UNIVERSITY OF HUDDERSFIELD

DOCTORAL THESIS

Archaeogenetics of Southwest Europe

Author:
Gonzalo Oteo García

Supervisor:
Prof. Martin B. Richards

Co-Supervisors:
Maria Pala
Ceiridwen Edwards

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Archaeogenetics Research Group
Department of Biological and Geographical Sciences

September 14, 2020

Declaration of Authorship

I, Gonzalo Oteo García, declare that this thesis titled, "Archaeogenetics of Southwest Europe" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: 

Date: 14/09/2020

"The study of history is the best medicine for a sick mind; for in history you have a record of the infinite variety of human experience plainly set out for all to see; and in that record you can find for yourself and your country both examples and warnings; fine things to take as models, base things, rotten through and through, to avoid."

Titus Livius. *Ab Urbe Condita, Praefatio.*

UNIVERSITY OF HUDDERSFIELD

Abstract

School of Applied Sciences
Department of Biological and Geographical Sciences

Doctor of Philosophy

Archaeogenetics of Southwest Europe

by Gonzalo Oteo García

This thesis consists on three chapters that investigate the genetic past of Iberia using modern and ancient DNA.

The first part offers a snapshot of the current mitochondrial diversity in the Iberian peninsula based on a newly generated dataset with over one thousand fully sequenced mitochondrial genomes. The genetic depth and resolution of this dataset allowed to date the arrival of the vast majority of mitochondrial lineages to Iberia at the time of the Neolithic. It also made possible to describe patterns in some lineages, like U6, that were shaped by Medieval and later population movements which were not considered significant until now.

The second part explores the evolution and transformations of the population in the east of Iberia from the late Neolithic to the Middle Ages through the genomes of twenty ancient individuals sequenced to varying depths. The prehistoric individuals indicate little genomic contribution from local Iberian hunter-gatherers by the end of the Neolithic and beginning of the Copper Age. I also found evidence for two important admixture events whose genetic legacy has been lost. The first event is evidence of genomic influx of ancestry from North African and eastern Mediterranean sources into the local late Roman population. The second admixture event I detected is heavy and widespread admixture with North African migrants that settled in the region during the Islamic period.

The last part focuses on how ancient genomes can be interrogated in a comprehensive way using modern machine learning techniques to better understand what ancient individuals looked like. For this purpose I developed an alternative method using pseudo-phenotypic data to recreate, in an comprehensive way, a polygenic trait: skin pigmentation. My results confirmed that indigenous European hunter-gatherers had darker skin tones than later European populations while also explaining the different polygenic base of similar phenotypes.

Acknowledgements

I would like to use these lines to thank all the people that have walked alongside me, personally and professionally, since I first moved to the UK six years ago and later during my PhD. I am going to try to keep it cliche and just say that they all know who they are. Having said that, I am immediately going to start violating what I just said in the previous sentence because there are things I feel I must acknowledge professionally... I hate putting names black on white because I feel I am always going to leave someone out unintentionally.

Firstly, I want to thank all my three supervisors, Martin, Maria and Ceiridwen, and the Leverhulme Trust who gave me the opportunity to carry out my research on a topic that has fascinated me ever since I was eighteen or nineteen years of age, and especially Martin for also giving me the freedom to be the maverick I like to be sometimes.

Secondly, I have to thank my hardcore Huddersfield gang embodied by George, Ale, Marina, Bobby, Rita, Pierre, Franci and Simão for being the best mix of fun and science I will ever find myself into. It has been a hell of a ride starting an ancient DNA laboratory from the scratch. We started at the bottom and made it through thick and thin.

I am also very thankful to all the museums, institutions and people behind them (Museu Arqueològic Municipal La Vall D'Uixó, Museo Municipal de Arqueología y Etnología de Segorbe, Museu de la Valltorta, Museu Arqueològic de Gandia, Museu de Belles Arts de Castelló, Conselleria d'Educació, Cultura i Esport, Ajuntament de València, SIAP, SIAM) that trusted me and collaborated with precious samples from my hometown and other locations to be studied by me. Special shout-out to Marisa and Josep from el Museu de La Vall D'Uixó for the early support, to Vicente Palomar from el Museo de Segorbe, to Albert Ribera for opening the door of the SIAM to me, to Amparo, Gustau, Arturo and the people at the SIAP, to Llorenç and Josep, to Rafa Marínez from the ICV+R, and to the many more archaeologists I have crossed paths with during my periplus across Valencian museums. Thanks to Pedro Jiménez for allowing me to use his amazing illustrations of Valencia in this thesis.

It is beyond words trying to explain how much of an honour it has been to piece together the genomes of the people that made our cities and landscapes. Hopefully, this is not the end of the line and there will be more to come.

Finally, last but not least, big thanks to my dad for being the best scientific sparring ever, and also for always being there no matter what.

A mi madre.

Archaeogenetics of Southwest Europe

Mediterranean Iberia through the ages

Gonzalo Oteo García

December 9, 2020

Contents

1 General Introduction

1.1 The Fields of Archaeogenetics and Paleogenomics

The discipline of archaeogenetics studies the past of human populations using techniques of molecular genetics (Sokal, 2001; Pala et al., 2014), originally using data from modern populations (Bodmer, 2015). One of the pioneers in the field was Luigi Luca Cavalli-Sforza, who in the decade of 1960s started working with simple human genetic markers available at the time and laid the foundations for the archaeogenetics discipline (Cavalli-Sforza and Edwards, 1967; Cavalli-Sforza et al., 1994; Cavalli-Sforza, 1998; Cavalli-Sforza and Feldman, 2003). Cavalli-Sforza worked on the development of evolutionary trees building phylogenies and also made the first inferences about Neolithic migrations and the wave of advance based on gradients observed in classical genetic markers like blood groups, and *Fst* statistic measurements, from so-called unadmixed European populations (Figure 1). Although valuable, especially at a time when technical resources and data were scarce, these early ideas missed some important details about replacements, migrations, and mixtures that were impossible to see before aDNA became a reality. Some of the conclusions were wrong and are no longer accepted (e.g. Bertranpetit and Cavalli-Sforza (1991)). One of such missed details is that the European Neolithic genetic makeup did not remain undisturbed, later migrations that were unsuspected at the time, have modified the European genetic pool in complex ways (Richards, 2003; Haak et al., 2015).

The first characterization of maternally inherited genetic markers (mitochondrial haplogroups) while studying Native American populations (Wallace and Torroni, 1992; Torroni et al., 1993) yielded a much better proxy tool to infer migrations and provide an age estimate for such population movements (Richards and Macaulay, 2001; Richards et al., 2002, 2016). The nature and time of the Neolithic transition could now be much better understood and allowed to define the maternal genetic contribution of the European Neolithic (Richards et al., 2000; Richards, 2003; Olivieri et al., 2017; Pereira et al., 2017).

Current research has benefited immensely from the notable advances in molecular techniques in the last 40 years. The invention of the Polymerase Chain Reaction (Mullis and Faloona, 1987; Saiki et al., 1985, 1988), the discovery that ancient DNA (aDNA) can be retrieved and is viable for sequencing, and more recently the new high throughput sequencing technologies together the rise of bioinformatics have formed the core of a new discipline that interrogates the past about the same questions Cavalli-Sforza did, but with more powerful tools. This new field in charge of studying the past uses genetic material retrieved from ancient remains is what we call paleogenomics. This new discipline took advantage of aDNA to advance a further step, although the goals remain the same. The borders between the archaeogenetic and paleogenomics have become somewhat blurry since archaeogenetics has started to benefit from information provided by ancient genomes.

An important name in the field during these pioneering years was Svante Pääbo, his early works in the 1980s exploring the possibility of recovering DNA from human mummified tissue, helped to pave the way for the current research being done with ancient DNA (Pääbo, 1985; Pääbo et al., 1988; Pääbo, 1993). However, these early attempts with mummies relied on molecular cloning and in the end, none of the recovered fragments was actual aDNA. We only know that now since the fragments were too long to be ancient, and therefore nowadays this is accepted to had been bacterial DNA contaminants instead. Molecular cloning soon gave way to the newly developed and more convenient PCR techniques, which became standard for the next two decades. The PCR approach was not free of problems itself since it required extreme care to validate results free of modern DNA contamination, especially for anatomically modern human remains. The claim that a mitochondrial fragment from skin of an extinct quagga had been characterized in 1984 (Higuchi et al., 1984) is thought to be the earliest successful attempt to recover ancient DNA. Nonetheless, the research remained confined to the study of mitochondrial DNA or fragments of particular genes (Llamas et al., 2017) until the advent of the next-generation (or 2nd generation) sequencing revolution (van Dijk et al., 2018) and establishment of current criteria of

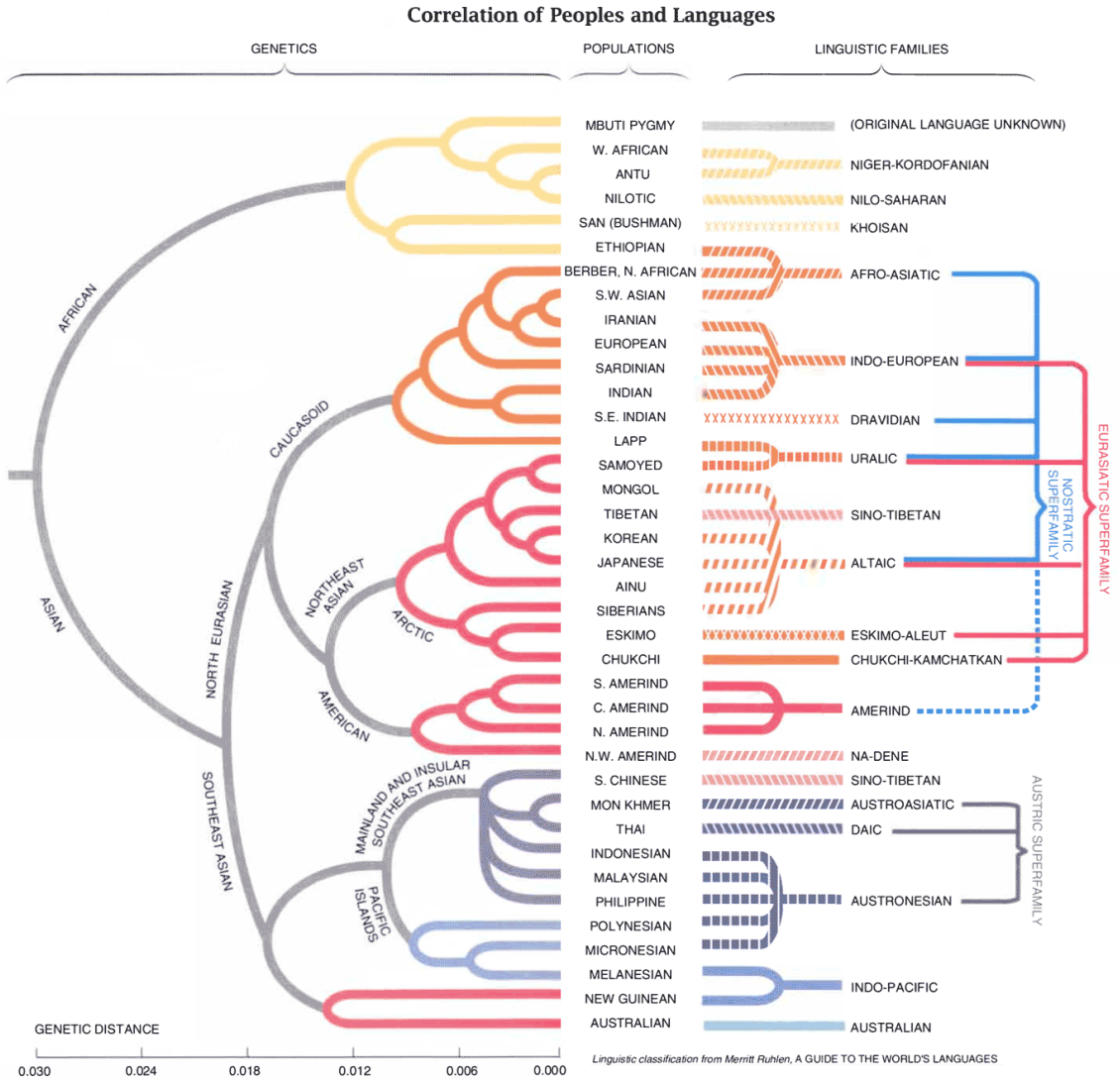


Figure 1: Cavalli-Sforza phylogentic tree based of *Fst* values contrasted to a linguistic tree (Cavalli-Sforza et al., 1994). From Cavalli-Sforza obituary in medium.com by John Hawks.

aDNA authentication (Leonardi et al., 2017).

The new era of high-throughput sequencing opened the door for the publication of the first human genome with Illumina technology (Rasmussen et al., 2010). This honor was bestowed on a 4.000-year-old Paleo-Eskimo individual (Saqqaq genome). The sample turned out to have very low levels of contamination by exogenous DNA which favoured a final genome coverage of around 20X. This milestone work also set the tone for the dominion of Illumina technology in the field, because since Saqqaq, the vast majority of ancient whole genome sequencing has been generated using Illumina technology. The realization that DNA was recoverable even from bones (Hagelberg et al., 1989) made possible that projects like the Neanderthal Genome Project (NGP) were attempted. The NGP culminated in 2010 with the publication of a draft genome from three 38.000-years-old Neanderthals from Vindija Cave in Croatia (Green et al., 2010). In the same year, an unexpected finding also saw the light: the first genetic evidence of Denisovans (Reich et al., 2010, 2011), an unknown population until then and a ghost from then onwards (Sawyer et al., 2015).

The Neanderthal Genome Project was probably the first big project to face the problems associated to work with highly fragmented old DNA and contamination by modern DNA (bacterial or otherwise) (Llamas et al., 2017). The novel criteria of authentication developed during the time of

the NGP standarized a set of rules and robust protocols currently used. It also became clear that it was necessary to establish dedicated ancient DNA facilities (Cooper, 2000; Lalueza-Fox, 2013), as echoed in the preceding years (Richards, 2005).

Ever since the publication of the NGP draft genome and the Saqqaq genome the work with modern humans has proliferated dramatically. The study of ancient DNA combined with modern data has achieved several scientific milestones such as confirming the origin of anatomically modern humans (AMH) burying the multi-regional hypothesis in the process (for the moment at least, see Xing et al. (2019)), and also an approximate dating for the time of the Out-of-Africa dispersal: roughly around 60,000 years ago. Although there is increasing debate about the possibility of an earlier dispersal into the Arabian peninsula from East Africa (Fernandes et al., 2012). We know today that almost all indigenous sub-Saharian Africans mitochondrial lineages belong within haplogroups L0 to L6. Major Eurasian basal lineages derive from macro-haplogroup M and N which both stem directly from the L3 lineage, which is widespread in Africa. L3 is therefore the root for almost all mitochondrial lineages found outside Africa (Cann et al., 1987; Soares et al., 2009).

There has been room for surprises in the study of the past of humanity (Leonardi et al., 2017). The detection of a previously unexpected signal of hybridization between anatomically modern humans and Neanderthals (Green et al., 2009; Meyer et al., 2012) was an idea widely rejected or not contemplated before the advent of ancient DNA (Reich, 2018). Today it is an accepted reality. We have also gained some interesting insights into the peopling of the Americas through the Mal'ta boy, a 24,000-year-old genome from Siberia. This individual appears to be ancestral to both Europeans and Native Americans, and somehow his ancestry type was contributed to the first peoples that crossed into the American continent via Bering around 20.000 years ago and became the ancestors of modern Amerindians (Raghavan et al., 2014).

In his time, Cavalli-Sforza was able to reconstruct a broadly accurate human phylogeny using the paltry data available at the time, but he probably did not imagine how much detail the field would achieve in few years. Perhaps, one of Cavalli-Sforza's most notable contributions is the introduction of the Principal Component Analysis (PCA) to the field of genetics (Figure 2). Now, thanks to bioinformatics, Cavalli-Sforza's idea and the information contained in thousands of genetic markers from individual genomes obtained with high-throughput sequencing have been brought together to build genetic maps that can reveal differences between populations. PCA has some intrinsic limitations and has to be interpreted carefully, but it is in its own right a very powerful tool to present visual information in a comprehensive way. This is probably one of the reasons why it remained a popular tool for analysis in population genetics until the rise of STRUCTURE (Pritchard et al., 2000), which was a more sophisticated K-means clustering approach with a stronger theoretical base. However, a milestone work by Patterson et al. (2006) was a key factor that sparked the re-enhancement of the popularity of PCA thanks to the incorporation of statistical backing suited for the inference of population structure using genetic data (Patterson et al., 2006; Novembre et al., 2008). The revival of statistically solid PCA along with ADMIXTURE (Alexander et al., 2009) became the canonical standard combo analysis in population genetics. ADMIXTURE is a more powerful STRUCTURE-like software that has ended up surpassing STRUCTURE itself. The Procrustes transformation method to superimpose genomes with missing data into a PCA shape space (Schönemann, 1966), and later the ability to mathematically project multiple samples onto a PCA are contributions that have allowed researchers to incorporate into the analyses huge numbers of genomes that include missing data. This was a crucial step forward, given the incompleteness of genetic data retrieved from ancient genomes (Olalde et al., 2015). These techniques allow to visualize ancient individuals with missing information on top of modern populations (Patterson et al., 2012), before these contributions there were severe limitations to the number of ancient genomes that could be analysed along with modern genomes as it can be seen in works from only a few years back (e.g. Sánchez-Quinto et al. (2012)).

PCA and ADMIXTURE are indeed useful tools to detect structure within and between populations to some degree, but they do not allow researchers to formally test admixture events. The

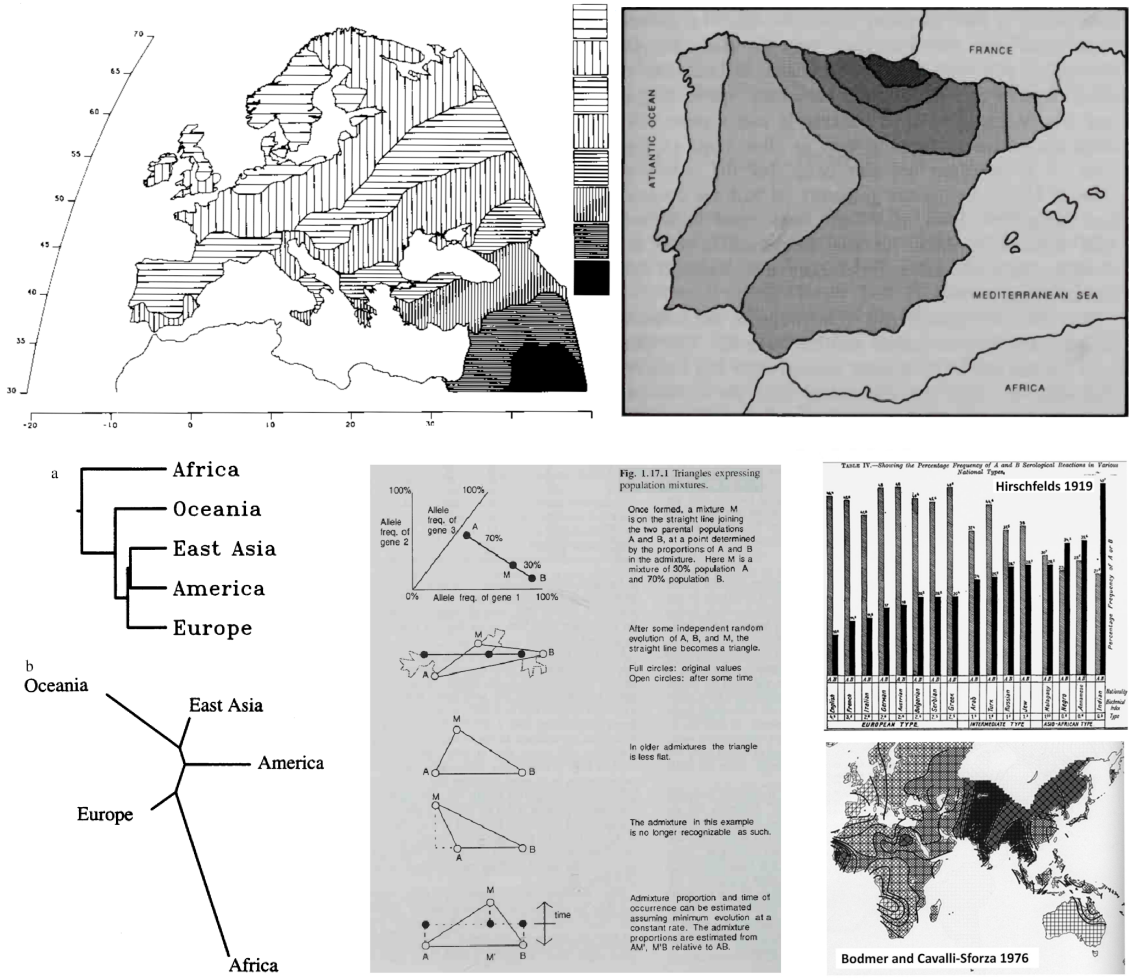


Figure 2: Early examples of PCA usage and *Fst* based phylogenies by Cavalli-Sforza and other iconic figures produced by him, such as geometric interpretation of admixture events and a map of ABO blood groups frequencies. From Cavalli-Sforza et al. (1994) and Bodmer (2015).

