 27(14), pp. 2185–2193.e6.

Sánchez-Quinto, F. *et al.* (2019) 'Megalithic tombs in western and northern Neolithic Europe were linked to a kindred society', *Proceedings of the National Academy of Sciences of the United States of America*, 116(19), pp. 9469–9474.

Scheib, C.L. *et al.* (2019) 'East Anglian early Neolithic monument burial linked to contemporary Megaliths', *Annals of human biology*, 46(2), pp. 145–149.

Schiffels, S. *et al.* (2016) 'Iron Age and Anglo-Saxon genomes from East England reveal British migration history', *Nature communications*, 7, p. 10408.

Schroeder, H. *et al.* (2019) 'Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave', *Proceedings of the National Academy of Sciences of the United States of America*, 116(22), pp. 10705–10710.

Schubert, M., Lindgreen, S. and Orlando, L. (2016) 'AdapterRemoval v2: rapid adapter trimming, identification, and read merging', *BMC research notes*, 9, p. 88.

Schwarze, K. *et al.* (2020) 'The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom', *Genetics in medicine: official journal of the American College of Medical Genetics*, 22(1), pp. 85–94.

Seguin-Orlando, A. *et al.* (2014) 'Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years', *Science*, 346(6213), pp. 1113–1118.

Seguin-Orlando, A. *et al.* (2021) 'Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France', *Current biology: CB*, 31(5), pp. 1072–1083.e10.

Sellevold, B.J. and Skar, B. (1994) ‘The first lady of Norway’, *NIKU*, 1999, pp. 6–11.

Sheridan, A. (2010) ‘The Neolithization of Britain and Ireland: the “big picture”’, *Landscapes in transition*, pp. 89–105.

Shi, S. *et al.* (2018) ‘Comprehensive Assessment of Genotype Imputation Performance’, *Human heredity*, 83(3), pp. 107–116.

Sikora, M. *et al.* (2014) ‘Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe’, *PLoS genetics*, 10(5), p. e1004353.

Sikora, M. *et al.* (2017) ‘Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers’, *Science*, 358(6363), pp. 659–662.

Sirugo, G., Williams, S.M. and Tishkoff, S.A. (2019) ‘The Missing Diversity in Human Genetic Studies’, *Cell*, 177(4), p. 1080.

Skoglund, P. *et al.* (2012) ‘Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe’, *Science*, 336(6080), pp. 466–469.

Skoglund, P. *et al.* (2013) ‘Accurate sex identification of ancient human remains using DNA shotgun sequencing’, *Journal of archaeological science*, 40(12), pp. 4477–4482.

Skoglund, P., Malmström, H., *et al.* (2014) ‘Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers’, *Science*, 344(6185), pp. 747–750.

Skoglund, P., Northoff, B.H., *et al.* (2014) ‘Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal’, *Proceedings of the National Academy of Sciences of the United States of America*, 111(6), pp. 2229–2234.

Soares, P. *et al.* (2011) ‘The Expansion of mtDNA Haplogroup L3 within and out of Africa’, *Molecular biology and evolution*, 29(3), pp. 915–927.

Sohail, M. *et al.* (2019) ‘Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies’, *eLife*, 8. Available at: https://doi.org/10.7554/eLife.39702.

St. George Gray, H. (1908) *Report on the Excavations at Wick Barrow, Stogursey, Somersetshire*. Taunton Castle.

Stoddart, S. *et al.* (2009) ‘The Tarxien phase levels: spatial and stratigraphic analysis and reconstruction’. Available at: https://www.um.edu.mt/library/oar/handle/123456789/47480.

Stringer, C.B. and Andrews, P. (1988) ‘Genetic and Fossil Evidence for the Origin of Modern Humans’, *Science*, pp. 1263–1268. Available at: https://doi.org/10.1126/science.3125610.

Sul, J.H., Martin, L.S. and Eskin, E. (2018) 'Population structure in genetic studies: Confounding factors and mixed models', *PLoS genetics*, 14(12), p. e1007309.