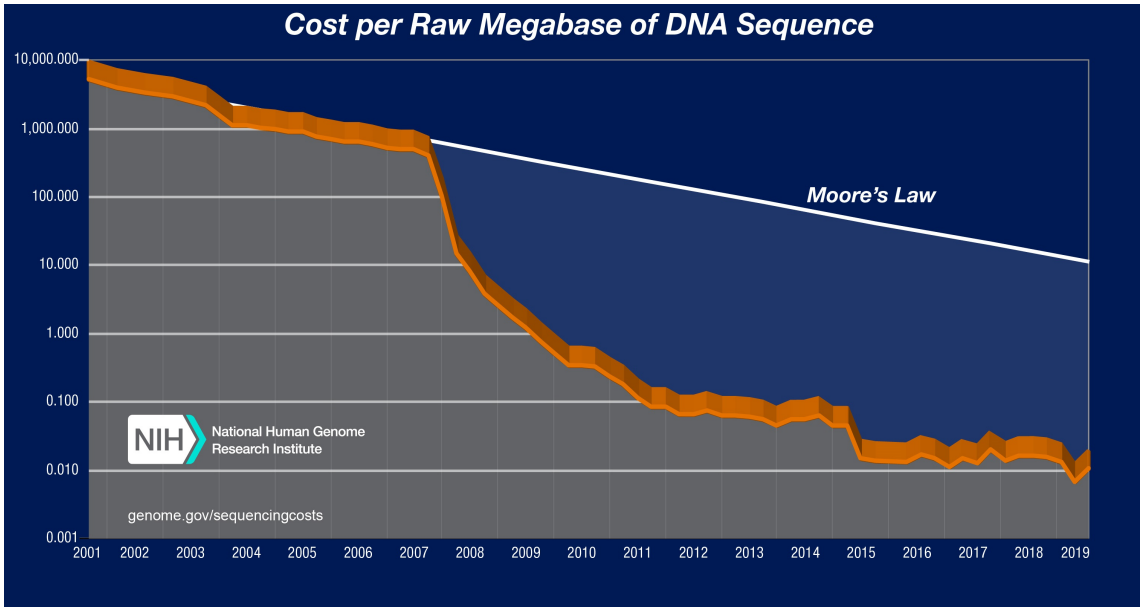


Figure 3: Cost of genome sequencing per year. Data from the NHGRI Genome Sequencing Program. Available at: www.genome.gov/sequencingcostsdata (Wetterstrand, 2018).

differences observed in PCA and ADMIXTURE results can arise via different mechanisms or population histories. Another push to consolidate the methods to study the history of human populations was provided new formal methods including f-statistics and D-statistics tests, which have become fundamental tools to investigate signals of shared drift and admixture episodes in past and present populations and individuals (Patterson et al., 2012). These techniques are among the most popular and familiar analyses currently being applied to study ancient and modern genomes, but like any other field that is fast developing and improving they only represent a fraction of the body of techniques currently being used.

Over the last ten years, since the introduction of the second-generation sequencing technology developed by Roche (454 pyrosequencing) in 2005 and Illumina (Solexa sequencing) in 2007 the cost of genome sequencing has seen an exponential reduction and more recently a steady stabilization in price (Figure 3). These innovations and competition between companies have reduced the cost of genome sequencing greatly since the beginning of the 2000s decade. This trend has democratized the field and allowed small labs to dare attempting more genome sequencing. While generating high-coverage complete genomes from ancient and modern samples can still be challenging and expensive depending on various factors, low-coverage genomes still can be a source of valuable information at low cost about, gene content, polymorphisms, repetitive elements and admixture among various other items (Rasmussen and Noor, 2009).

Since the entry of Biology into the era of bioinformatics, the challenge now lies on how to fully extract the huge amount of information derived from full genomes. In the coming years we should expect more and more publications with ever growing numbers of ancient genomes (Olalde et al., 2019) but focusing on smaller and smaller geographical regions across the globe (Antonio et al., 2019).

1.2 Ancient DNA

Ancient DNA (aDNA) is any genetic material extracted from samples whose age ranges from decades to thousands of years old, the upper limit is thought to be around one million years of age (Willerslev and Cooper, 2005). Such samples typically come from archaeological sites or museum collections (Figure 4). The main feature of aDNA is that it has fragmented into short sequences. In practical terms, the nature of aDNA is more related to the state of the DNA in the sample, rather than to its age (Leonardi et al., 2017; Llamas et al., 2017).

The work with aDNA has emerged as a new dimension that allows studying ancient human history and long gone populations but it presents some inherent difficulties. As mentioned above, all aDNA is found in very fragmented pieces due to degradation (typically 20-80bp long), and it is not equally well preserved in all parts of the skeleton (Pinhasi et al., 2015). Post-mortem fragmentations acts mainly through depurination. However, the second-genertion sequencing technologies have turned this highly fragmented state into an advantage since this technology requires shearing DNA into a myriad of short fragments for sequencing. On top of this ancient DNA typically displays a certain damage patterns at the ends of the surviving fragments (Briggs et al., 2007). Ancient sequences usually have single-stranded ends, which favours faster rates of cytosine deamination. The result is higher C-to-T transitions at the ends of reads. These deaminations can be a means to validate genuine endogenous ancient DNA (Hagelberg et al., 1989; Hagelberg et al., 2015) acting like a flag that distinguishes the endogenous ancient sequences from modern contaminant DNA (Figure 4).

Ancient DNA specific protocols (Adler et al., 2011; Dabney et al., 2013; Allentoft et al., 2015; Damgaard et al., 2015; Gamba et al., 2016) demand, among other things, the use of UV light to irradiate the surface of the samples and deep cleaning of the working facilities to reduce bacterial or exogenous DNA contamination. Other requirements are the need for physical separation between the ancient DNA extraction room and further steps, facilities with positive air pressure, and other related profilactic measures. Researchers also have to wear sterile costumes and masks to minimize exposure, and two layers of gloves. Validation of results should include independent replication and sequencing the researcher's own mitochondrial DNA to check for eventual undesired self contamination. However, the varying numbers of mitochondria present in different tissues might introduce bias when estimating levels of contamination so it is also recommended to use autosomal chromosomes to estimate contamination rates (Green et al., 2009).

Very much like the science of archaeology, the extraction of ancient DNA is an invasive and destructive process that generally ends with a partial or total destruction of the archaeological material and that is why it is important to know what material has to be targeted for the most efficient extraction. The petrous bone, which is the inner part of the ear in the skull, it is a very dense bone and for this reason it favours better DNA preservation Pinhasi et al. (2015). Molars are good candidates too, as long as the roots are not open and exposed. Phalanxes and the epiphysis of long bones can be used for ancient DNA extraction as well (Damgaard et al., 2015). In the case of the petrous bone it is necessary to cut or drill a wedge in order to obtain sufficient amount of bone powder from which DNA can be extracted, and in the case of molars, roots are cut off to be powdered (Rohland and Hofreiter, 2007).

Preservation of aDNA can be affected by several factors and depending on the circumstances some will have a bigger or smaller impact than others. Intuitively it is natural to think that age is the main factor affecting aDNA survival, but although it is important, it is not the main factor. One of the variables of importance for aDNA preservation is the climate, more specifically the temperature of the environment (Damgaard et al., 2015; Hansen et al., 2017). Colder regions, such as northern Europe, favour an optimal aDNA preservation, Siberian permafrost being the paradigmatic example. In extreme cold conditions, there is likely to be less bacterial contamination because the environment is harsh for them to proliferate. Warmer places tend to have lower successful DNA recovery rates. However, warm places that favour natural mummification and survival of tissue can also be good. Other factors include the characteristics of the site and the

Figure 4: Stages of sample collection, laboratory workflow and bioinformatic processing.

burial conditions, such as the type of soil where the sample is buried and how it was buried (Figure 5). Acidic soils for example, tend to dissolve the bones until they disappear, which obviously extinguishes any chances of recovering DNA (Figure 5). Where the samples are buried is also important: caves for example tend to be better for aDNA preservation since they act as a box protecting the sample from potential external high temperatures, rain or water reaching the bones, alkaline conditions, etc... On average, open air burials tend to yield less aDNA than what can be recovered from samples in caves. Samples affected by fire or water are not good omens for aDNA recovery, water tends to wash away or destroy the surviving DNA sequences within the bones, by acting as a catalyst for enzymes or bacteria activity. On the other hand, fire destroys the native structure of DNA. Available data indicate that the successful recovery rate of mitochondrial ancient DNA, more readily recovered due to its high copy number, from the south of the Iberian Peninsula is between 20-45% whereas in central and northern Iberia the rate is around 70-80% (Szécsényi-Nagy et al., 2017).

Even though damage patterns are crucial to verify the authenticity of the aDNA generated it is still necessary to prevent non-real mutations entering downstream analyses or at least reducing the impact they can have. This can be achieved before sequencing with different approaches or after sequencing by trimming the ends of the read or downgrading base quality scores according to their

probability of being an artificial error (Llamas et al., 2017). The end result is that the quality of an ancient genome after an extensive curation process, however incomplete, can match the quality of a modern one (Figure 4).

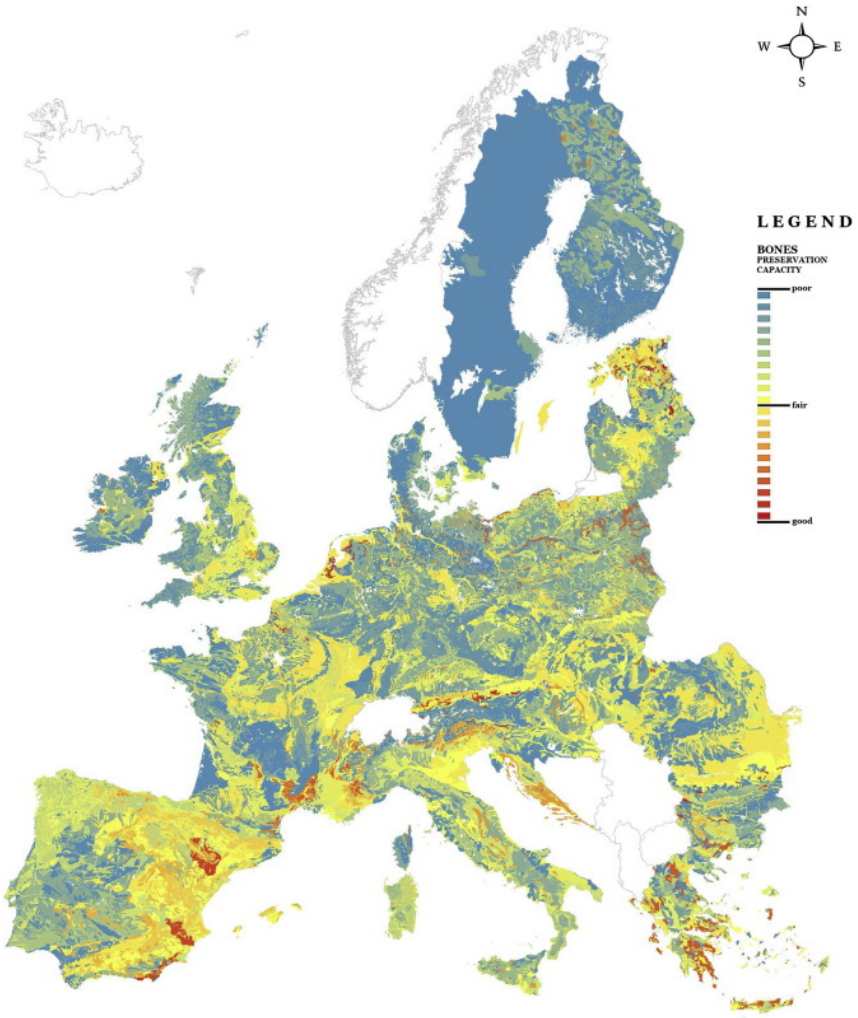


Figure 5: Map of Europe showing the bone preservation capacity of different regions based on soil type which affect the amount of DNA that survives. It may seem a counter-intuitive map at first since it does not take the important role of temperature into account. Taken from (Kibblewhite et al., 2015).

.

1.3 Genomes, Uniparental Markers and Phylogeography

Mammalian species share an important feature in the architecture of their genomes: an XY sex-determination system. This system sets a distinction between two categories of chromosomes; autosomes and allosomes (or sex chromosomes). In humans, allosomes represent the 23rd pair of chromosomes in the genome. Females are carriers of two chromosomes of the same type (XX) and have 23 homologous chromosome pairs, whereas males carry two different ones (XY) and only have 22 homologous pairs. However, X and Y chromosomes still retain a small homologous region that can be used for recombination. This is important because when studying a human genome we make use of a somewhat arbitrary division on the genome into three parts. The first two correspond to the autosomal genome which includes the 22 autosomes of the human genome, and the Y chromosome that is treated independently. The third part is a very small part of our genome that we inherit maternally because it is found in the mitochondria of cells, and therefore it is called the mitochondrial genome.

1.3.1 The Autosomal Genome

The completion of the Human Genome Project between the years 2001 and 2004 was the outcome of more than a decade of research and technical innovation, an herculean task that started with Sanger sequencing and by the time of its completion had seen the emergence of shotgun sequencing technology, with public and private institutions involved in the endeavour (Consortium, 2001; Craig Venter et al., 2001).

The human nuclear genome comprises 23 pairs of chromosomes and its total size makes for approximately 3 billion base pairs (Jobling et al., 2013). The autosomal genome covers most of that size since it is constituted by chromosomes 1 to 22 out of the total 23 pairs. The order of the chromosomes is not arbitrary because autosomes are ordered roughly by size (Figure 9). Chromosome 1 with 250 million base pairs is the biggest and chromosome 21 (45 million base pairs) is the smallest. The overwhelming majority of genetic variation that is present in the genome is neutral and has no evolutionary, functional or phenotypic impact (Auton et al., 2015). On average a human genome can have between four and five million variant positions (Figure 6). However, the amount of existing variants in a population depend on several factors like the time elapsed since the founding event of the population, so the older a population is the more variants it will contain. This explains the great diversity found in sub-Saharan African populations compared with the rest of the continents (Figure 6). Eurasians and Native Americans are ultimately derived from more or less severe bottlenecks and founder effects, as reflected in their genomic fingerprints (Figure 6). Admixture events also contribute to expanding the genetic pool of variants of populations. These mutations present in the genome can be of different types: insertions, deletions, single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), and copy number variants (CNVs).

Deciphering the human genome has not lead to find the cure for all the diseases the scientific community thought it will, but it has helped to build human reference genome and unveiled a complexity in our genetic makeup much deeper than previously thought (Akey et al., 2004). Currently there are two main version of the human reference that are used: GRCh37 (released in 2009) and GRCh38 (released in 2013). Over time patches have been added to both versions in use (GRCh37.p13 in June 2013 and GRCh38.p13 in March 2019). GRCh37 was the nineteenth version of the reference and that is why sometimes it might be referred to as *hg19*. Despite being the older version, *hg19* is still widely used because the transition to GRCh38 has not been fully embraced by everyone yet. To preserve anonymity, the reference genome was made from many different individuals that volunteered their DNA. Although efforts were made to keep the identities of the donors private it is clear that the majority of the reference genome came from an individual of African origin (Auton et al., 2015). The curation of the reference is carried out by the Genome Reference Consortium (GRC) which is made up of various institutions worldwide.

As mentioned, whole genome sequencing (WGS) is the most powerful method when studying

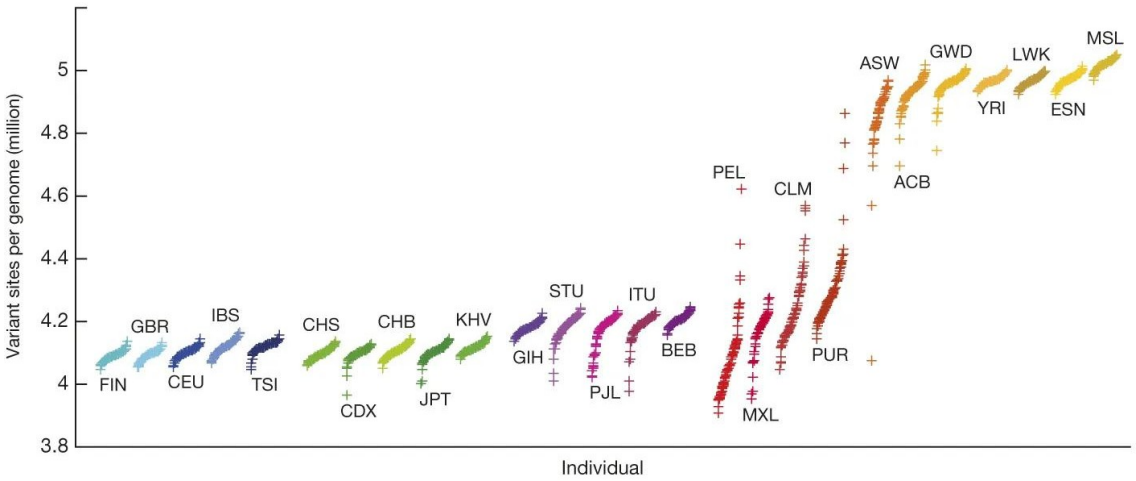


Figure 6: Number of variants found in the genomes of individuals from various population from all continents (Auton et al., 2015). Europeans in blues, East Asians in greens, South Asians in purples, Amerincans in reds, Africans in yellows.

ancient and modern samples (Nielsen et al., 2017), but it has some limitations. WGS is not only economically expensive, albeit much more affordable than a few years ago (Figure 3).

Nevertheless, the study of human evolution and its populations can be approached in alternative ways that have enough resolution to distinguish between samples in detail (Sudmant et al., 2015; Hellenthal et al., 2014). For this instance there are two approaches. One is using a Single Nucleotide Polymorphism (SNP) capture-genotyping technique (Nielsen et al., 2011), the other is by shotgun sequencing aiming to cover the whole genome at random. SNP genotyping targets certain predetermined positions known to be variable in the human genome using a chip designed *ad-hoc* for this purpose. This approach allows us to differentiate between populations at a high level of resolution but the downside is that it misses on other genomic information due to the fact that each chip consistently looks at the same positions over and over. Also, two chips designed by different laboratories can have very little overlap, making the possibility of merging datasets very difficult or not worthwhile. The approach is based in the assumption that a high degree of redundancy exists in the genome due to linkage disequilibrium. The level of resolution depends on the number of SNPs available. Currently there are SNP capture chips (Carpenter et al., 2013) designed to be able to target up to 1.2 million of known SNPs, across all autosomes, sex chromosomes and mitochondrial DNA in humans. Shotgun sequencing on the other hand, allows us to obtain random short chunks of the genome with the information from the SNPs and neighboring bases. This allows for imputation of the missing gaps in the ancient genome and opens the door to more refined types of analysis such as runs of homozygosity (RoH) (Ceballos et al., 2018) or even exploring haplotype sharing (Cassidy et al., 2016; Martiniano et al., 2017; Cassidy et al., 2020).

In the field of archaeogenetics, the availability and use of genomic data, in both forms, has proved itself the best way of identifying ancestries of past populations, measuring dissimilarity and affinities with other populations and individuals, detecting mixture between populations, measuring shared drift and predicting (or at least trying) phenotypic traits (Mathieson et al., 2015; Cassidy et al., 2016; Brace et al., 2019; Jensen et al., 2019).

Sub-Saharan Africans are a deeply divergent branch to any non-African population, and at the same time many surviving African populations are old and deeply divergent between themselves (Figure 7). Some Oceanian groups (Melanesians, Australians and Papuans) are also very distinct from other world populations due to early isolation of their islands. On the other hand, in the continental ground of Eurasia we found a progressive gradient of genetic change from Western Europe to East Asia. As expanded on further below, the broad genetic ancestry of modern Europeans can be model as mixture of three main components in different proportions depending on

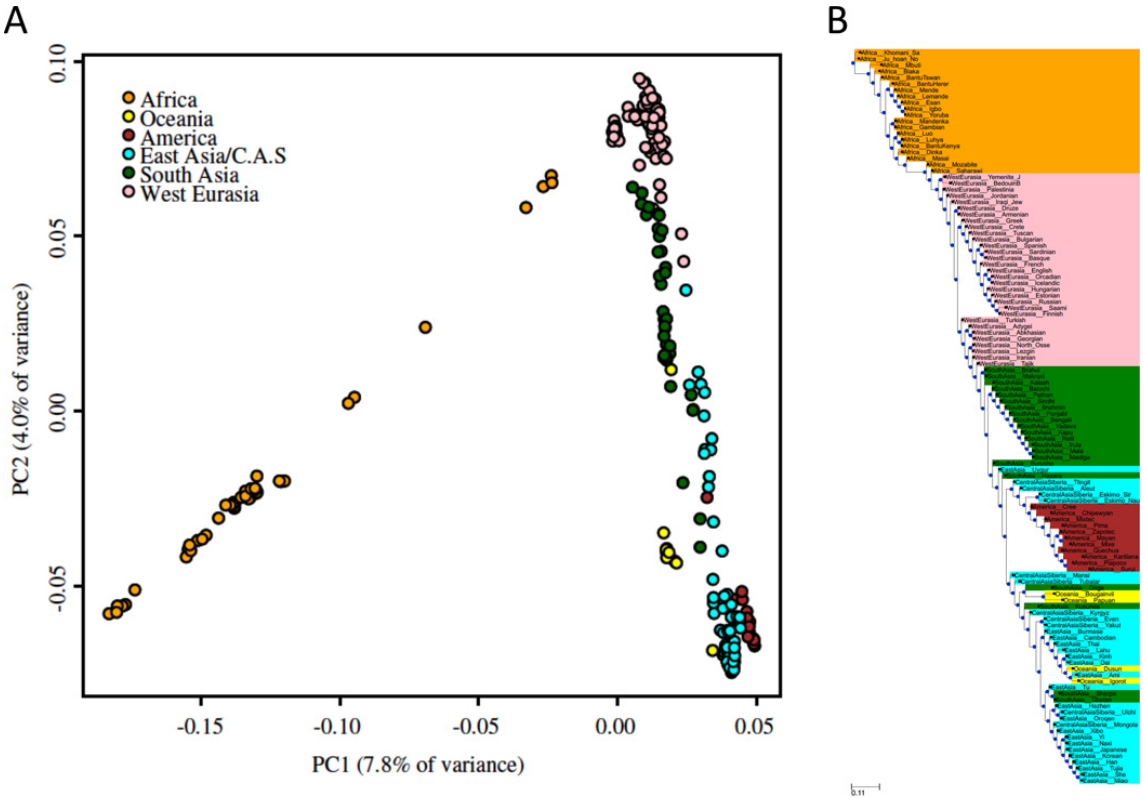


Figure 7: PCA (A) and NJ tree (B) made with the individuals from the world populations sequenced in the Simons Genome Diversity Project Mallick et al. (2016).

geographic location. The first genetic component is Mesolithic hunter-gatherer related ancestry, the second component was introduced with the onset of the Neolithic by early migrating farmers from Anatolia (Feldman et al., 2019), and the last genetic component is a Pontic-Caspian steppe related ancestry whose trademark is derived from Caucasus hunter-gatherers (Jones et al., 2015).

1.3.2 The Mitochondrial Genome

Mitochondrial DNA (mtDNA) has been one of the most widely used markers in the study of human past, and one that has given many answers on the history of human migrations (Richards et al., 2016). In order to understand the phylogenetic and phylogeographic methodologies (Richards et al., 1998, 2002), a brief introduction to the mitochondrial genome is presented (Figure 8).

The mitochondrial genome (or mitogenome) is a circular piece of DNA found in the mitochondria organelles in eukaryotic cells whose main role is the production of energy for the organism. The mitogenome consists of 16,568 pairs of bases, which is a small size and relatively easy to handle. There are currently two reference mitochondrial genomes. One is the arbitrary but widely used, rCRS (revised Cambridge Reference Sequence) (Figure 8). The rCRS reference corresponds to an haplogroup found in a branch of the H lineage (H2a2a1). The other reference sequence is the RSRS (Reconstructed Sapiens Reference Sequence) (Andrews et al., 1999; Behar and et al, 2012*a*), which sits at the base of the mitochondrial tree as root for all the modern human branches that exist or have existed. In the literature it is sometimes referred to as the *Mitochondrial Eve*. Note that this does not mean the woman bearing this sequence was the only woman alive at the time. The coalescence time for the RSRS has been estimated using the molecular clock at around 190kya (Soares et al., 2009). The human mitochondrial genome has some characteristic features. The existence of a non-coding region of 1120bp, which is in turn divided into the hypervariable region I and II (HVS-I, HVS-II, there is also a short HVS-III) (Figure 8). This region is also called the control region, and was the one used by the pioneer studies with mtDNA to characterize lineages (Macaulay et al., 1999). Hypervariable regions I and II are located in the 16,000-to-600bp interval. By convention, the position of the first base was allocated in the control region. The control region evolves at a faster rate than the rest of the mitogenome and accumulates around a quarter of the variation and many of the diagnostic mutations (Figure 8). The high rate at which mutations arise in the mitochondrial genome allows us to link haplogroups to populations or geographical locations and infer demographic events in the past. This is the basis of the phylogeographic approach. It is belived that the accumulation of variation in the mitochondria is largely neutral but other factors such as climate may have an influence (Balloux et al., 2009).

Since mitochondrial lineages are always maternally inherited, the information obtained only represents the variation in the matrilineal genealogy. When the global phylogenetic tree was reconstructed with mitochondrial DNA, it was established that the most common recent ancestor (mt-MRCA), had lived some 200,000 years ago (Cann et al., 1987; Soares et al., 2009) (Figure 11). She is the most recent woman from whom all humans today descend in an uninterrupted line purely through their mothers (Richards and Macaulay, 2001).

In Europe for example, between 40 and 50% of the mitochondrial diversity (Carter, 2007; Pereira et al., 2009) is represented by lineages of haplogroup H, although it is unclear exactly how it reached such a high frequency. The coalescence age of H is estimated to be around 20kya, after it branched off from the HV root. In the late Neolithic, H experienced a burst in diversity that is reflected by the distinctive star-like branching in the tree. The most important sub-branches of H in the European population are H1, H2, H3 and H5, which account most of that 40-50%. Nevertheless, lineages of haplogroup H are not endemic to Europe because prehistoric migrations of people that affected Europe that reached other regions of Eurasia also introduced these lineages into the new lands. Such is the case of the Indian subcontinent, where we can find sub-lineages of H2 and H13 that were likely introduced during the Bronze Age (Silva et al., 2017).

Other relevant haplogroups that are part of the European maternal diversity include U (including K), J, T, and X. Haplogroup lineages within U like U2, U5, U8 and probably even U6 (Peştera Muierii 1) were introduced in Europe with some of the first AMH that colonized the territory, since they have been recovered amongst the oldest European genomes sequenced to date (Posth et al., 2016). Haplogroups descended from the maternal lineages the first hunter-gatherers brought with them can still be found in the modern European population, mostly in the form of U5 lineages that have survived until the present. Haplogroups within U4, U5a, and U5b occur in modest

frequencies but are commonly found. U5b is distributed across all western Europe, U4 and U5a are more typically found in Eastern Europe. The remaining haplogroups belonging to lineages K, J and T are also typically Near Eastern, European and North African, and are accepted to be the core of the mitochondrial package introduced by the migrating early farmers during the Neolithic from Anatolia. Some of these Neolithic lineages identified so far include haplogroups K1a, K1b, K2a of the K family, which can be found in frequencies oscillating between 5 and 15% depending on the region in Europe; T haplogroups like T1a and T2b in frequencies between 3 and 10%; and J1a, J1b and J2b lineages, with frequencies of J lineages across Europe varying between 5 and 20%. There are also other diverse haplogroups common throughout Europe that belong to HV, HV0, W, X and I with peaks of 5% in particular regions. N1a1a is an example of a lineage that was quite common during the Neolithic in central Europe but today is virtually extinct in Europe (Haak et al., 2005).

At the other end of the Eurasian continent, we found that the lineages of the mitochondrial tree that dominate the maternal diversity in East Asia and South Asia belong to haplogroups within the macro lineage M, including haplogroup D but also haplogroup B, although R and N are ubiquitous too.

In Africa we find the most deeply divergent lineages, including the L3 which is the common root for all mitochondrial lineages originating outside Africa. These are labelled as L haplogroups, going from L0 to L5. Evidently, these haplogroups harbour the greatest diversity.

Native Americans on the other hand, carry a reduced number of different haplogroups: D1, C1a, C1b, C1c, A2, B2 and X2a. This is unsurprising given that the ancestors of modern indigenous nations that colonized the American continent for the first time, experienced a genetic bottleneck due to a founder effect that greatly reduced the diversity of their uniparental marker types.

Finally, it is also important to note that due to the large numbers of mitochondria present in the cells, the mtDNA survives better than nuclear DNA after the death of the organism. This represents an advantage since ancient mitochondrial DNA is more likely to be recovered compared to nuclear ancient DNA. However, contrary to the rest of genomic DNA, expression of mitochondrias in the petrous bone is depressed and the recovery rate is lower compared to teeth (Pinhasi et al., 2015).

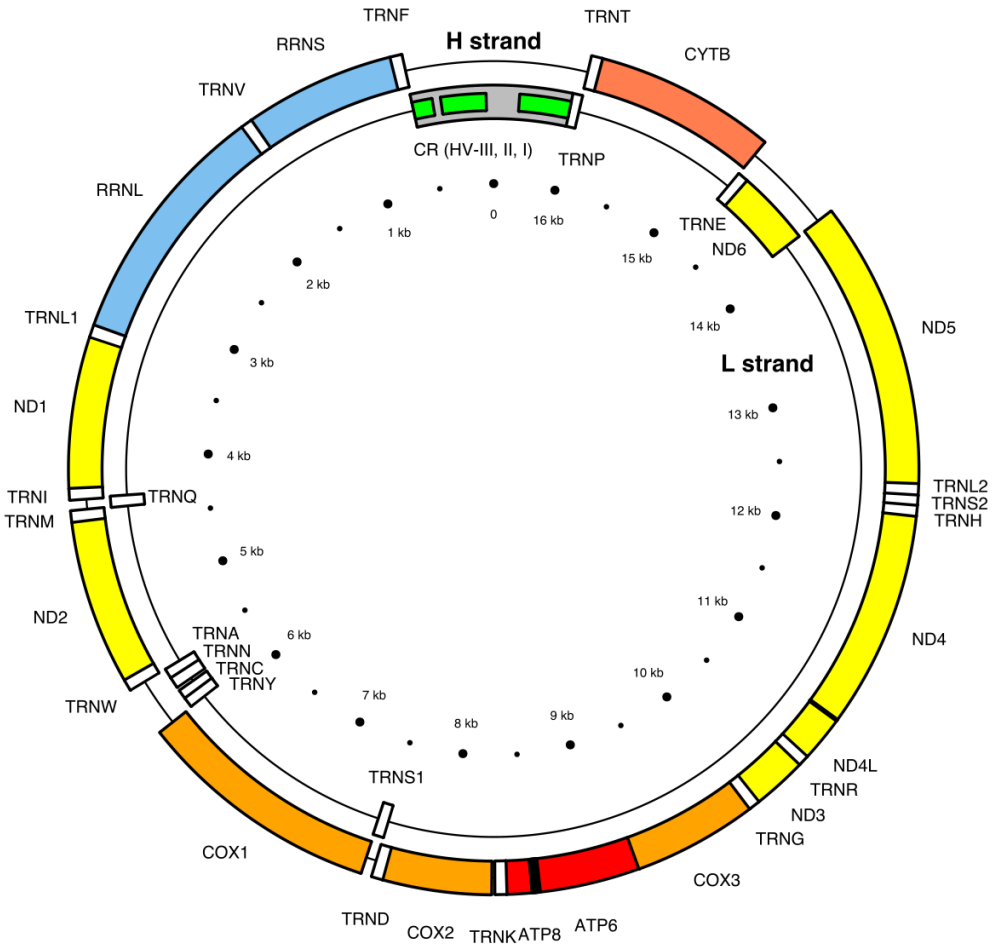


Figure 8: Representation of the mitochondrial genome showing the Control Region and all coding regions for tRNA and other genes.

1.3.3 The Y Chromosome

During the evolutionary trajectory of the sex chromosomes, the Y chromosome lost the majority of its homology with the X chromosome and is currently a dwarfed pseudo-homologous pair for the X (Figure 9). As mentioned above, a Y chromosome is inherited paternally but an X chromosome is inherited in the same fashion as the autosomes; the only difference is that half of the time the X pairs with a non-homologous chromosome. This affects its recombination behaviour because it is not possible to form cross-over points with Y (Jobling and Tyler-Smith, 2017). Although X is officially a sex chromosome, it sits halfway between autosomes and allosomes since chromosome Y actually has two pseudo-autosomal regions (*PAR1* and *PAR2*) that effectively recombine during meiosis. The reduced size of Y also implies a reduced number of genes. The Y chromosome has little more than ten genes of known function, and they all relate to mechanisms of fertility in males, like *USP9Y*. Mutations in the *USP9Y* gene that cause loss of function are associated with abnormal sperm production (Tyler-Smith and Krausz, 2009). On the other hand, the 155 million base pairs of the X chromosome make for thousands of genes.

Figure 9: Digital karyotype of the human genome as displayed in the NCBI Genome Data Viewer. The relative size of each autosome can be appreciated along with the dwarfism of Y compared to X.

In the case of males, another implication of having one copy of Y instead of an extra X chromosome, is that X-linked alleles that cause a certain phenotype or disease will always manifest themselves, even when recessive, since there is not an alternative allele present to mitigate such effects. Recessive diseases that are linked to the X are therefore more common in males than females. Perhaps, the most well known example of this is the "royal disease" or the famous cases of haemophilia in European royalty, which was introduced by Britain's Queen Victoria to other royal houses, including Spain and the Tzars of Russia (Stevens, 1999).

The great majority of mutations that occur on the MSY (the male-specific part of the Y chromosome) are neutral and since there is no recombination they are also the only source of variation (Jobling and Tyler-Smith, 2017). This is a reason why the Y chromosome makes for such a good candidate as a molecular clock (Underhill et al., 2001; Wei et al., 2013). Useful Y-chromosome genetic diversity presents itself in two forms: single nucleotide polymorphisms (SNPs) and microsatellites or short tandem repeats (STRs). Currently SNPs are the favoured tool to define Y-chromosome lineages (Jobling and Tyler-Smith, 2003).

Large numbers of SNPs have been discovered in the MSY region, for example, the 1000 Genomes Project that analysed genomes from more than 20 different populations around the world, and between SNPs, STRs and copy number variants (CNVs) identified over 65,000 variants were described.

In phylogeography the male specific Y segment or non-recombining Y (MSY or NRY) is the reference. It covers 95% of the length of the Y chromosome and is passed exclusively through

the generations from fathers to sons. Because the different haplogroups in the Y chromosome tree display geographical structure they can be studied with a phylogeographic approach. The MSY is longer than the mitochondrial genome (55 Mb pairs) but has a genetic nature consisting of tandem repeats that hamper the sequencing process which initially made it less popular to work with (Bachtrog and Charlesworth, 2001). The nomenclature has been traditionally confusing due to alternative mutation naming, tree updates and new methods to define the same mutation has finally standardized since the whole genome sequencing era (Jobling and Tyler-Smith, 2017), and the study of Y phylogenies is a fundamental complement to mtDNA (Figure 10).

Like the mitochondrial genome but even larger, the different haplotypes of the Y chromosome phylogeny accumulate mutations that define haplogroups which in turn are inherited uniparentally and therefore become a source of information about the past of the male line of some groups (Finocchio et al., 2018). A calibrated tree set the time of the TMRCA at 190,000 years ago and the TMRCA of all non-African lineages at 76,000 years (Hammer et al., 1998). A more recent work sets the TMRCA to be ~110,000 years, and the lineages found outside Africa dated to 65,000 years (Wei et al., 2013). One of the most recent reconstructed phylogenies of Y however, in Karmin et al. (2015), pushed back the estimate for the TMRCA at around 250,000 years ago (Jobling and Tyler-Smith, 2017).

In the diversity of European Y lineages, haplogroups I and G have been very common in the past, in the form of I2a and G2a haplogroups, and today we still find haplogroups like J, G and I, but they are very minor compared with the current abundance of haplogroup R1. This lineage which has two major branches, R1a (Underhill et al., 2001) and R1b (Solé-Morata et al., 2017). The coalescence time for R1b is estimated to be around 20,000 years ago (Poznik et al., 2016) and appears in high frequencies in Western Europe (Karmin et al., 2015; Solé-Morata et al., 2017) (Fig. 8). The sub-branch R1b-M269, whose age is around 8,500 years, is relevant because there is strong genetic evidence that it was largely introduced by a male-driven migration during the Bronze Age, although this did not become clear until genomic studies and aDNA shed light on it (Myres et al., 2011; Haak et al., 2015; Finocchio et al., 2018). As mentioned above, male lineages like I and G were common in European prehistorical individuals (Battaglia et al., 2009; Batini et al., 2017; Lipson et al., 2017) but they experienced a sharp decline during the Bronze Age, almost completely replaced in some regions, and this decline coincides with the arrival of R1b to Western Europe some 5000-4000 years ago. The frequency of R1b-M269 across modern populations of Western Europeans ranges from 50% to over 90%, but the average is closer to 75% in the continent (Figure 10). The North African and Sub-Saharan landscape paints a similar picture: the Y-chromosome diversity is overwhelmingly dominated by one major haplogroup: in this case, E1b. The coalescence time of E1b is just over 10,000 years. The patterns of E1b and R1b in the phylogenetic tree are both strikingly similar: relatively young lineages that experience a rapid growth. This rise to high frequency likely reflects successful population expansions by particular groups. The picture in Asia is different though. In South Asia, we find a greater diversity of haplogroups representing Y-chromosome haplogroups present in the population (C, G, H, J, R, L). The presence of R1a, however, which is found in frequencies of typically 20-40%, reveals connections to Europe in a similar fashion to mitochondrial haplogroup H. R1a also spread into South Asia following Bronze Age migrations from the Caucasus region and the Eurasian steppe (Figure 10). In East Asia, although the majority of male lineages belong to haplogroup O, and the colescence age for these O lineages is older than in the E1b and R1b, closer to 30,000 years in most cases (Figure 10). Among Native Americans if we exclude lineages of European origin, almost all individuals are carriers of haplogroup Q1a. As discussed above, this is due to an ancestral genetic founder effect occurred during the original colonization of the Americas (Figure 10).

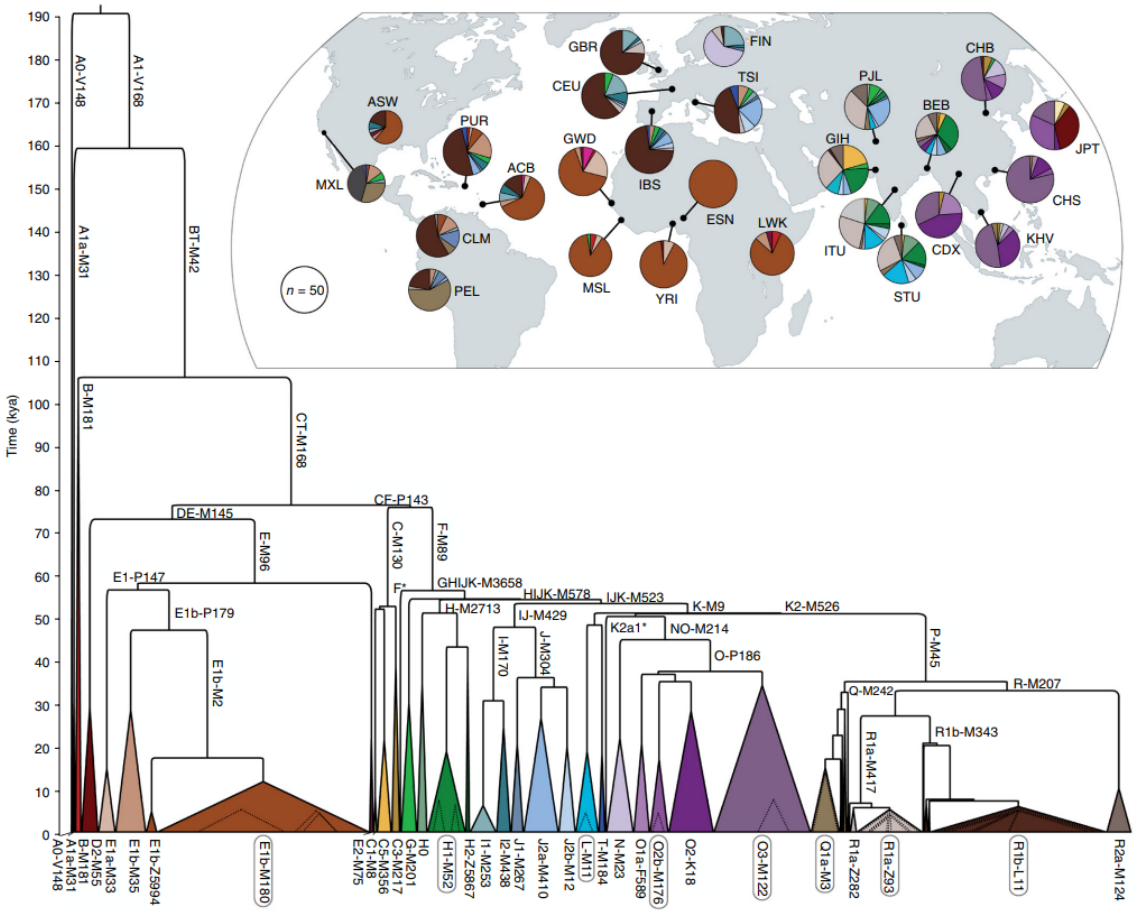


Figure 10: Phylogeographic tree of human Y chromosome with the continental affiliation of their highest occurrence. Sourced from Poznik et al. (2016).

1.3.4 The Phylogeographic Approach

Phylogeographic analyses combine phylogenies of uniparental markers (Lippold et al., 2014) with geography, in an attempt to allow inference and reconstruction of past demographic events and migrations (e.g. Soares et al. (2016)). In addition, if the molecular clock (Kimura, 1980) is brought into the equation, genetic distances can be converted into dates and provide age estimates for events such as the time of emergence of uniparental lineages, mitochondrial or Y chromosome (Karmin et al., 2015; Poznik et al., 2016), and sometimes their ancestral regions of origin (Figure 11). Interpretations based on individual uniparental lineages have to be treated with care since they cannot act as representatives of whole populations because they only account for a small fraction of the genetic diversity. Unless the sampling has been comprehensive and deep, most mtDNA datasets reflect just a subset of all maternal variation present in a population. The peopling of the Americas is an example for in which the use of uniparental markers have proved themselves a good tool to make demographic inferences about the past.

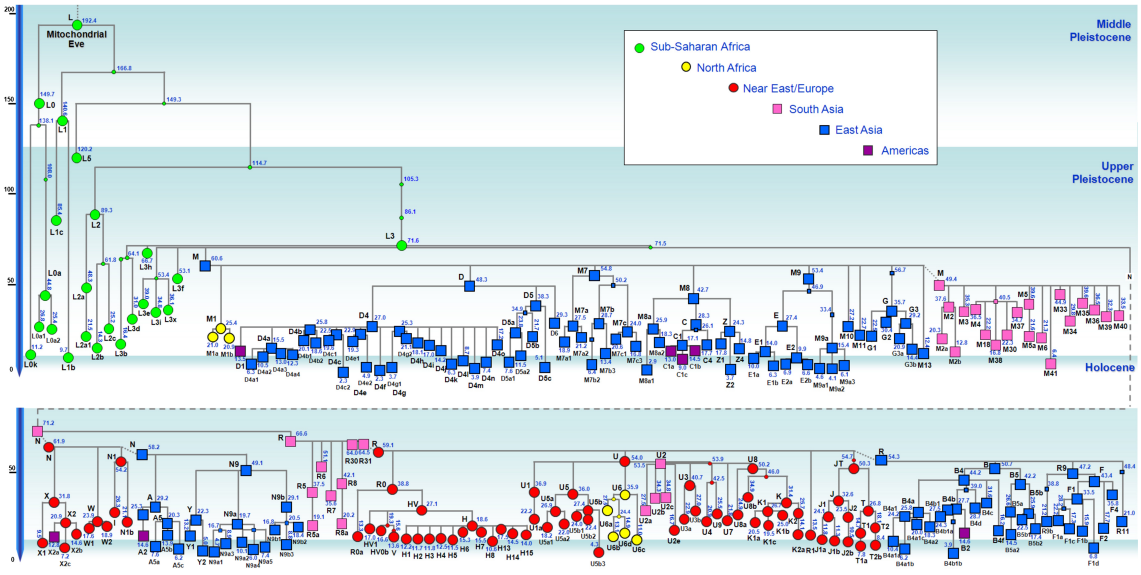


Figure 11: A phylogeographic tree showing all major mitochondrial lineages and their broad distributions in Africa, Asia and Europe. Taken from Soares et al. (2009).

Mitochondrial DNA and the MSY are exclusively transmitted maternally and paternally respectively and because of this, they are immensely helpful for research about past human migrations. The almost complete absence of recombination and negligible heteroplasmies facilitate interpretations. In this way we can reconstruct and trace phylogenies thousands of years back in time (Figures 10, 11). It is also possible to date the age at which different uniparental lineages appeared in combination with a calibrated molecular clock and a model of DNA evolution. The Jukes-Cantor model from 1969 (JC69) and the later contributions by Kimura (K81), Tamura, Nei, Hasekawa among others were the first attempts to provide the tools for this (Jukes and Cantor, 1969; Kimura, 1981; Soares et al., 2009).