Teasdale, M.D. *et al.* (2015) 'Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1660), p. 20130379.

Thompson, E.A. (2013) 'Identity by descent: variation in meiosis, across genomes, and in populations', *Genetics*, 194(2), pp. 301–326.

Thompson, J.E. *et al.* (2020) 'Placing and remembering the dead in late Neolithic Malta: bioarchaeological and spatial analysis of the Xagħra Circle Hypogeum, Gozo', *World archaeology*, 52(1), pp. 71–89.

Torroni, A. *et al.* (1998) 'mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe', *American journal of human genetics*, 62(5), pp. 1137–1152.

Tykot, R.H., Freund, K.P. and Vianello, A. (2013) 'Source Analysis of Prehistoric Obsidian Artifacts in Sicily (Italy) Using pXRF', in *Archaeological Chemistry VIII*. American Chemical Society (ACS Symposium Series), pp. 195–210.

Valdiosera, C. *et al.* (2018) 'Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia', *Proceedings of the National Academy of Sciences of the United States of America*, 115(13), pp. 3428–3433.

van der Valk, T. *et al.* (2021) 'Million-year-old DNA sheds light on the genomic history of mammoths', *Nature*, 591(7849), pp. 265–269.

de Vareilles, A. *et al.* (2020) 'One sea but many routes to Sail. The early maritime dispersal of Neolithic crops from the Aegean to the western Mediterranean', *Journal of Archaeological Science: Reports*, 29, p. 102140.

Venter, J.C. *et al.* (2001) 'The sequence of the human genome', *Science*, 291(5507), pp. 1304–1351.

Vercellotti, G. *et al.* (2009) 'Stature estimation in an early medieval (XI-XII c.) Polish population: testing the accuracy of regression equations in a bioarcheological sample', *American journal of physical anthropology*, 140(1), pp. 135–142.

Vianello, D. *et al.* (2013) 'HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment', *Human mutation*, 34(9), pp. 1189–1194.

Vidale, M., Micheli, R. and Olivieri, L.M. (2016) *Excavations at the protohistoric graveyards of Gogdara and Udegram*. Sang-e-Meel Publications.

Villalba-Mouco, V. *et al.* (2019) 'Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula', *Current biology: CB*, 29(7), pp. 1169–1177.e7.

Waddington, C. and Bonsall, C. (2016) *Archaeology and environment on the North Sea littoral: a case study from Low Hauxley*. Archaeological Research Services Ltd.

Walter, R.C. *et al.* (2000) 'Early human occupation of the Red Sea coast of Eritrea during the last interglacial', *Nature*, 405(6782), pp. 65–69.

Wilhelmson, H. (2017) *Perspectives from a human-centred archaeology : Iron Age people and society on Öland*. Lund University. Available at: https://lup.lub.lu.se/record/c6d33164-dd18-466b-9936-44156f3e3ea4 (Accessed: 16 March 2022).

Yang, D.Y. *et al.* (1998) 'Technical note: improved DNA extraction from ancient bones using silica-based spin columns', *American journal of physical anthropology*, 105, pp. 539–543.

Yengo, L. *et al.* (2018) 'Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry', *Human molecular genetics*, 27(20), pp. 3641–3649.

Yengo, L., Wray, N.R. and Visscher, P.M. (2019) 'Extreme inbreeding in a European ancestry sample from the contemporary UK population', *Nature communications*, 10(1), p. 3719.

Zilhão, J. (2014) 'Early prehistoric navigation in the Western Mediterranean: implications for the Neolithic transition in Iberia and the Maghreb', *Eurasian Prehistory* [Preprint]. Available at: https://www.researchgate.net/profile/Joao_Zilhao/publication/273127196_Early_prehistoric_navigation_in_the_Western_Mediterranean_Implications_for_the_Neolithic_transition_in_Iberia_and_the_Maghreb/links/54f82fcb0cf28d6deca07e8e.pdf.