Until the year 2000, these studies relied on mitochondrial haplogroups based on the hypervariable region and/or restriction site variation in the coding region, thanks to a manageable size, but the possibility of sequencing complete mitochondrial genomes has provided an improvement in the resolution that can be worked with (Behar and et al, 2012*a*), allowing better age estimations and inferences (Richards and Macaulay, 2001). Combining lineage-based approaches with analysis based on genomic ancient data, has confirmed patterns previously hinted at uniparental phylogenetic trees (e.g. Neolithic demic diffusion), but it has also revealed other invisible ones that may or may not have an echo in the uniparental markers (Patterson et al., 2012; Brotherton et al., 2013).

1.4 The Prehistory of Europe and Iberia

The genome wide composition of modern European populations can be broken down into three main ancestral components (Lazaridis et al., 2014): a European hunter-gatherer contribution from Mesolithic times; a Neolithic contribution from Asia Minor; and a Bronze Age contribution from the Pontic-Caspian steppe (Allentoft et al., 2015; Haak et al., 2015).

1.4.1 The Palaeolithic and Mesolithic

The first anatomically modern humans (AMH) to inhabit Europe were nomadic groups with a lifestyle based on a hunting and gathering (Figure 12). It is estimated that these groups arrived to the European peninsula around 45,000 years ago (Hoffecker et al., 2008). Although AMH are known to have interbred with Neanderthals in the Near East soon after leaving the African continent 60,000 years ago (Green et al., 2010; Fu et al., 2016), their arrival to Europe coincides with a sharp decline in Neanderthal presence in the archaeological record that eventually lead to their disappearance (Benazzi et al., 2011, 2015). Radiocarbon dates from archaeological remains found in Gibraltar suggest that the south of the Iberian Peninsula was the last Neanderthal stronghold (Finlayson et al., 2006; Zilhão et al., 2017). Chronologically this makes sense, since the southernmost tip of Iberia would be the last point to be reached by a population advancing by foot through the continent. However, dates of AMH arrival have been challenged (Cortés-Sánchez et al., 2019). The first evidence of Neanderthal occupation in Gibraltar stems from the finding of a skull, known as Gibraltar 1, by an officer of the British Royal Navy. The importance of the find was not fully understood at the time because Neanderthals had not been described as a species yet. Nevertheless, it draw enough attention at the time for Charles Darwin and Thomas Huxley to examine the skull. Upon examination of the skull they hypothesized that it must have belonged to an already extinct branch of the human evolutionary line (Darwin, 1871; Klein, 2009).

Following this successful colonization of continental Europe, modern human hunter-gatherers populated most of Europe (Fu et al., 2016; Kashuba et al., 2019) during the Paleolithic developing a number of different material cultures over time (Aurignacian, Gravettian, Epigravettian, Solutrean, and Magdalenian), some of them are regionally specific. Aurignacian is one of the oldest, dating to around 40 kya, while the Magdalenian is one of the more recent (18-12kya) and links the Upper Paleolithic with the Mesolithic. In Iberia, the only genomic data available for the Upper Palaeolithic comes from a Magdalenian individual (18.7kya) found in El Mirón (Ramales de la Victoria, Cantabria), a cave in northern Spain. Skull and long bones were missing from the burial that had traces of ochre and was covered by a stone. It was possible to recover genetic material from a toe bone and sequence the genome (Fu et al., 2016). The results revealed that she was a woman who carried a U5b mitochondrial lineage. The traces of ochre suggest she was painted in red at the time of burial and this is the reason why she is referred to as *the Red Lady of El Mirón*. The Solutrean (22-18kya) is a tool-making style intermediate in time that is characteristic of Iberia and France after replacing Gravettian lithic industry (33-22kya).

Around 25-21kya world temperatures decreased and the ice sheets covered most of Europe, this period is known as the Last Glacial Maximum (LGM). During this time, some human groups went extinct or retreated to territories in Europe not covered by the ice. The European hunter-gatherers survived the last Ice Age largely restricted in the climatic refuge of the southern peninsulas at the zenith of the ice sheets during the LGM. Iberia, Italy and the Balkans were these refugia. Here, humans, animals and plant species sheltered and later repopulated Europe when the ice started to melt (Gamble et al., 2004; Cardoso et al., 2011, 2013; Behar and et al, 2012*b*; Fu et al., 2016).

The reduction and fragmentation of suitable habitats during Last Glacial Maximum was an event that caused a major genetic bottleneck, reflected not only by the Mesolithic hunter-gatherer societies emerging at the end of the Ice Age (Posth et al., 2016), but also in many animal and plant species (Tallavaara et al., 2015). The groups that recolonized Europe from the southern refugia

had a narrow maternal diversity compared to pre-LGM times (Fu et al., 2016). Almost exclusively, all European Mesolithic individuals studied to date belong to mitochondrial haplogroup U5, either U5b or U5a (Bramanti et al., 2009; Posth et al., 2016). Mitochondrial lineages U5a and U5b exist in a gradient, U5a being most common in prehistoric eastern Europe and U5b more common in the west. The coalescence time for haplogroup U is consistent with the timing of the arrival of AHM into Europe (Richards et al., 2000), which is consistent with the idea that they were the first to introduce these lineages. On the paternal side of uniparental markers post-LGM, hunter-gatherers were carriers of lineages belonging to branch I, although there was more diversity prior to the LGM. Within this branch, haplogorup I2a is typically found in Mesolithic individuals.

From a population genetics perspective, European Mesolithic populations of hunter-gatherers can be separated in different sub-groups (labelled Western, Scandinavian and Eastern) because they display a degree of genomic differentiation following a soutwest-northeast gradient. Despite not being sedentary, this is likely due to limitations in mobility and the resulting isolation by distance (Sánchez-Quinto et al., 2012; Brace et al., 2019; Kashuba et al., 2019).

Along with a genetic bottleneck, the post-LGM period of recolonization of Europe also had a profound impact in the *modus vivendi* of hunter-gatherers. Uncharted landscapes and a warmer climate demanded adaptation to face new challenges (Figure 12). For example, the post-LGM period witnessed the extinction of the megafauna (Barnosky et al., 2004), probably linked to over-exploitation due to the better conditions for humans to live and hunt, and therefore precipitated the end of big-game hunting and the shift to small-game and aquatic resources.

The study of Mesolithic individuals from Europe through ancient DNA has also revealed some previously unsuspected physical features among European hunter-gatherers. The idea that western hunter-gatherers could have phenotypes combining dark skin and light eye colour has gained momentum following works by Olalde et al. (2014); Mathieson et al. (2015); Brace et al. (2019); Jensen et al. (2019). This is a combination of traits previously thought not to have existed at the time. The Mesolithic individuals LaBraña1 (7800 years) from LaBraña-Arintero (Spain) and the Cheddar Man (9000 years) from Cheddar Gorge (England) have become iconic examples of the hunter-gatherer population thanks to their facial recreations (Brace et al., 2019).

Other hunter-gatherer groups also existed in the fringes of Europe. These populations were genetically distinct from what we have referred to here as European hunter-gatherers and will play a role in later genetic re-shaping events of Europe (Pinhasi et al., 2012).

One of these groups inhabited the extended region around the Caucasus mountains. This is why this new group is commonly referred to as Caucasus hunter-gatherers (CHG), although the genetic cluster originated among prehistoric inhabitants of the Iranian plateau that later expanded north of the Caucasus. The genetic makeup of CHG will impact the genetic pool of European populations in the coming millenia through migrant Bronze Age steppe herders (Jones et al., 2015; Haak et al., 2015).

Geographically close but genetically very differentiated to CHG were the Natufians. This archaeological culture was discovered and characterized in the 1930's by Dorothy Garrod, a British archaeologist working in the Levant at the time. The Natufian hunter-gatherers of the Levant were a semi-sedentary population from the Near East that existed between 13kya to 7.5kya. They are considered the putative ancestors of the first Neolithic communities in the Levant. Based on data from ice cores from Greenland, the apparition of the Natufian culture coincides in time with the Bølling-Allerød warming, a period of sudden rise in temperature and moisture that happened towards the end of the LGM, just before the Younger Dryas (Platt et al., 2017).

Another group important for understanding the genetic diversity in Europe was the human population associated to the Iberomaurusian lithic industry in North Africa. The Iberomaurusian culture appeared during the LGM (25-22kya) spreading across Morocco, Algeria, Tunisia, and Libya, and disappeared by 11kya. From the limited genomic data available, a group of Iberomaurusian individuals from Taforalt (Morocco) (van de Loosdrecht et al., 2018; Fregel et al., 2018) appears to be genetically differentiated hunter-gatherer group with affinities to Natufians and a

lost sub-Saharan group. Their genetic makeup is still present in native North African populations, commonly referred as Amazigh folk or Berbers, and would also impact Europe in later times.

Figure 12: Levantine Rock Art depictions of hunting and honey gathering scenes from Cova dels Cavalls (top left) and Cova de la Aranya (top right) and its distribution in Spain (bottom).

1.4.2 The Neolithic

The transition from the Mesolithic to the Neolithic represented a great revolution in human prehistory (Alday Ruiz, 2009). The advent of the Neolithic is characterized by the adoption of agriculture as the primary subsistence strategy and use of pottery by groups that also became sedentary. The first peoples to adopt this innovation were communities in the coastal Near East with genetic affinities to the Natufians, who then partially influenced Anatolia gentically (Haak et al., 2010; Lazaridis et al., 2016; Kılınç et al., 2016; Feldman et al., 2019). In Iran, groups related to CHG started to domesticate goats and also adopted agriculture (Broushaki et al., 2016) via diffusion of ideas across the two regions, but without genetic mixing initially (Lazaridis et al., 2016; Reich, 2018). Development of agriculture in the region of the Fertile Crescent happened without much genetic interaction between two genetically distinct groups. However, later in the Neolithic period there was Levantine contribution to Iranian farmers and *vice versa* (Shinde et al., 2019). These new farming communities also started to domesticate other animals, such as aurochs, as the predecessors to modern cattle (Verdugo et al., 2019).

The Neolithic revolution is thought to have started in the Near East 12,000 years ago, but in Europe appeared later (Figure 13), since it took time for the new technology to spread to Anatolia and from there into Europe (Battaglia et al., 2009; Lazaridis et al., 2016; Feldman et al., 2019). The expansion of Neolithic groups can be explained thanks to the socioeconomic innovation that the development of agriculture brought. In biological terms, what the agricultural production system did was to raise the ceiling of maximal load of the carrying capacity in farming populations, therefore allowing an unprecedented population growth. Increasing population densities caused Neolithic communities to spread in all directions (Racimo, Woodbridge, Fyfe, Sikora, Sjögren, Kristiansen and Linden, 2020; Gamba et al., 2012; Feldman et al., 2019). Pioneer groups of Levantine farmers migrated towards Europe and Iranian farmers expanded further into the Asiatic steppes and the Indian Subcontinent (Silva et al., 2019; Narasimhan et al., 2019; Shinde et al., 2019). The Neolithic expansion was slow; the rate of expansion is estimated at 1 kilometer/year (Gangal et al., 2014; Isern et al., 2017) and likely unintentional, only driven by the need for new land to cultivate and to accommodate the growing number of people (Figure 14).

Since the beginning and inherited from the archaeological debate, there were conflicting ideas about how the spread happened. Some authors proposed a small contribution and others defended a significant influence of a Neolithic migration (Richards, 2003; Chyleński et al., 2017; Furtwängler et al., 2020). The views are based on two antagonistic hypotheses about the Neolithic; the Demic Diffusion model (DDM) and the Cultural Diffusion Model (CDM) (Diamond and Bellwood, 2003), which can be traced back to Cavalli-Sforza theories. However, in practice both models are oversimplifications because they describe very specific colonization models, and neither fits the reality as it happened. The reality was messier and likely more driven by deliberate directional pioneer colonisation towards favourable areas coupled with other local dynamics (Gamba et al., 2012; Isern et al., 2017). Nevertheless, it is worth explaining the two radical opposing views of this particular archaeological debate.

The CDM argues that the diffusion of the Neolithic was merely a transmission of ideas rather than people. The new life style arrived in Europe and was assimilated by the local hunter-gatherer populations who fully adopted agriculture towards the Middle Neolithic (Brandt et al., 2013; Haak et al., 2010) without genetic interaction, much like how agriculture developed in the Fertile Crescent amongst different groups. The DDM argues that the spread of the Neolithic cultural package was linked to migrations of farmers from Anatolia to Europe via the Balkans from where it split into two routes; a land route and a sea route (Olalde et al., 2015). The land route stretched from the Balkans to Central Europe. These pioneering early farmers settled in the new territories (Richards et al., 2000; Currat and Excoffier, 2005; Soares et al., 2010; Olalde et al., 2015; Isern et al., 2017). On the other hand, agriculture reached the southwestern part of Europe via the Mediterranean sea.

In Iberia the Neolithic began around 7500 years ago, ending around 5000 years BP. Note that there is never a clear boundary because change is not homogeneous (García-Martínez de Lagrán

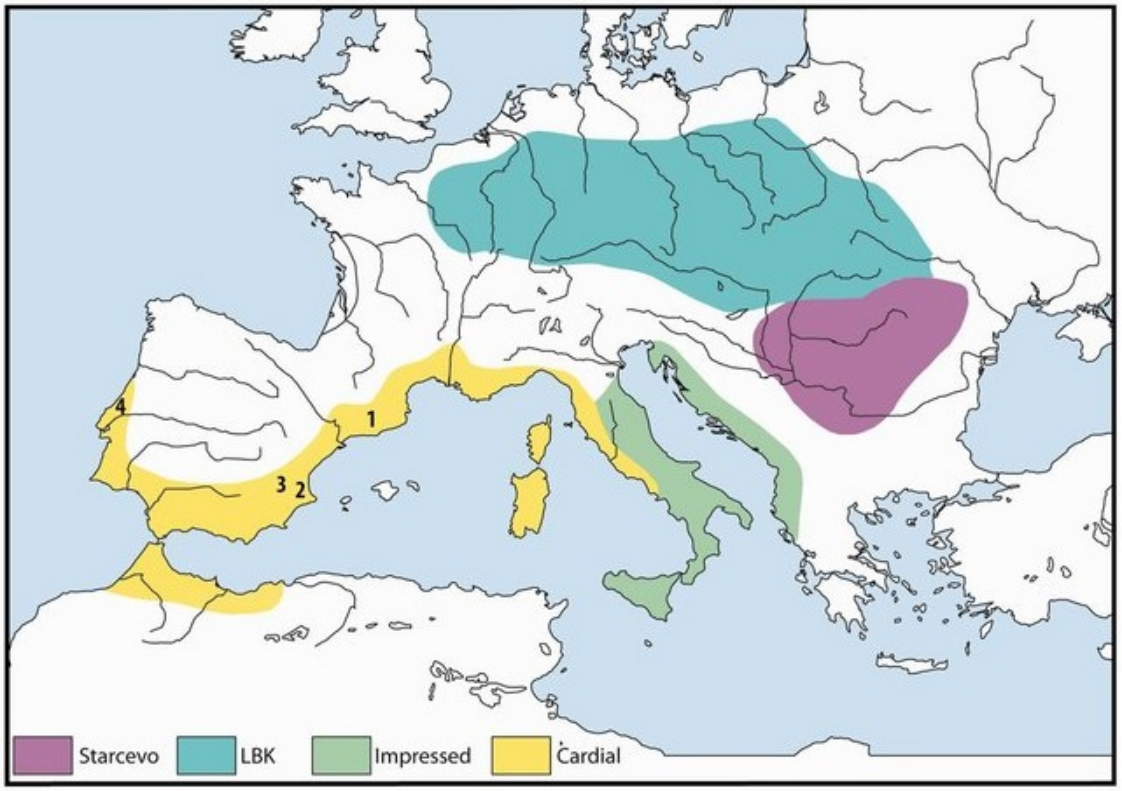


Figure 13: Distribution of the different major pottery styles and ceramic cultures that existed during the Neolithic in Europe. Taken from Olalde et al. (2014).

et al., 2018). Different locations within the same area might have been subjected to variable rates of innovation, even with coexisting groups with opposing subsistence strategies, namely foragers and agriculturalists. These changes allowed human populations to become sedentary and thus the foraging life-style was gradually abandoned. Mitochondrial data, genomic results and archaeology seem to fit better with a scenario closer to a pioneer colonization model (Gamba et al., 2012). Iberia in particular was reached and colonized by sea voyagers (Figure 14). These incoming farmers spread by establishing coastal settlements along the coast while sea voyaging in a leapfrog manner (Bernabeu-Auban, J. and Barton, C.M. and Pardo-Gordo, S. and Bergin, S.M., 2015; Isern et al., 2017). This is consistent with the idea that there were two routes with a common origin of Neolithic expansion into Europe from the Near East (Currat and Excoffier, 2005), Anatolia and the Balkans. The land route into Central Europe represented by the LBK material culture, and the Mediterranean route via navigation, represented by the Cardial pottery culture (Olalde et al., 2015, 2019).

The population dynamics are not always simple and there are other theories about the expansion and exceptional cases where there was no genetic input by farmers from Anatolia. Such is the case for the Baltic region (Saag et al., 2017). In any case, for the most part it is accepted that the arrival of agriculture to Europe during the Neolithic was linked to substantial migrations of people from the Fertile Crescent via Anatolia, the Balkans and the Mediterranean Sea (Figure 13). These early farmers settled in the locations and lived independently from indigenous foragers during the early stage of the Neolithic. It is only by the Middle and Late Neolithic when we start seeing significant admixture between farmers and hunter-gatherers, in the fashion of local women and men joining farming groups that practiced female exogamy as suggested by genetic data and uniparental markers (Goldberg et al., 2016; Knipper et al., 2017; Fernandes et al., n.d.; Schroeder et al., 2019; Furtwängler et al., 2020). Although in more complex Megalithic societies exceptions have been found (Cassidy et al., 2020).

Mitochondrial macrohaplogroups like K, J, T and H appear for the first time in western Europe during the Neolithic. However hunter-gatherer men and their Y chromosome markers were also incorporated to the resulting mixed societies. By the the late Neolithic the Y-chromosome diversity consisted mostly of I2a lineages of local hunter-gatherer origin and G2a lineages introduced by the farmers. The Tyrolean Iceman, Ötzi, is a natural mummy of a male individual found in the Italian Alps dated to about 3300 BCE. He was a carrier of an extinct or almost extinct mitochondrial K1 lineage and his Y chromosome lineage belongs to the G2a branch. Genomically he resembled modern Sardinians which are an isolated population that has changed little since Neolithic times. Both from the uniparental marker and the genomic point of view, Ötzi represents the epitome of a European Neolithic genome (Ermini et al., 2008; Keller et al., 2012; Sikora et al., 2014; O'Sullivan et al., 2016; Lugli et al., 2017).

The archaeologist Marija Gimbutas, who coined the Kurgan hypothesis, defended the idea that prehistoric Neolithic groups in Europe were peaceful and egalitarian. In her books such as *The Goddesses and Gods of Old Europe* (1974), *The Language of the Goddess* (1989), *The Civilization of the Goddess* (1991), she introduced the idea that Neolithic societies were matriarchal or female-centered with important female deities in close connection with the land. This idea about the Neolithic population of the Balkans came to be known as *Old Europe*, although sometimes can be loosely applied to an extended area. She also proposed the Kurgan hypothesis which suggested that *Old Europe* was conquered by proto-Indo-European speaking invaders from the Pontic-Caspian steppe (Mallory, 1991). These Indo-European migrants brought with them a male-dominated culture of horse-riding warriors that imposed patriarchy and war as basic elements of the new lifestyle (Anthony, 2010; Negrete, 2009). Ancient DNA can say little about how matriarchal a society was, but as we shall see below, her Kurgan hypothesis is now backed by genetic and archaeological evidence (Anthony and Ringe, 2015; Haak et al., 2015; Batini et al., 2015, 2017).

Figure 14: Neolithic and Megalithic areas of influence. Sourced from Atlas Nacional de España.

1.4.3 The Chalcolithic and Bronze Age

Archaeological evidence suggests that the story of metallurgy in West Eurasia starts in the region of Anatolia at some time during the eighth millennium BCE. The Çayönü Tepesi archaeological site in Turkey has been very important in regards to the issue of archaeometallurgy. In Mesopotamia and the Iranian Plateau, findings are less abundant and date to the seventh or sixth millennium BCE. However, what these regions have in common some form of early exploitation of native sources of copper. To the rest of the Mediterranean metallurgy spreads only after about 5500 BCE (Ruiz, 1993; Murillo-Barroso and Montero-Ruiz, 2012).

In Iberia the transition to the first of the Metal Ages, the Chalcolithic or Copper Age (CA), only started to develop around 5000 years BP but was independent of that in Anatolia (Murillo-Barroso and Montero-Ruiz, 2012). Iberia is a very copper-rich region, with deposits in the southwest (Rio Tinto), in the north (Asturias and Leon), in the southeast (Los Millares site in Almeria) and in the central Meseta.

The Chalcolithic is a period that is sometimes assimilated into the early Bronze Age because it did not developed in all regions in Europe. As mentioned above, in the Fertile Crescent, and unlike later bronze based civilizations, copper technology is linked to local metal deposits and mines, so it could only develop *in situ*. On the other hand, by the time of the Bronze Age, extensive trade networks had developed and allowed for the transport of raw materials to be worked elsewhere which eliminated the need for local deposits to be available. This allowed the Bronze technology to be fairly homogeneous across large parts of Europe. The complexity of trade networks (Ben-Tor, 2011) in Bronze Age Mediterranean is well exemplified in the book *1177 B.C.: The Year Civilization Collapsed* by Eric C. Cline (Cline, 2014) where it is explained how Egyptians, Hittites, Canaanites, Cypriots, Minoans, Mycenaeans, Assyrians and Babylonians were connected and how the disruption of these networks by the Sea Peoples and other natural events caused entire civilizations to collapse.

Besides the start of use of copper for tool making, in Iberia at least, there is strong cultural and genetic continuity with the populations of the preceding Neolithic. There are nevertheless, some hints of social differentiation. For example, the origin of the Bell Beakers phenomenon can be traced to Iberia, and the use of this type of pottery is linked to prestige individuals (Harrison, 1974; Doce, 2006). Los Millares is the most iconic example of a Chalcolithic society in Spain. The settlement of Los Millares had Bell Beaker pottery, influences of the Megalithic culture (Figure 14) and was fortified. The need for a fortification probably implies the existence of at least sporadic raids against the settlement. This is not unexpected, since these cultures had started to accumulate wealth in the form of resources like copper. The necropolis of Los Millares, located outside the settlement, had several tholoi (a type of large and sophisticated burial construction) that were dedicated to multiple inhumations. Each of the 70-80 tholoi contained between 20 and 100 individuals.

Bell Beaker pottery was a western European phenomenon characteristic of European post-Neolithic times. This material culture stared to appear by the end of the Late Neolithic and spread across Iberia and Europe. It is not considered a culture *per se* but there has been some attempts to interpret it as a form of incipient ancient religion (Harrison, 1974). Bell Beaker burials are identified by the styles of their pottery and other elements of their grave goods package. The name is derived from the vessels in the shape of an inverted bell that they used and that are very common in the burials (Figure 15). The key features of the Bell Beaker pottery are the bell-shaped body with a wide neck and profuse decoration with an incision technique. Geometric motifs are typical, made with lines and zigzags arranged in horizontal bands from the top to bottom of the beaker (Rojo Guerra et al., 2005; Alonso-Fernandez, C and Jimenez-Echevarria, 2015).

It is widely, although not universally, agreed that the earliest evidence based on archaeological remains point to an origin towards the end of the third millennia BCE (4900–4800 BP) in the southwest of the Iberian Peninsula. The Bell Beaker folk worked with metals, since many metal object can be found as part of the grave goods (Fitzpatrick, 2011), and the phenomenon quickly

spread thanks to the rich copper resources of the region (Vander Linden, 2007). Nevertheless, some alternative origins have been placed in The Netherlands and Hungary (Jeunesse, 2015).

The spread of the Bell Beaker package over Europe began roughly around 4500 years BP and its disappearance from the stratigraphic record occurs by 3700 years BP (Harrison, 1974). Before the appearance of the Beaker folk, there was a diversity of burial types and ceramics in the adjoining regions and some persisted after the consolidation of the Bell Beakers. In fact, the Megalithic culture which is of a middle Neolithic tradition coexisted and peaked with the Bell Beaker phenomenon across the Atlantic façade Schulz Paulsson (2019); Sánchez-Quinto et al. (2012); Cassidy et al. (2020), and other parts of Europe where the Bell Beaker phenomenon also spread (Figure 15). The Megalithic culture, however, is almost non-existent in the Mediterranean sphere of Iberia which had been previously dominated by the artistic expression of Levantine Rock Art. However, the Bell Beaker influence is present in Mediterranean regions like Sardinia and Sicily (Olalde et al., 2018). Bell Beakers have been the subject of a long debate because it was unclear whether they were a homogeneous entity genetically and/or linguistically, a pan-European multicultural trade network or just a domestic pottery fashion among ruling elites. Most information about the Bell Beaker expansion throughout Europe derived from the archaeological pottery culture, found in funerary contexts from Iberia, western Mediterranean islands, France, Benelux, British and Irish isles, and Central Europe (Vander Linden, 2007; Olalde et al., 2018) (Figure 15). The vessel and funerary goods indicate cultural affinities between the groups, and their presence is also related to the diffusion of the copper metallurgy throughout Western Europe.

A great number Bell Beaker burials have now been genetically characterized and their affinities unravelled (Szécsényi-Nagy et al., 2017; Olalde et al., 2018). It is clear now, that the Bell Beaker phenomenon was very complex and many factors and interactions influenced its spread, moving first out of Iberia and later into the Atlantic archipelago via the Netherlands (Olalde et al., 2018). Like the culture of the Megaliths that links Neolithic genomes from Spain to Ireland (Cassidy et al., 2016), it lasted almost a millennium and influenced large areas in Atlantic territories and Central Europe. However, claims that the Beaker folk spread was linked to a population expansion throughout western Europe from Iberia do not find support of aDNA based on the publication by Olalde et al. (2018). No evidence was found that significant contributions of Iberian related ancestry to Bell Beaker individuals in Central Europe. They found however, evidence for the opposite, some Iberian Beaker individuals display affinities with Central European samples that carry the new genetic component that arrived from the Pontic steppe. This indicates that the movement of people actually happened in the reverse direction, from the continent to the Iberian peninsula. Despite that the hypothesized migration from Iberia does not have an echo in the genomic data, these results suggest that road networks established during this time played a role modelling the current European genetic pool (Brotherton et al., 2013; Olalde et al., 2018, 2019) by facilitating the spread of Steppe migrants and their associated ancestry (de Barros Damgaard and et al, 2018*a*).

It is in the time of the Corded Ware culture during the European Copper Age that significant changes start to surface and when the modern European genetic pool is formed 16. Culturally, the new era is characterized by an increasing complexity of the economy, increasing social stratification (García Sanjuán, 1999) and the appearance of ruling elites based on the control of resources (Murillo-Barroso and Montero-Ruiz, 2012). The shift is best reflected in the abandonment of settlements, change of pottery and the burial styles. Burial customs change from multiple inhumations outside settlements in megalithic-like structures or caves, to individual burials protected inside the cities. This is likely due to existence of social elites and the desire to protect the valuable grave goods. The old abandoned settlements are replaced by new ones but not necessarily in the same locations. For example, Los Millares disappears and gets replaced by newly founded cities and societies like El Argar (Gilman, 1976) in which metallurgy was very relevant but with a bronze technology. The basic agricultural system shifted towards more specialized jobs, such as intense exploitation of metal deposits (copper, silver, gold) (Kristiansen et al., 2017), as shown

by evidence of environmental contamination in Iberia, which supports existence of mines in those times. Warfare becomes obvious for the first time in the archaeological record (e.g. the Tollense battlefield, proliferation of bronze swords, etc.) (Jantzen et al., 2011).

The introduction of these new customs is probably best reflected by the iconography of the Mycenaean material culture, much more war-like than its predecessors in Crete, the Minoans. But it is perhaps the tales and lore referring to this period that have survived until our days that best reveal how human societies had morphed into something more relatable to the thinking of our present day and age. After all, Homer's epic poem of the Trojan War in *The Iliad* is the first depiction of highly organized warfare and it is set during the Bronze Age, although it is likely to have been written later in the early Iron Age which explains the anachronisms.

Another example is the antagonism between the Titans and the Olympians in Greek mythology, which could represent nothing but the struggle between the gods of *Old Europe* and the new Indo-European gods. Where Gaia represents the Neolithic traditions of *Old Europe* as a prominent mother deity closely linked to the Earth in opposition to Zeus, a sky god less connected to earthly matters. The idea is that this change in the belief system represents a shift in the priorities and ethos of society, from farming to less nature-dependent tangible problems (agriculture) and more male dominated societies ruled by issues like trade, horse-riding and warfare (Negrete, 2009; Anthony, 2010). This change of ideology during the Bronze Age did not occur without genetic mixing of culturally differentiated groups (Immel et al., 2020). The work by Haak et al. (2015) confirmed with ancient DNA the Kurgan hypothesis because a genetic change in the early Bronze Age in Europe is palpable. However, from the little genomic evidence there is available from the late Bronze Age in Greece, it does not seem to be a great genetic differentiation between individuals ascribed to the Minoan culture discovered by Arthur Evans and Mycenean individuals (Lazaridis et al., 2017) despite using completely different languages (written in the Linear A and Linear B scripts). Linear B, deciphered by the British architect and philologist Michael Ventris, is a form of proto-Greek, and therefore part of the Indo-European family language tree. Linear A on the other hand, is pre-Indo-European and has not been deciphered yet. Although there are a lot of elements of truth in Gimbutas's idea of *Old Europe*, perhaps violence was not completely absent in late Neolithic Europe. We have already mentioned that fortified settlements existed, and there is evidence that Ötzi suffered from interpersonal violence which resulted in his death.

The shift in the genetic pool and lifestyle occurred at different times across Europe (Brandt et al., 2013; Allentoft et al., 2015; Cassidy et al., 2016) but mostly during the 3000 BCE to 2000 BCE time window, and during this period two important episodes occurred. One of such events was the advent of the Yamnaya pastoralists arriving from the Pontic steppe (Figure 16). This is arguably the defining moment of the European Bronze Age because they are responsible for introducing the third major genomic component that made the modern European genetic pool, as mentioned above. Archaeological evidence shows that these herders from the steppe brought axes, horses, carts, and the early proto-Indo-European speech (Haak et al., 2015; Anthony and Ringe, 2015). A clear genetic break occurs in the archaeological horizon between pre and post Bell Beaker central Europe, giving way to novel Bronze cultures. The same trend was observed by studying Neolithic and Bronze Age genomes from Ireland (Cassidy et al., 2016), and later in a publication with dozens of genomes from the island of Great Britain (Olalde et al., 2018). However, such a change was not so immediate in Iberia during the Chalcolithic (Martiniano et al., 2017; Olalde et al., 2018), and it only becomes evident and widespread few centuries later. It appears that initially only Iberian Bell Beakers display steppe ancestry which is absent in non-Beaker burials, at least in the early stages (Figure 16).

Iberia became an important source of metals during the Bronze Age and continued to be so in later times (Gilman, 1976; McConnell et al., 2018). There was early mining activity in the Tagus estuary and it increased in the Chalcolithic. The region of the Iberian Pyrite Belt (in southern Portugal) is where Iberian metallurgy began in the first place. As mentioned above, archaeological records like the one from La Molina (Spain), show that there was mining-derived pollution in

northern Iberia around 5000 BP (Martínez Cortizas et al., 2016) but early Iberian metallurgy was more invested in tool-weapon making rather than accessories (Murillo-Barroso and Montero-Ruiz, 2012), although this changes later in the Bronze Age. During this period, many open settlements were abandoned or replaced by fortified villages which are not common before this time. Regarding funerary practices, cave burials are typically practised during the Chalcolithic, as opposed to open air circular inhumation wells which are more common in some early farmer communities. This period also saw the onset of large-scale trade, by land and sea (Jesse et al., 2011; Lacan et al., 2011). In opposition to the prolific archaeological evidence of metal working and material culture in the Bronze Age (Gilman, 1976), human remains became scarce due to changes in funerary practices because of a shift towards cremation, anticipating the Iron Age rituals. The adoption of Bronze culture is accompanied with the introduction of the steppe genomic component, which started to infiltrate during the Bell Beaker period. This is well exemplified by a dramatic turnover in the composition of Y-chromosome haplogroups. Typical male lineages of the Neolithic gave way to the previously unseen R1b-M269 (Olalde et al., 2018).

The origin of the Yamnaya or steppe component introduced into Europe at the beginning of the metal ages by migrants from the Pontic-Caspian region (Anthony, 2010) can be traced to samples of a population of eastern hunter-gatherers and Caucasus hunter-gatherers ancestry (Jones et al., 2015; Haak et al., 2015) (Figure 16).

Among other issues related to this period, we can count the Celtic from the West hypothesis that points to a correspondence between the distribution of ancient Indo-European languages (Bouckaert et al., 2012) in western Europe and the Atlantic Bronze Age culture (Koch and Cunliffe, 2013). This hypothesis challenges the traditional view of Celtic language and people arriving to Iberia in a later migration (Brandt et al., 2015). The substrate for the Iron Age Celtic identity and language would have spread with the Bell Beakers starting in southwest Iberia. This idea is based on the fact that the areas under Bell Beaker influence match the later Celtic speaking territories. However, this idea is hard to reconcile with the recent genetic results suggesting that Indo-European languages arrived with the Yamnaya from the Pontic steppe (Bouckaert et al., 2012; Haak et al., 2015; Lazaridis et al., 2017). Some academics like John Koch in his Celtic from the West hypothesis, locate the origin of Celtic in Ireland because it is a better fit based on funerary cist archaeological evidence and linguistics than the case of Iberia.

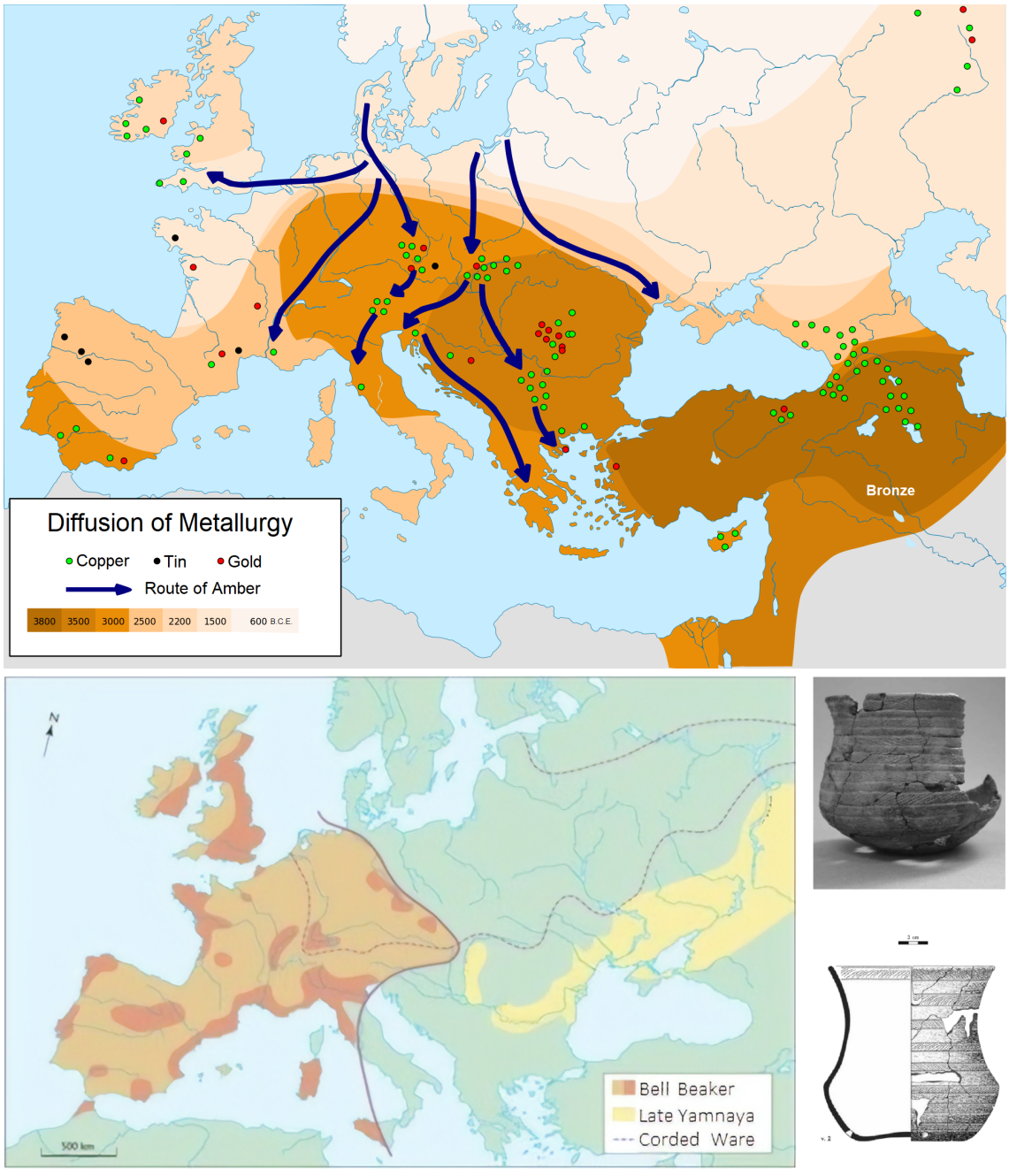


Figure 15: Diffusion of metallurgy across Europe (Top). Distribution and reconstruction of the iconic Bell Beakers. This beaker was found in one of the burials from the La Vital (Gandia) site as grave good. The bell beaker was part of the grave goods of the inhumations along with some metal objects, as typically seen in the bell beaker package. Source: Pérez Jordà et al. (2011) (Bottom).

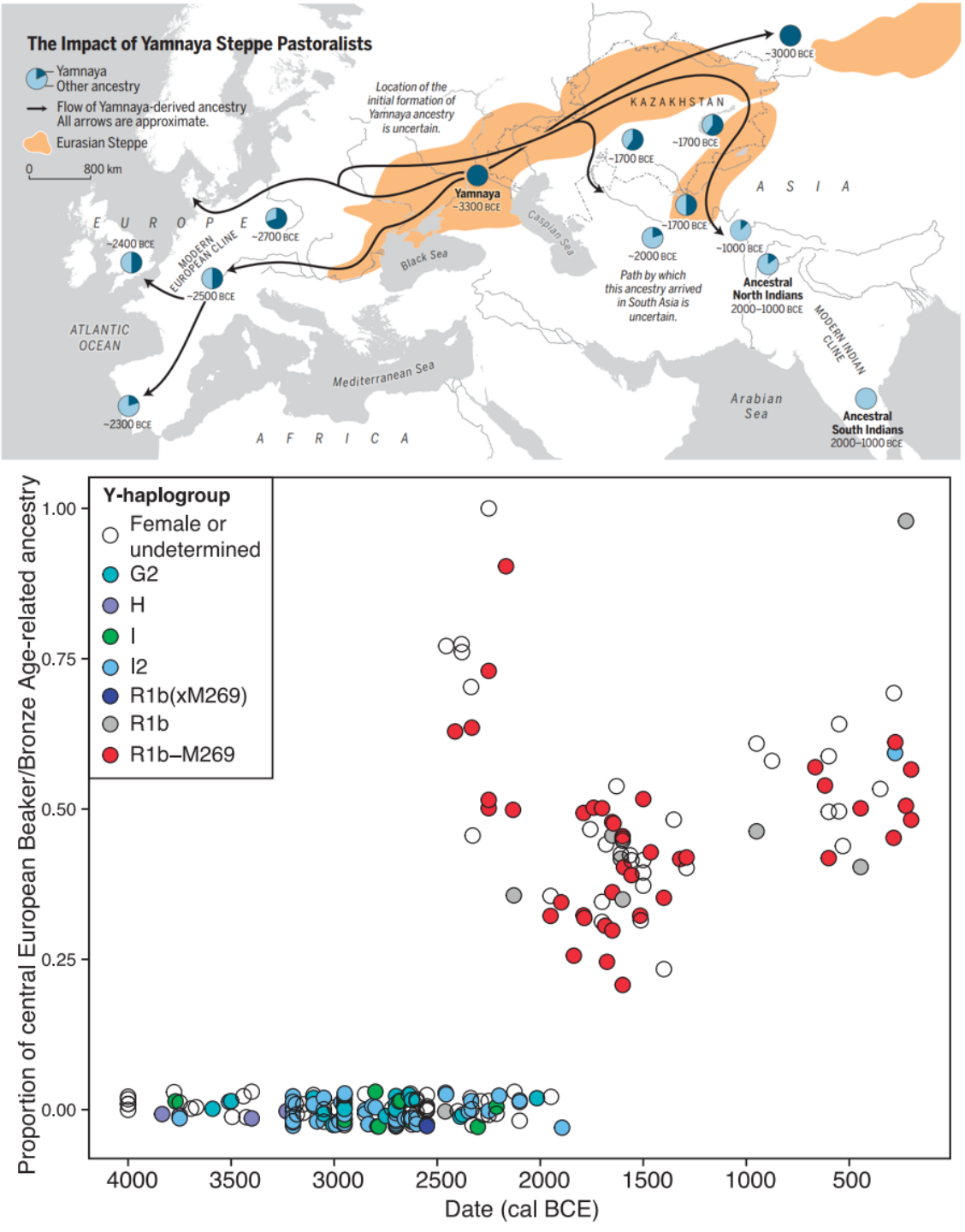


Figure 16: Schematic representation of the Indo-European expansion through migration out of the steppe of Yamnaya pastoralists. Taken from Narasimhan et al. (2019) (top). Amount of ancestry derived from central European Bronze Age populations in Middle Neolithic to Iron Age individuals from Iberia. Taken from Olalde et al. (2019) (bottom).

1.4.4 The Iberian Iron Age

In summary, the genetic evidence for the period previous to the Iron Age has shown that a genetic component arriving in Europe just prior to the Bronze Age is not widespread in samples of Chalcolithic Iberians from Spain and Portugal (Gómez-Sánchez et al., 2014; Martiniano et al., 2017; Olalde et al., 2018). This component only gradually enters Iberia towards the end of the Bell Beaker period along with the new technology of bronze. The migration was markedly male mediated and perhaps the late impact in Iberia has an echo in the survival of non-Indo-European languages (Haak et al., 2015) that characterized Iron Age Iberia (proto-Basque and the now-extinct Iberian language) (Figure 17) (Grau-Mira, 2019).

Figure 17: Examples of Iberian art and metal work. Top left: *Dama de Elche*, an Iberian art masterpiece. Top righ: *Guerrer de Moixent*, votive figurine yielding a *falcata*. Middle: detail of a *falcata*, the Iberian iconic sword that bears some resemblance to the Greek *kopis*. Bottom: lead plaque from Bastida de les Alcusses written in Iberian language. (image by Tautintanes, CC BY-SA 3.0.

Another question about the period concerns whether the intrusion of the Urnfield culture facilitated further genetic contribution from Central Europe. This issue is difficult to approach from the standpoint of genetics since the very trait that defines the Urnfield culture makes it impossible to study using aDNA: funerary cremation. Although, there are a few Iberian Iron Age genomes available (Núñez et al., 2016; Olalde et al., 2019) which actually show an increase in the steppe-related

ancestry, the sample size is currently too small to draw conclusions. This is relevant because there may have existed genetic differences between the cultural identities inhabiting Iberia before the Roman conquest, namely the Indo-European speakers and the Iberians which had a non-Indo-European language. This paints a somewhat complex scenario because the Urnfield tradition has a clear continental origin and it is the predecessor to the Celtic Hallstatt culture, but in Iberia it intrudes into the non-Indo-European-speaking area along the Mediterranean coast. This situation echoes previous archaeological debates. Was the spread of the Urnfield practice in Northwest Spain linked to ideas or movement of people?

The Iron Age was a time when population structure and languages arguably started to form groups that we recognize nowadays. In Iberia we encounter the Celtic-Iberian language duality but there were other cultural identities like Tartessos, the Phoenicians (Matisoo-Smith et al., 2018; Haber et al., 2017) and the Greeks (Figure 18). The exact nature of the genetic makeup of the Celtic populations in Iberia remains a mystery since most of Iron Age samples analysed in Olalde et al. (2019) belonged to the Iberian-speaking culture (Figure 17). Only three Celtic-speech associated individuals were genetically characterized in Olalde et al. (2019), two females and one male. The male does not carry an R1b haplogroup in the Y chromosome, and instead has a marker that indicates continuity with pre-existing paternal lineages.

Human remains from this period are scarce since cremation became a common practice already in the 3rd millennium BCE, and not only in Iberia. People abandoned burials in caves and started cremating the dead to bury the ashes and bones in urns. Such rituals are costly and only wealthy inhabitants must have been able to afford it. A remarkable exception to the cremation rule common in many Iron Age cultures across the whole of Europe can be found in the Iberian tribes of the Mediterranean edge. When infants died in their early childhood, they were not cremated. An anthropological explanation for this behaviour is that young children probably did not qualify yet to be considered members of the adult society and were subject to a different ritual treatment. Instead of being cremated, infant remains were placed in jars and buried underneath the households. The funerary process that the bodies received as to how they were placed inside the urn is unknown. It is common to find more than one infant in these jars; three, four and even up to five individuals together are not rare (Martínez Valle and Guérin Fockedey, 1987).

In the same spirit as the Bronze Age, Iron Age people lived in fortified settlements (*oppidia*) that controlled large areas of territory and often had other smaller and subsidiary settlements. The Iberian culture (Figure 17) can be considered truly urban and examples of these indigenous examples of early cities in the Mediterranean basin are La Bastida de les Alcusses (Moixent), Edeta/Tossal de Sant Miquel (Llíria), Arse (Sagunto), Ullastret (Girona) and Els Vilars (Arbeca). The main tribes that inhabited the East, descending from north to south were: the Ilergetes; Ilercavones; Edetanii; Olcades, and Contestanos among others. On the other hand, the hinterland and the west were populated by Tartessians, Celts, Celtiberians, Cantabros, Carpetanos and Turdetanos among other groups (Almagro Gorbea, 2004). The northeastern coast was more heavily influenced by Greek merchants as examplified by the colony of Ampuries due to its proximity to Massalia. On the other hand, the south and southeast coast were under Phoenician and later Punic influence, as exemplified by the colonies of Gadir (Cadiz) and Qart Hadasht/Cartago Nova (Cartagena) 18. The Phoenician influence can also be observed in the religious sphere with examples of syncretism and adoption of foreign deities such as Tanit, Astarte and the cult of Hercules (Aubet, 2001). These pan-Mediterranean influences shaped the Iberian identity in contrast with the Celtic interior and Atlantic side. The economy was trade-oriented because the southwest was a region of mineral wealth. Iberia was a mine of silver for the Phoenician capital cities of Tyre and Sidon. The contact of Iberians with these more advanced civilizations resulted in the development of their own writing system, which was adapted from the Phoenician alphabet. This reveals a heavy influence by Greek and Phoenicians maritime traders (Figure 17). In the later times, Iberia was a source of mercenaries for Carthage during the wars for the control of the Mediterranean against Greeks and later Romans (Ruiz Zapatero and Almagro Gorbea, 1992; Dominguez Monedero, 1996; Negrete,

2009).

One of the Iron Age cultures of Spain referred to above as Iberians or *Iberos* in Spanish, are known for speaking a non Indo-European language that survived until Roman times. However their language is yet to be deciphered. Recovered iconic material culture produced by Iberians includes falcata swords, small ex-voto metal figurines, and most famously the Dama de Elche, which might have Punic influences (Figure 17). Like many Iron Age cultures there has always been a halo of mysticism surrounding their origins and it is debated whether women retained a central role in a society surrounded by Indo-European influences (Currás and Sastre, 2019). This is something that is currently being addressed from the genetic standpoint with aDNA.

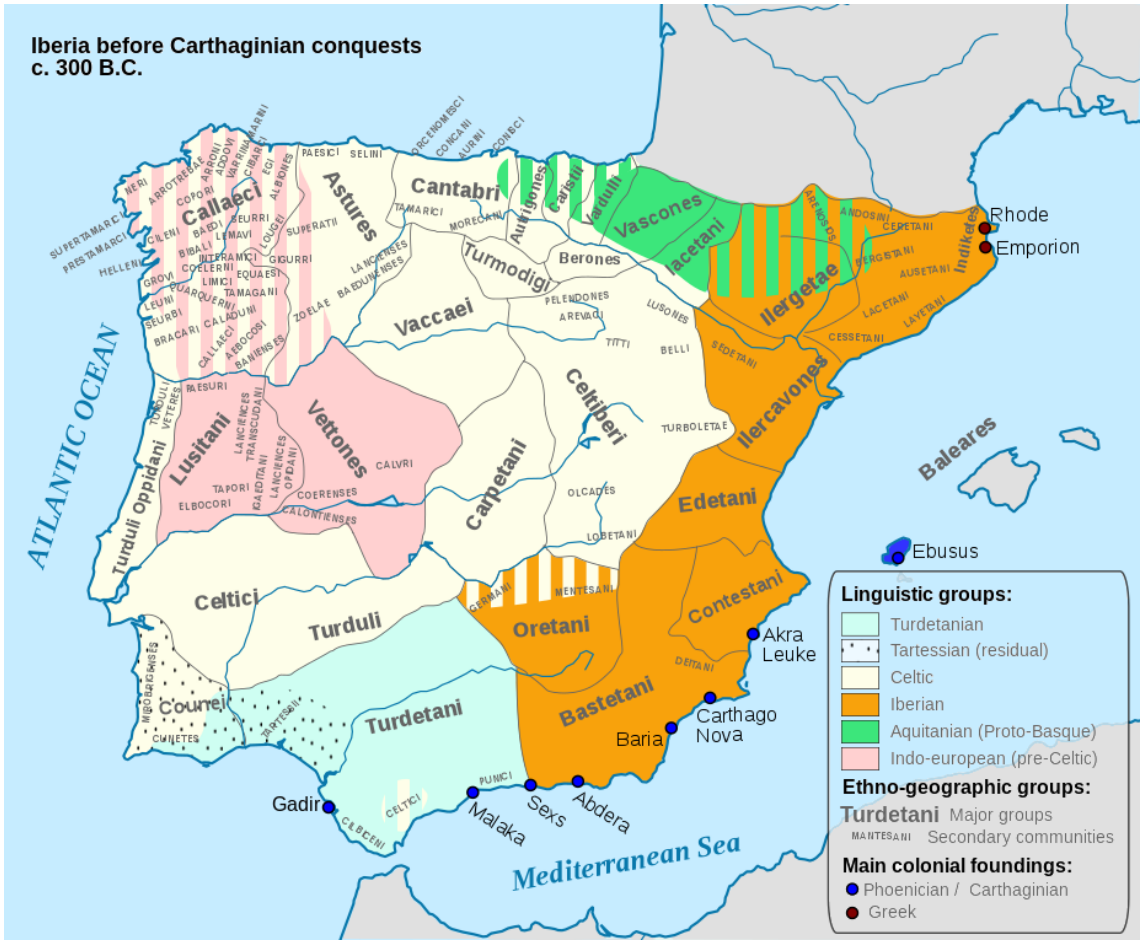


Figure 18: Ethnographic and Linguistic Map of the Iberian Peninsula at about 300 BCE (before the Carthaginian conquests). Based on the map done by Portuguese Archeologist Luís Fraga, from the "Campo Arqueológico de Tavira". The reference map can be found at this location. Author: Alcides Pinto, taken from Wikipedia. Alternatively, an identical map in lower resolution and in Spanish can be found in the book *The Roman Conquest of Hispania* by Javier Negrete.

1.5 The History of Mediterranean Iberia and the Making of a Modern City

The official arrival of the first Roman pro-consular armies to Iberia under the command of the brothers Publius and Gnaeus Cornelius Scipio in 218 BCE at the beggining of the Second Punic War, marks the start of a reliable recorded written history about Iberia. From this moment Iberia is no longer a land of mystery and myths, Prehistory ends and History starts (Figure 20). The wealth of information about Iberia, archaeological and otherwise, increases greatly compared to prior eras. It is precisely for this reason that is impossible that the sole work of a doctoral thesis could cover successfully the whole Peninsula. In order to keep coherence and to narrow the scope of the research, from Antiquity to Medieval times this introduction about the History of Iberia will be restricted to the East, or what is commonly referred to as the Iberian *Levante* or Valencian region. The territory intended to cover overlaps with the modern administrative provinces of Valencia and Castellón but not Alicante, and most of the information gravitates around the city of Valencia, which was founded as a Roman settlement (Figure 21).

Unraveling the genetic story behind the modern city of Valencia is one of the goals and compare to the surrounding its rural population. In a similar style to what the work by Antonio et al. (2019) did with the city of Rome. The project therefore, aims to follow the development of the rich and complex urban genetic history of the city since its foundation as Roman colony of *Valentia Edetanorum* until its consolidation in Islamic and Late Medieval times as *Balansiya*. This project is a unique opportunity to explore historical dynamics and population turnovers, of which we have written references of indirect evidence. The conclusions may be extrapolated to the rest of Iberia but such is not the aim of this work. Along with written records, we also have a rare gift in the shape of very a robust archaeological record for a city, since excavations though the years have found vestiges of the cultures that inhabited Valencia: Romans, Visigoths, the elite of Islamic rulers, as well as jews and common people from both Islamic and later Christian times (Ribera i Lacomba, 1998; García-Prósper and Polo-Cerdá, 2020).

1.5.1 Roman Hispania

After Phoenician influence over the south of Iberia disappeared around the 6th century BCE, the maritime supremacy of Carthage in the western Mediterranean allowed the Punic rule to naturally take over the former commercial domains of Tyrus. The territory under Punic influence in modern Spain was the South and the East. The river Ebro was the natural frontier to the north. The Greek influence remained in the east-northeast, best represented by the site of Emporion. Barter trade economy remained the rule until the time of the Second Punic War when coins become common under the Barcids (Figure 19) (García-Bellido, 2014). It is not until after this twenty-year-long war when local mints fully develop and indigenous names start to feature in the coins (Arse, Baskunes, Lauro, Barkeno and Kelin among others).

Following the Roman conquest of Mediterranean Hispania (Figure 20), the territorial structure of the Iberians collapses bringing the the end of long established urban centres (Cadiou, 2008). The Iberian cities start to be replaced politically and physically as urban centres by newly founded Roman colonies. The colony of Valentia is probably one of the best examples of this process. Its foundation in 138 BCE coincides with the decline of the Iberian city of Edeta in the nearby hinterland (Ribera i Lacomba, 1998, 2003; García-Prósper and Polo-Cerdá, 2020). The process of Romanization accelerated with the advent of the Empire which brings the completion of the conquest of Iberia as a whole. Romans developed a dense network of roads across a heterogeneous territory. The South and East assimilated Roman life style totally, were agricultural, and urbanized. The central plain, the West and North were less Roman and more pastoral (Butzer et al., 1985; Butzer, 1988).

Figure 19: Silver double shekel from Carthage, part of the Moixent Hoard, kept at the British Museum. On one side there is one of the Barcids (possibly Hamilcar Barca or Hasdrubal the Fair) represented with the atributes of Melqart (Punic equivalent of Hercules) and presumably a North African war elephant, species nowadays extict (top). The silver coins and pieces from the Moixent Hoard (Valencia, Spain), this treasure is dated to have been buried around 230 BCE, which coincides with the period of total domination of the region by the Barcids of Carthage. These foreing coins are an example of the very first monetary system that circulated in Iberia, and that whose introduction was there to stay (bottom) (García-Bellido, 2014).

Valentia: Birth, Fall and Rise of a City

The city of modern Valencia was founded as *Valentia Edetanorum* in the year 138 BCE, by the consul Decimus Junius Brutus Gallaicus as mentioned in the *Periochae* (a summary of the lost chapters of Ab Urbe condita by Livy) (Figure 21) (Ribera i Lacomba, 1998, 2003). The name *Valentia* roughly translates as *city of the braves*. Some sources point to the presence of former legionnaires among the settlers who were honorably discharged from the army after the recent Gallaecian War in northwest Spain and Portugal. However, this is not confirmed, and many versions about the original settlers exist, including defeated soldiers from the Lusitanian leader Viriatus rebel army. *Edetanorum* refers to the indigenous name of the land, a nearby settlement and its people (Edetanii tribe). The likely reason behind the colony was to gain direct control of the fertile area amidst the loyal native cities of Arse (besieged by Hannibal Barca a century earlier, and later known as Saguntum), Edeta and Saetabis. Those cities nowadays correspond to Sagunto, Lliria and Xativa respectively. Valencia was a typical Roman city since its conception, strategically

Figure 20: Stages of the advance (in green) of the Roman conquest of Hispania through the centuries (Artola Gallego, 1993). Years 218 BCE, 155-133 BCE, and circa 30 BCE from left to right.

located in a river island by the mouth of the river Turia.

Another advantage of the colony was that it was traversed by the Heraklean Way which later became the Via Augusta, and connected the South of Iberia with the Pyrenees. The debate of whether a previous Iberian settlement, identified as Tyris, existed or not in the surroundings during the 3rd Century BC was reignited by the excavations in Ruaya Street in 2008 (Albelda Borrás, 2015). There were even some early claims in the media of a the site being a military camp used by Hannibal's army on his march to the north but there is no solid archaeological evidence backing this, beyond the coincidence in dates of the material culture found. However, it is clear there was human activity in the area prior to the roman settlement but very little is known about it. Indeed, the earliest reference to the city in Livy's *Ab Urbe Condita (Periochae 55.4)* where he says: *"In Hispania, consul Junius Brutus gave land and a town, called Valentia, to those who had fought under Virtiathus"* (a rebel leader from the province of what is nowadays Portugal).

In the early moments of the city, as revealed by the different funerary traditions recovered from excavations of the oldest Necropolis, heterogeneous rituals coexisted. This is reflected in the indigenous-styled cremation urns as opposed to Roman inhumations. Roman burials are easily identified by the *tegulae* covering them (García-Prósper et al., 2003, 2010; García-Prósper, 2016). As the romanization advanced the two groups probably merged giving birth to the hispano-roman population.

There are plenty Roman inscriptions recovered from excavations in the City through the centuries, and some of them were found and are known since Medieval times (Pereira Menaut, 1979; Martínez Valle, 1998; Corell, 2012; Abascal Palazón and Cebrián Fernández, 2015). However, these Latin inscriptions have shed little light over the issues that interest most of the archaeological research and the questions surrounding the exact date of foundation of the city, the magistrate who ordered it, whether the first colonists were Romans or Lusitanians, and the circumstances that led to the concession of the status of colony so shortly after being founded. It is agreed that by the year 60 BCE Valentia had a colony status and was not a *peregrine* settlement. This is based on a inscription found in Asculum (Italy) dedicated to consul (in the year 60 BCE) Lucius Afranius by the colony of Valentia in Hispania. The status of the city at the moment of birth relies of the nature of its first inhabitants: if they were barbarians (meaning foreigners) then the settlements must have been considered a *peregrine* city (a foreging city, lesser status than a colony or an municipium, without rights to Roman or Latin law); if the first settlers were Roman citizens then there is room to make a case for Latin or Roman law rights.

The first inconsistency derives from the fact that Livy credits Junius Brutus (Grandfather of Marcus Junius Brutus, one of the leaders of the assassination of Julius Caesar) as the founder of the city in 138 BCE (Figure 21). At the time, Junius Brutus was the governor of the province of Hispania Ulterior, so it is hard to reconcile with the foundation of a city outside his jurisdiction in Hispania Citerior which is where Valentia was located. Scholars also find difficult to accept the fact that Junius Brutus granted such prosperous lands to former Lusitanian enemies. If this was indeed the case, it raises the question of how the city acquired colony status after only 70 years,

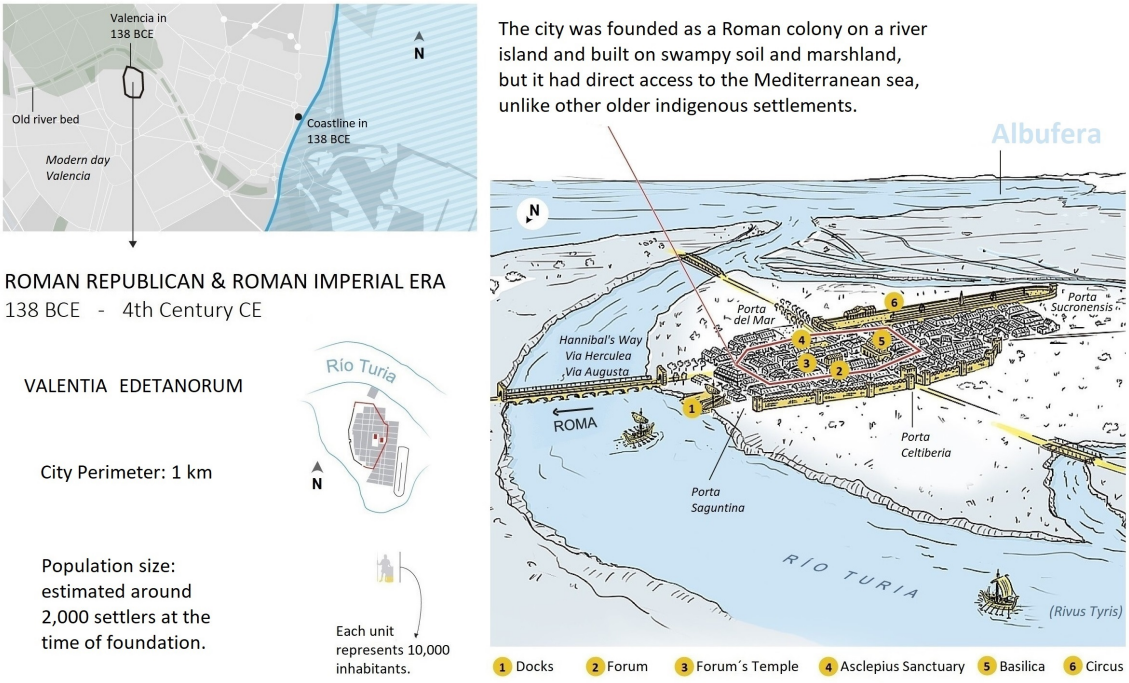


Figure 21: Infography depicting a reconstruction of the Roman colony of Valentia Edetanorum. The illustration was originally published in an article under the name *La Valencia por desenterrar* by *El Mundo* newspaper on the 28th of October 2018. The artist and illustrator is Pedro Jiménez (@PedroJimnez), and the material is accessible at *infografia-pedrojimenez.blogspot.com*. The image has been modified to translate and adapt the text with permission of the author.

since the founding settlers were non-Roman citizens. Another surprising fact is the evidence that Valentia started minting typical standard Roman coins very early on (Figure 22) (Ripollès i Alegre, 2002).

To reconcile Livy's report with all the inconsistencies that it implies, it has been proposed that the settlement of Lusitanians by Junius Brutus and the actual Roman foundation of the city are separate events, and that the summarizer of the Periochae got it wrong. However, Ribera i Lacomba, one of the most authorised voices on the topic, disagrees with the need for such explanations because, as it shall be further explained below, the city is destroyed in 75 BCE during the Sertorian War (Figure 22). Ribera i Lacomba thinks the degree of destruction was enough to urge Rome to repopulate the city with new Italian or Roman colonists (Ribera i Lacomba, 2008). Given the origin of the new inhabitants, the acquisition of colony status by 60 BCE is easy to explain since these new settlers had rights to Latin or Roman citizenship and not just *peregrine* status. This is in agreement with the timing of the Asculum inscription to Afranius, who fought and razed Valentia under Pompey during the Sertorian War. Judging from the inscription, it would seem that Afranius, a client of Pompey, not only had a role in the sack of Valentia but also in repopulating the city afterwards. The city is therefore reborn under the patronage of Pompey, likely to help expand his client network in Hispania.

As referenced above, following the Social War and shortly after Sulla's Civil Wars, Valentia finds itself involved in the Iberian branch of the conflict: the Sertorian War. For the city, the conflict culminated in the form of a sacking. The city was under control of general Quintus Sertorius who lead the last remnant of the *Populares* cause in Hispania. The Sertorian War was a byproduct of the first Roman Civil Wars waged between the supporters of Gaius Marius and Sulla, that immediately followed the Social War against the Italian allies. Sertorius was sympathetic with the *Populares* reformist current and had previous connections with the *Marian* faction.

Sulla's first march on Rome forced Marius into exile but after Sulla's departure to Asia in 87

BCE to join the First Mithridatic War (89-85 BCE), Marius is named consul by the Sanate and returns from his African exile. At this moment, Sertorius joins the *Marian* faction encouraged by the consul Cinna. Despite being a *populare*, Sertorius never was much appreciative of old Marius nor his behaviour after retaking Rome from Sulla. Sertorius went as far as retaliating against Marius army of freedmen army that had been carrying out the purges in Rome. During this brief reign of terror in a Rome controlled by Marius and Cinna, Sertorius took on a praetorship in Hispania (83 BCE). However, he found himself seeking exile refuge in North Africa when Sulla marched for a second time on Rome and finally took back control of the eternal city in 82 BCE. After returning to Iberia from Africa in 80 BCE invited by local tribes to fight against Roman power, Sertorious fiercely opposed the Sullan *Optimates* for ten years. Eventually, he is assassinated by his own men in 72 BCE. Only then could the *Optimates* finally regain full control over Hispania. However, the issues underlaying the conflict were never settled and Sulla's legacy and reforms will desintegrate within a generation. This fracture in the Roman elite, *Populares* and *Optimates*, was the ultimate cause of the fall of the Roman Republic, culminated with Julius Caesar's crossing of the Rubicon and the war against Pompey the Great and the Senate.

During the Sertorian War a young Pompey the Great is sent to Iberia (in 77 BCE) to defeat the long standing rebellion of Sertorius, who had created a parallel state in Hispania rallying the unhappy local elite to his side. According to classical sources, in the year 75 BCE, Valentia is sacked in the aftermath of a battle between Pompey and a Sertorian army lead by Gaius Herennius and Marcus Perpenna Vento (lieutenants of Setorius) in the strip of land between the river and the northern the wall of the city (Alapont Martín et al., 2009). The written evidence for this episode comes from Plutarch's *Parallel Lives* chapter about Pompey, where he wrote: *"Near Valentia Pompey crushed the generals Herennius and Perpenna, men of military experience among the refugees with Sertorius, and slew more than ten thousand of their men"*. In fact, Herennius is killed in the battle and it is believed that during the sack the city is destroyed (Figure 22), only the Temple of Asclepius and perhaps other religious buildings were spared (Alapont Martín et al., 2009). Upon hearing the news, Sertorius who was fighting Quintus Caecilius Metellus Pius in Hispania Ulterior returns to the coast to face off Pompey and Afranius. The subsequent battle happened near the river Sucro in the same year 75 BCE, south of Valentia. The sources indicate that Pompey's defeat was only avoided by the arrival of Metellus army, which forced Sertorius to retreat. Perpenna on the other hand, who managed to escape the defeat in Valentia became the instigator of the assassination of Sertorius three years later. However, the rule of Perpenna did not last long, as he was finally defeated by Pompey shortly after the power grab.

Archaeological proof of such destruction has also been found in the last decades in excavations throughout the ancient city nucleus, although nothing that proves the extent of Plutarch's claim of the killing of 10,000 people. In 1985 and 2002, excavations in the area of Plaza de l'Almoina found evidence of this layer of destruction along with the remains of at least 7 Roman legionnaires (Figure 22) (Alapont Martín et al., 2009). These soldiers died or were executed *in situ*, during or after the battle. The area of this site has long been identified as adjacent to the *forum* of the old Roman city. All of the soldier remains studied except one were young males in their 20's, whose limbs were traumatically amputated and affected by fire in some cases. This is interpreted as evidence of torture. Such extreme cruelty is not a total surprise, it is merely the trademark of a period marked by the climate of civil war and the purges and counter-purges of Marius and Sulla in their struggle for power in Rome. It is worth noting that the one of the victims found is a middle aged man who was tortured and impaled with a *pilum* (Figure 22). There is some degree of speculation among local archaeologists about the identity of this man, his age and the treatment received in defeat may indicate he was a high ranking officer. Hence, the individual is informally referred to by the name of the commander in chief known to have been killed in the fight.

The inscriptions do, however, reveal another important archaeological question regarding the early inhabitants of the city: the existence of two groups of citizens, the *old ones* and the *veteran ones*, as mentioned in an inscriptions recovered in excavations (*Valentini Veterani et Veteres*)

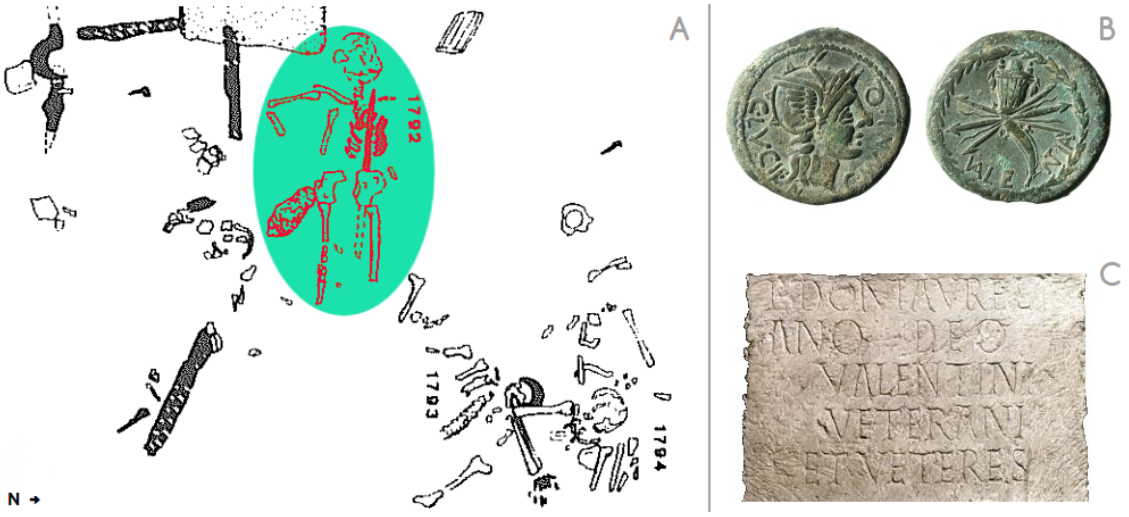


Figure 22: A) Archaeological context of the Sertorian soldier (UE1792) found tortured and impaled by a *pilum* in the Forum (modern Plaza de l'Almoina of Valencia, (Ribera i Lacomba et al., 2010)). The *pilum* can still be appreciated. Other badly preserved individuals can be observed here, as well as some weaponry. B) Roman as (a type of coin) minted in Valentia with the name of the magistrate in charge of the treasury of the city: C. Lucien(us) Muni(us) Q(uaestor). C) One of many Roman inscriptions co-commisioned by the two political bodies of Valentia: the *Veterani* and the *Veteres*. This one in particular is dedicated to Emperor Aurelian (3rd century CE).

(Figure 22). These two groups might represent pre and post destruction settlers but the question remains unclear (Pereira Menaut, 1979).

This duality is seen as the imprint of a second settlement in the city, of licensed Roman soldiers, which could explain the incongruity mentioned above that a city supposedly founded with the remains of the enemy army obtained the status of a colony Roman so soon (Casson and Wilson, 1966). Also, from that moment there was a separation between those who already lived there (the veterans, descendants of the first inhabitants, Lusitanians or not) and newly settled Roman army veterans. However, some voices argue that the inscriptions that mention the *Valentini Veterani et Veteres* cannot be taken as testimony of such an early second colonization, because the dates of a part of the inscriptions appear much later (1st century CE). The inscriptions indicate at least that there were decisions made by the two groups jointly, for example in dedications to emperors and their families the city appears as the dedicator. *Veterani and Veteres* behaved like a unit.

It is possible that the second settlement (*deductio*) was related to some disaster that made new citizens necessary to restart the city life. Such a catastrophe seems testified by one of the most important inscriptions from the 1st century CE found in Valencia. The text is a large block of limestone which was part of the frieze of a public temple with dedication to a divinity asking or thanking for his/her help. The word *clades* is preserved, which means calamity, misfortune, or catastrophe of any kind. The fact that such inscription from a public building has been found within the scope of the Forum implies that it refers to a local catastrophe, such as floods of the Turia river. Floods in Roman times are attested by layers of mud in Roman strata, as suggested by the illustrious Domingo Fletcher. However, it is not possible to exclude that the inscription refers to a general catastrophe in the Roman Empire. Perhaps, it may even refer to the destruction of the city by Pompey a century earlier? According to the quality of the writing of the inscription, it probably dates to the 1st century CE.

So whatever the length of the hiatus in the activity of the city after the events of 75 BCE, the dates of the inscriptions show that the city was already back to live by the time of Augustus. There are no more historical or important written references to Valentia corresponding to the Imperial, but the city must have been an important commercial hub. There is archaeological evidence of a

Circus, roof-tile wrokshops, a recently excavated perfume house, etc... reveal an important socio-economic activity. The location of several urban and periurban necropolis and another group of inscriptions draw a perimeter of 500m of radius around the Forum that corresponds roughly with the limits of the city in Imperial times. If we relate the character, origin of the inscriptions and the topography of Valentia, the most important findings are concentrated in the scope of the Forum; which confirms it as the centre of public life in the city. This area was the highest part of the old Valentia, now occupied by the Cathedral, the Almoina archaeological centre and adjoining places (Figure 23) (Ribera i Lacomba et al., 2010). Pedestals have been found with dedications to different emperors of the 1st to 3rd centuries CE. In fact, most of the inscriptions found in Valencia are from the Imperial period, which likely was its era of splendour as a Roman city.

This is in sharp contrast with next six centuries that will follow after the fall of the Western Roman empire. The crisis of the third century must have sparked a decay in the population, because many *latifundia* surrounding Valencia were abandoned during the fifth century (Butzer et al., 1985). The activity area of Valentia regressed from an area of 50 hectares during the 2nd century to about 15 during the 4th, as indicated by the walls of circa 300 CE (Jiménez Salvador et al., 2014). The lack of sources prevents us from knowing how badly the turbulences of the late Roman Empire affected Valentia, but the archaeological record of ceramic material from the foundational necropolis (C/Quart-Cañete) indicates that the city was active during the 1st century BCE and the necropolis continues to be used until at least the 3rd century CE in Roman tradition (García-Prósper et al., 2003, 2010; García-Prósper, 2016; García-Prósper and Polo-Cerdá, 2016, 2020).

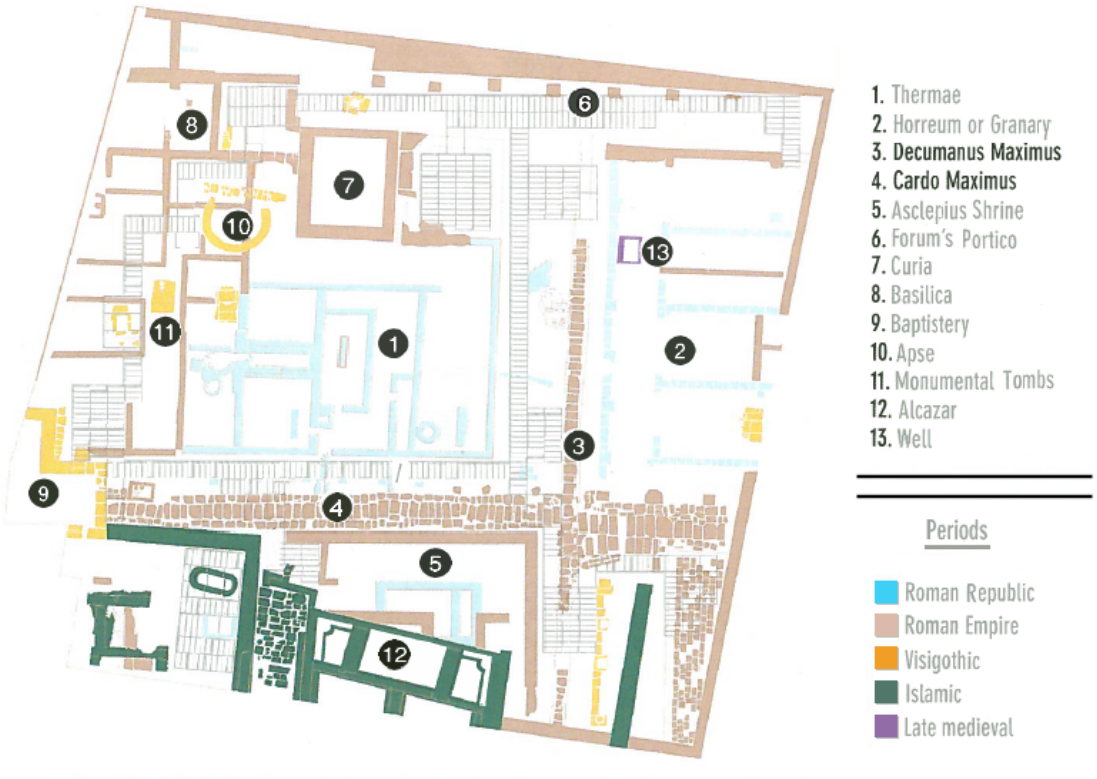


Figure 23: Current outlook of the archaeological site of l'Almoina. Highlighted in different colours are all the structures built by different cultures over the centuries. L'Almoina was originally the Forum of the early Roman city, later became the Citadel (al-Qasr or Alcazar) during the late Islamic rule. Interestingly, Romans, Visigoths and Islamic rulers used the location partially as cemetery. Today is a public square and archaeological centre. Source: SIAM archive (Ribera i Lacomba et al., 2010).

1.5.2 Visigothic Spania

Visigoths were a branch of the Gothic people, but any further knowledge about their origin beyond this statement becomes cloudy, like for most of the Germanic tribes of the Migration Period. Their name (West Goths) is a mere Roman classification void of any ethnic meaning, used for practical reasons to distinguish them from their Gothic cousins in East Europe (Ostrogoths). This makes the issue of understanding their genetic origins challenging. Not only the original homeland of the Goths is uncertain, although Götaland (south of Sweden) is generally assumed to be the place (Augustyn et al., 2020). From there, Goths first migrated to Poland and then scattered across Europe. During the barbarian invasions of the Migration Period, the Goths split into different groups. This period of instability that lasted more or less from the 2nd to the 6th century saw the arrival of new tribes that took over the areas of the Empire that were no longer under direct control of Rome. These movements of people were so important that even some territories were renamed after the newcomers. This is not the case of Iberia where Vandals, Alans, Suevii and Visigoths invaded, but for example France, formerly Gaul, derives its names from the Franks that arrived during this time, and same applies for England, formerly Britannia, whose name is derived from the Anglo-Saxons invaders (Bogucki and Crabtree, 2004).

Figure 24: Map of the consolidation Visigothic kingdom and military campaigns (red arrows), extent of the Suevii kingdom until 585 CE (in green), northern rebellious tribes (in yellow), and Byzantine invasion (in brown). Map available online in Atlas Nacional de España.

The Visigoths probably derive from a branch of the Thervingi tribe that inhabited the Danube plains a few centuries before their arrival to Iberia. One of the earliest and most significant interactions Visigoths had with Rome was in the Battle of Adrianople (378 CE) that occurred in the Roman province of Thracia. In this battle they fought alongside the Alans against the Roman Empire after crossing the Danube river border. At this time, many Germanic tribes were already pushing South and West as a consequence of the invading Huns, which kept pushing them further West during the next century. However, in another dramatic example of the shifting tides of His-

tory, we find the Visigoths less than a century later fighting along the Western Roman Empire and other Germanic tribes in the Battle of the Catalaunian Plains (451 CE). This battle was a turning point because it was the first time the Hunnic horde was stopped in their advance under Attila.

The struggling Western Roman Empire called upon the Visigoths to fight the Vandals and Suevii (both Germanic tribes), and Alans (Iranian nomadic pastoralists) that had invaded Hispania early in the 5th century. The Visigoths crossed into Hispania as *foederati* (allies without citizenship) of Rome and defeated the Alans and a faction of the Vandals only to be relocated into Aquitania (southwestern France) while Rome recovered control of Mediterranean Iberia. By the end of the 5th century and after further military interventions and the vacuum of power left by the fall of Rome, the Visigoths took over most of Iberia, while the Suevii remained in control of the Northwest (Figure 24). The Vandals on the other hand, had crossed to North Africa and embarked in a military conquest. It is not until the 7th century that the Visigoths find themselves in full control of Hispania and part of France after defeating the invading Byzantines in the Southeast. The Visigoths had to face a Byzantine invasion orchestrated by Justinian I in 552 which lasted until 624 CE (Figure 24). A century of relative peace and political unity followed thanks in part to the abandonment of *Arrianism* in favour of Catholicism after the Third Council of Toledo (589 CE). The Visigothic kingdom in Spain of the 6th-7th centuries was characterized by a rural pagan society, intruded by urban Christian institutions (Hillgarth, 1980), but also by a impoverished population and a rudimentary economy. These issues were probably inherited from the crisis of the third century. There are examples in Spain showing how the unstable political climate in the third century drove people to bury their wealth to protect it; for example the *tesoro de Valsadornín*.

In the year 711 CE the Visigoths lost the battle against a combined force of Arabs and Berbers lead by Tariq ibn Ziyad, following Musa bin Nusair command, acting governor of North Africa. The military defeat is followed by a total collapse of their rule in the kingdom, and the Iberian peninsula is rapidly conquered by the Umayyad Caliphate (Figure 25). In the Valencian region, Theodomir (Tudmir in Arabic, a major landowner in the area and political governor), and Abd al-Aziz (son of the governor in North Africa) signed the Treaty of Tudmir in Orihuela (713 CE) (Rubiera Mata, 1985). This treaty concluded the diplomatic conquest and peaceful capitulation of major cities, including Valencia. Helped by the political unrest of these years, Tudmir (whose palace was near Valencia), had become the *de facto* king of the Valencian region. In contrast to how modern historiography has portrayed the event, it is far more likely that the new conquerors did not represent a cultural earthquake, or at least they must have not been regarded very different from what the previous Germanic invaders meant to natives at the time. After all, the majority of the Islamic newcomers originated from a North African world that was largely romanized culturally and linguistically.

In general, Germanic tribes moved extensively over a relatively short period of a few centuries. This makes it interesting to see if they assimilated other peoples or remained a unit. At the same time, despite their historical importance, they were small in number making it difficult to leave a significant archaeological footprint in the populations conquered or visited.

Hispano-Roman Valentia in No Man's Land: between Visigoths and Byzantines

It is suspected that since the mid 4th century CE a primitive Christian community might have existed, organized around the cult and memory of a local saint (San Vicente Martir) who suffered martyrdom in Valencia in the year 304 CE. This community seems to be attested archaeologically since through a paleochristian lead coffin burial has been found dating to around this time.

A century later, during the end of Roman times and the beginning of the Migration Period, Valencia regressed into a minimal state and entered in decadence, entire boroughs were depopulated and infrastructure networks were abandoned. The city does not gain any historical mentions either during this time, because it is also disputed whether it is one of the mentioned locations in the list of the Treaty of Tudmir. Interestingly though, the palace of Tudmir has been recently discovered in Riba-roja del Turia. It consisted of a magnificent villa erected upstream by the same

river that crossed Valencia but only a few kilometres away from the city. The city is actually said to have been governed by Agrescio when it was sieged by the Islamic army of Tariq. Legend goes that after a steadfast defence, Agrescio and Tariq agreed on a friendly surrender that guaranteed respect to native religion and traditions in exchange of acceptance of Islamic political and military rule (Coscollá Sanz, 2003).

Coincinding with the first barbarian invasions and the onset of Visigothic rule, the Church takes over and city becomes a diocese governed by a bishop that started replacing Roman temples with building for the Christian cult. The city remains relatively independent *de facto* but nominally is a vassal to the Visigothic king in Toledo (Spain), capital of the kingdom.

The shrinking population of Valencia abandoned the extramural necropolis and new ones appear at the old heart of the city (García-Prósper, 2016). A change in the funerary practices can be observed in the relevant Visigothic families of the city. This represents a clear break with the immediately previous tombs in Roman style at the same location, characterized by the presence of tegulae and amphoras. The new elite starts using family pantheons to bury family members in consecrated land next to the Cathedral. An area linked to the shrine honouring the martyrdom of the local saint, San Vicente Martir. This necropolis was located in what nowadays is la Plaza de l'Almoina, originally the early Roman Forum (Figure 23) (Ribera i Lacomba et al., 2010).

From the 6th century (527 CE), under the rule of a local bishop named Justinian (circa 492-547 CE), and coinciding with Byzantine invasion the city experienced some recovery and even organized a regional council. This is likely because from 554 CE the city became strategically important in the Visigothic-Byzantine war and military forces fortified and settled in the city (Coscollá Sanz, 2003). After the final defeat of the Byzantines in 625, the city enters again a dark age with reduced urban activity of which very little documentation is known or has survived. Interestingly, a recent study searching for the traces of the First Pandemic, also known as the Justinian Plague, that included and screened 36 samples from this necropolis (6th-7th century CE) found a strain of the pathogenic genome in one of the individuals. Based on the radiocarbon dates of the sample and historical records this would correspond with the first outbreak in Iberia in 543 CE (Keller et al., 2019).

By the time of the Islamic conquest in the 8th century, the city was merely a near-depopulated coastal settlement, the urban activity in Valencia experienced a decay while Christian religion took over completely. The state of decadence of early Medieval Valencia it is perhaps best reflected by the loss of its latin name in the Arab sources, so deprecated that the Islamic conquerors initially referred to the city as Madinat al-Turab (City of Dirt) (M. Eplaza in Simon (1996)). Same applies for the nearby Sagunto which came to be known by the Arabs as Murvidero, meaning *old walls*.

1.5.3 Islamic Al-Andalus

The world of Al-Andalus in Iberia was the product of mixing Greco-Latin, Arab and Berber cultures and the root of this origin has its explanation in the composition of the new settlers (Watt and Cachia, 1965). Two main different ethnic identities formed the core of the new conquering force of Iberia (Figure 25).

Figure 25: Territorial evolution of Islamic Iberia (in green) from the 8th century CE to the 11th century CE. Emirate period (left), Caliphate period (centre), and Taifas period (right). From Atlas Nacional de España.

One the one hand there were the Arabs (Mackintosh-Smith, 2019), a group so incredibly diverse in origin that, the Yemen-based British Arabist, Tim Mackintosh-Smith says in his book *Arabs: A 3,000-Year History of Peoples, Tribes, and Empires* the label *arab* is so broad and slippery that is almost impossible to define, since its meaning has shifted over three millennia that cannot refer to one particular group. However, the term in its earliest forms referred to nomadic groups that existed as pastoralists transhumants from the desert at the fringes of civilization, since at least the times of the Eastern Roman Empire they were known for their raiding activities. Arabs however, as we understand them today following Muhammad and the advent of Islam, are an ethnic blend of peoples from the Arabian peninsula with diverse origins: agricultural sedentary South Arabian tribes (*qabīla*) from what today is Yemen, and semi-nomadic tribes from the North and the desert. Their ultimate origin is the prehistoric Fertile Crescent and their language belongs to the Semitic linguistic family. Although there have been later assimilation of other peoples, language is the most defining feature that binds the Arab identity together.

On the other hand there were the Berbers, North African natives, whose correct endonym is Amazigh and Tamazight their language. They formed the bulk of the first armies in Iberia due to the geographical proximity of their homelands (Watt and Cachia, 1965), and also because the restrictive patriarchal way of inheritance of Arab identity made Arabs always a minority. Berbers started a religious transition to Islam after the Arab conquest of North Africa. Such conquest was relatively quick, over a period of fifty years (647-709 CE) the Arabs and other Near Eastern peoples (Schuenemann et al., 2017; Skourtanioti et al., 2020; Agranat-Tamir et al., 2020) already integrated in the Caliphate conquered the Maghreb after defeating the Eastern Roman Empire and the Berber kingdoms. The Berbers encountered by Islamic forces were Christians and romanized but from this moment entered a process of arabization that changed their religion, culture and language forever. However, despite arabization, Berbers remained far from homogeneous. The deep division between nomadic and sedentary groups (Watt and Cachia, 1965) was never erased. The latter group constituted the majority of settlers in the rural areas of Spain and Valencia.

The Islamic conquest of Iberia was so swift that the Caliphate crossed the Pyrenees into France, but their advance was stopped by the Franks in the Battle of Tours-Poiters (732 CE) (Figure 25) (Gleize et al., 2016). After this battle, the Islamic policy shifted from territorial expansion to territorial consolidation. The native Iberian population had been granted *dhimmi* status under the new rulers. *Dhimmi* was a term used to describe the status of non-Muslim inhabitants under the protection of Arab rulers. This treatment was reserved only for other *people of the book* (monotheists), while polytheists faced the choice between *Islam or the sword* (Watt and Cachia, 1965). This

somewhat familiar treatment granted to the local Iberian people could be also explained by the fact that the majority of newcomers were North Africans and Berbers whose language, the long-extinct Afro-Romance speech, was the same or very similar to what was spoken in Iberia at the time (Wright, 2012). There are hints that some dialectal peculiarities in Spanish were also present in the Afro-Latin language (Adams, 2007; Wright, 2012), as inferred from the misspellings in the texts of the *Albertini tablets* from the Vandal period in North Africa (435-534 CE). As it seems, the *v* and *b* sound confusion in modern Spanish already existed in Roman Africa according to Isidore of Seville, but later becomes a Spanish trait as exemplified by the quote *"Beati Hispani quibus bibere vivere est"*.

The Arabs, who had been the driving force behind the wave of Islamic expansion, became the urban elite in the political landscape during the Umayyad Caliphate, that ruled from 711 CE until 1031 CE. This holding of the power was in spite not representing a majority of the new Islamic settlers. However, as decades advanced, marriages with non-Arab women by rulers in Spain and elsewhere in the Islamic empire started to introduce diversity in the Arab identity.

In terms of social structure, the social pyramid that originated following the Islamic conquest is perhaps more profuse in labels than other medieval counterpats of the same period. There were two main classes *Jassa* (elite) and *Amma* (commoners). Arabs belonged to the *Jassa* class, along with perhaps some Syrians and important Berber families. To the *Amma* class belonged the majority of Berbers, the *Muladis* (native Christians that converted to Islam), *Mozarabs* (Christians living in Muslim territory), Jews, and slaves (Figure 26).

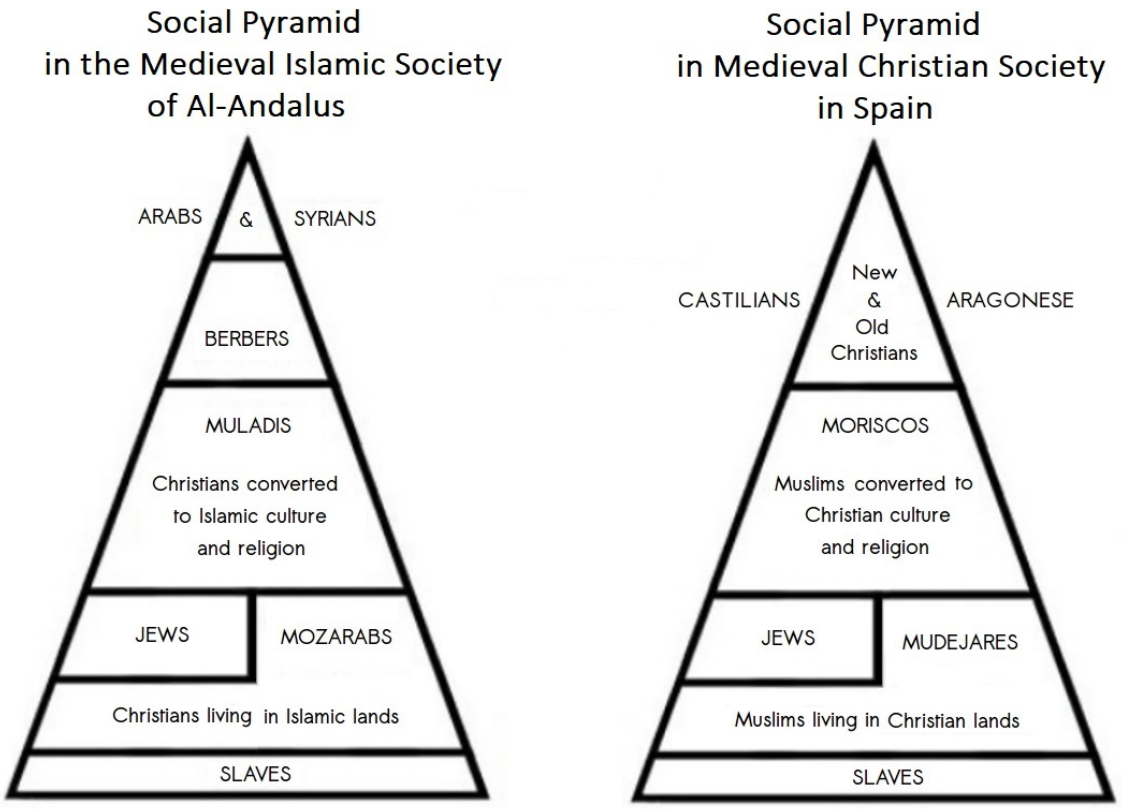


Figure 26: Schematic social pyramids of the two medieval cultures in Spain.

The start of the collapse, in 1009 CE, and fragmentation of the Caliphate of Cordoba into smaller kingdoms (first period of *Taifas*, 1031-1090 CE) also brought about the end of Arab dynastic rule in Al-Andalus. Berber tribes, who had not wielded any significant political power initially, saw the picture change during the 11th-13th centuries and the *Taifa* mayhem. It is in this period when two successive Berber empires took control of the Maghreb and Iberia: the Almoravids

(1090-1145 CE) and the Almohads (1145-1223 CE) (Watt and Cachia, 1965). The Almoravids were a religious reformist movement turned military empire. It started amongst the ancestors of the modern Tuareg, the Sunhaya, a Berber nomadic tribe of camel herders of the Saharan steppe that extended to Senegal and the Niger river to the south. They succeeded in conquering and unifying the fragmentary states in the Maghreb in only a few decades. Their first emir, Youssef Ibn Tachfin, founded Marrakech in 1062 as an strategic outpost but it soon became the new capital. After securing North Africa, the Almoravids were called upon from Iberia by the Andalusian ruling elite that was struggling with the expanding Christian kingdoms at the time. After corroborating the weakness and division of the Taifas in Al-Andalus, the Almoravids conquered and unified the Taifas under their rule. The presence of Sunhaya groups in Spain is attested by the toponimy, a few villages carry names derived from this tribe (e.g. Soneja) that settled in the Valencian mountains (T. Glick in Simon (1996)). The Almohads followed a similiar trajectory, originating as a religious movement, but stemming from a rival Berber tribe from the Atlas mountains: the Masmuda (Watt and Cachia, 1965). The subsequent conquest of Marrakech by the Almohads in 1145 CE effictively ended the Almoravid empire. The Almohands turned their heads to the post-Almoravid divided *Taifas* of Al-Andalus after a succesful campaing in Algeria. They also successfully took control of Al-Andalus and held power in the South and East until the Christian kingdoms reduced the Islamic territory in Iberia solely to the kingdom of Granada.

Balansiya: from City of Dust to Garden of the Caliphate

Regarding Valencia, after the signature of the Treaty of Tudmir in 713 CE, the city starts to be known as Balansiya and it gets integrated to the Emirate of Cordoba that later became the Caliphate of Cordoba (Watt and Cachia, 1965; Coscollá Sanz, 2003) (Figure 27). During the Visigothic and Arab rule and until the 10th century, the regional population density was very low. In fact, Valencia was destroyed once more by the caliph after an revolt in 778 and remained almost ruined for some 20 years (Ubieto, 1975; Butzer et al., 1985). The neighbour city of Sagunto followed a similar fate, which delayed the growth of these urban centres until the 11th century. However the urban, commercial and general relevance of the city remained very modest during the early Islamic Caliphate since there is not much archaeological evidence of immediate transformation following the conquest. Perhaps, the best example are the gardens of Russafa (*al-munyah*), built at some point in the 9th century by Abd Allah al-Balansi, a son of the first independent Emir Abd al-Rahman I, but they were located outside the walled core of the city. This garden imitated the Munyat al-Rusafa gardens from Cordoba built by his father. The Rusafa from Cordoba was also built to recreate the original Rusafa garden from Syria. The original garden must have been part of an Umayyad palace with a central pavilion discovered in modern Resafa (Syria), originally built by the grandfather and Caliph, Hisham (724-743 CE) (Otto-Dorn, 1957; Ruggles, 2000, 2008). Other improvements of the Caliphate rule were the development of a perimeter of orchards in the area of the modern neighbourhood of El Carmen, and turning the old Visigothic episcopal area next to the modern Cathedral into a mosque and marketplace (*Souq*) vinculated to the governor. Besides political and urban developments, what was truly transcendent was entrance of the city into the Islamic world and the change of language, religion and customs of its inhabitants in a short period of time.

It is not until after the 10th-11th century CE when the centralized power of the Caliphate of Granada collapses that more relevant evidences of Islamic activity start to appear in the city. This coincides with the establishment of Valencia as a capital in the new period of Taifas. The first Arab elite does not settle until around 1011 CE rulers of slavic origin associated to the Amirid dynasty (allies and descendants of Almansur) took over the city, later they lend power to Almansur's grandson Abd al-Aziz ibn Amir, who was also the maternal great-grandson of the King of Navarra; Sancho Garcés II. However Berber agricultural and pastoral settlers remained majority among the newcommers, as reported by the geographer Yaqubi (Watt and Cachia, 1965). The time of Islamic relevance and splendour of the city also overlaped with the rule of the Almoravids that captured

Valencia in 1102 CE from the Christians. as an independent warlord *El Cid* had conquered and ruled the city for himself during a brief period of five years between 1094 and 1099 CE. In the next decades the power in the city was held by the family of the Banu Mardanix clan (1142-1172 CE) for whom a funerary stele has been recovered attesting their presence. Later on another Berber empire, the Almohads, extended their dominion to the region (Watt and Cachia, 1965; Simon, 1996).

During the second half of the Islamic period in Spain, the city flourished and the defensive walls were improved to face the Catalan-Aragonese advance. The 11th century walls covered an area of 45 hectares, and at the time of the reconquest in 1238, the area including suburbs was about 53 hectares (Butzer et al., 1985) (Figure 27). Important agricultural improvements must have occurred with the development of the *acequia* system (*al-saqiyah*) to irrigate the fields (T. Glick in Simon (1996)), orchards and farmlands of the surrounding hamlets known as *alquerias* (*al-qaria*) (Butzer et al., 1985; Coscollá Sanz, 2003). The Arabs incorporated to the local agriculture the cultivation of rice, sugar cane, citrus fruits, silkworms, as well as sorghum, some of which still remain very important nowadays (Butzer et al., 1985). Important necropoles develop north and west of the wall gates. The extramural settlements consolidated as boroughs (*rabaḍ*), one of them developing around the Russafa gardens to the south. Some, like the arrabal de la Alcudia, located north of the city are linked to pottery activities, and it is likely where many craftsmen resided. Archaeological excavations have revealed that a great number of wells were excavated to extract clay from the banks of the river (Coscollá Sanz, 2003).

The importance of water supply for human needs was critical in shaping the identity of Islamic societies in region of Valencia. The canal system around the city and the region was intricate and of great complexity, and it might have profited from pre-Islamic foundations (Butzer et al., 1985; Beltran Lloris and Willi, 2012). Before the mass drainages of coastal marshlands carried out during the 19th century, the Valencian irrigation system, covered over 100 km2, with water being supplied using six arterial canals radiating out onto the coastal plain like deltaic distributaries, in a similar fashion to the Nile Delta. Irrigation in this region had three different scales: *macro*, *meso*, and *micro*. Firstly at *macro* scale, when many towns jointly managed water distribution. It involved thousands of people, an elaborated canal system. Scondly at *meso* scale, a community of few hamlets, cotrols the water from one or few major springs, involving hundreds of cultivators and a small network of canals. Finally, at *micro* scale, one extended family managed a few irrigation ditches (covering no more than a hectare) fed by a small spring. Outside the cities, the rural Islamic world was structured in communities of small hamlets (*alquerías* of *meso* or *micro* systems) around a castle to which they were affiliated for tax collection and administration and defensive purposes (T. Glick in Simon (1996)), and of course water use management. Agriculture was not only important for the city of Valencia, it was central to the whole region. For readers who might not be familiar with it, the orography of this region consists of plains that extend from the Iberian Mountain Range until to sea for kilometres. Such territory is well suited for the practice of agriculture, which combined with the technical advances favoured the development of multitude of rural settlements in the shape of hamlets, most of which had not existed until this point, along this coastal strip of fertile land.

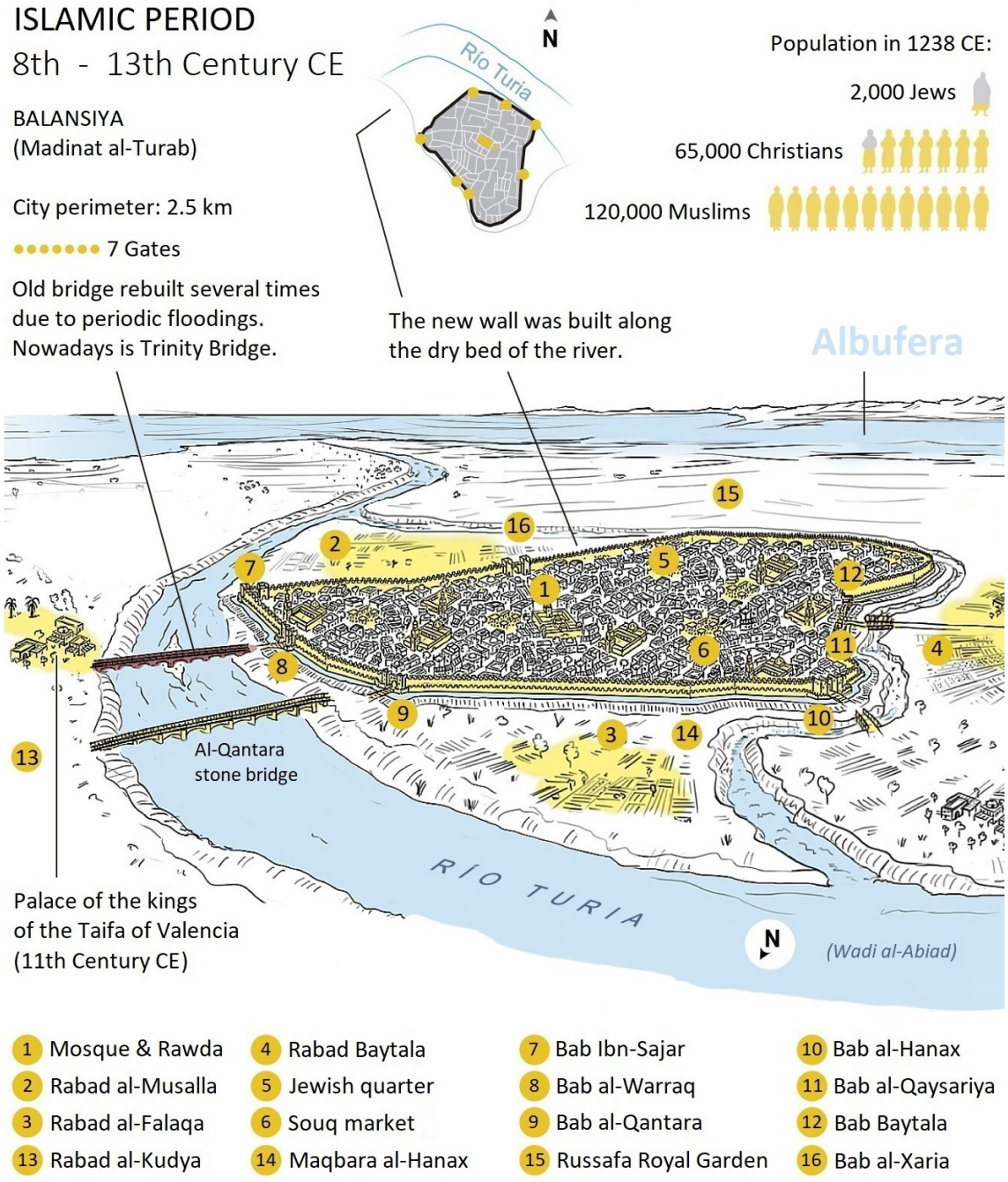


Figure 27: Infography depicting a reconstruction of the city of Valencia (Balansiya) during the Islamic era. *Rabad* means hamlet, *Bab* means gate, *Maqbara* and *Rawda* both mean necropolis and royal necropolis respectively. Rabad al-Kudya also had a maqbara from where genomes have been sequenced. This illustration was originally published in the article *La Valencia por desenterrar* in *El Mundo* newspaper on the 28th of October 2018. The artist is Pedro Jiménez (@PedroJimnez), the material is accessible at *infografia-pedrojimenez.blogspot.com*. Image has been modified to translate and adapt the text with permission.

1.5.4 Christian Medieval Spain

The so-called *Reconquista* is a period which comprises the centuries long expansion of the Catholic kingdoms southwards from 1223 CE after the defeat of the Almohads, until in 1492 CE when the last Islamic kingdom of Granada falls (Figure 28). However, the culmination of this conquest did not imply the end of moors or muslims in the peninsula. A large population of *Mudejares* (muslims that stayed in christian terrioty) later known as *Moriscos* (forcibly converted Muslims) persisted in Spain. A large population of *Moriscos* remained in Spain until 1609 CE, most notably in the east (formerly known as *Sharq al-Andalu*s), when they were expelled by force by order of the Spanish Crown and had to seek refuge in North Africa or the Near East (Watt and Cachia, 1965; Glick, 1970, 1977).

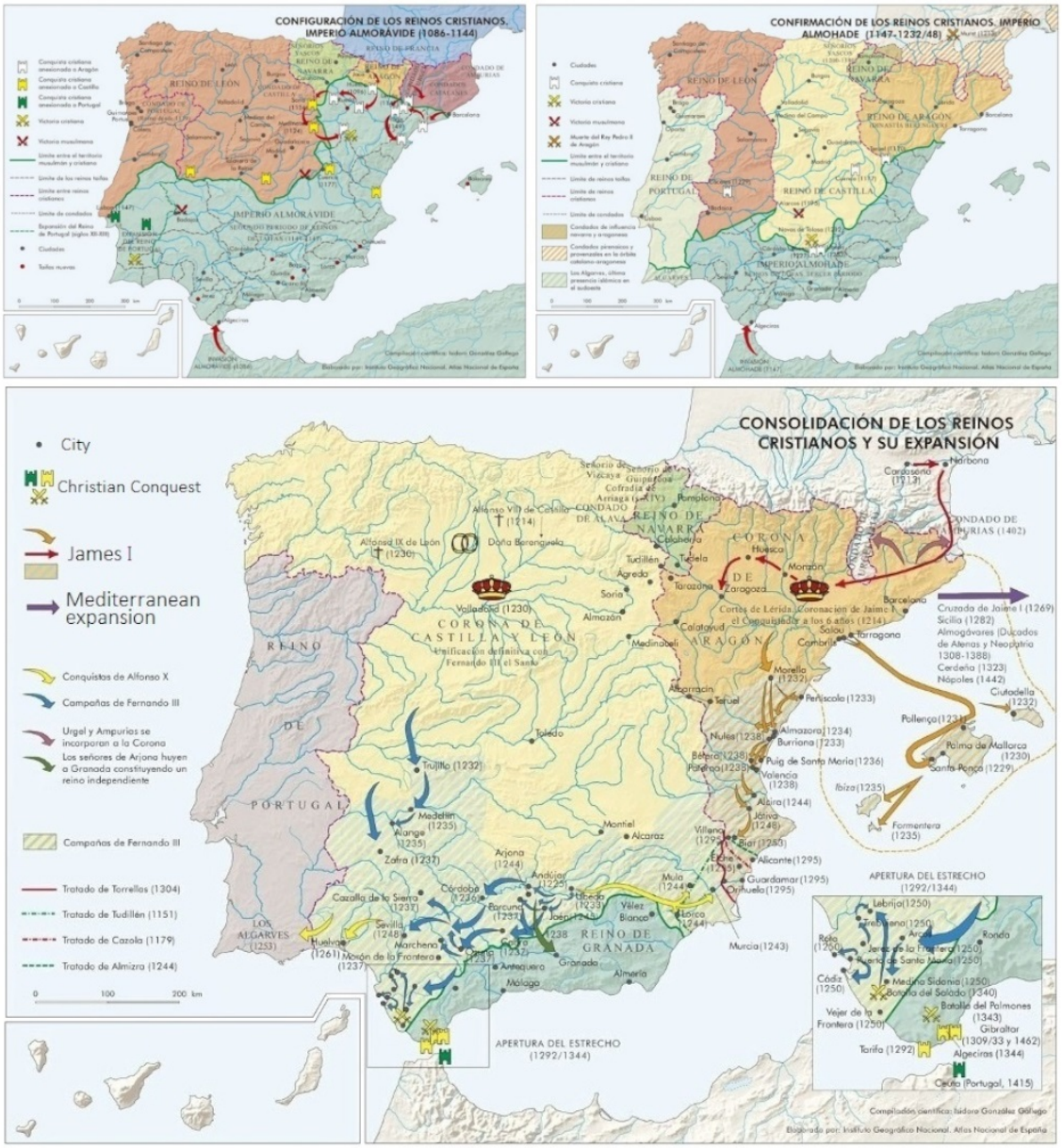


Figure 28: Territorial expansion of the Christian kingdoms in Iberia from the 11th century CE to the 14th century CE. Al-Andalus is in green in all three maps. The territory of Castilla is in pale yellow in the maps of the centre and right. The territory of Aragon appear in orange in all three maps. From Atlas Nacional de España.

The Conquest of Jaume I and the Golden Age

The end of the Islamic rule in the the Mediterranean region that would later become the Kingdom of Valencia (Torró, 2012) occurred in the first half of the 13th Century CE (Figure 29). For the city then known as Balansiya, the end of Islamic rule arrived in 1238 CE, when after a siege of six months Zayyan ibn Mardanish, the last Islamic king of an independent Taifa of Valencia capitulated to the forces of king Jaume I of Aragon, *the Conqueror*. In 1229 CE Zayyan had become king of Valencia and *de facto* ruler of all Sharq Al-Andalus, after a successful coup against the Almohad governor Zayd Abu Zayd. Zayd sought the help of Jaume I to regain power, but instead submitted in vassalage and converted to the Christian faith in exchange of lands in Aragon. However, this was an important turning point in the history of Valencia since it gave Jaume I a *casus belli* for his Valencian campaign of 1232 that culminated the 9th of October 1238 CE with the conquest of the capital city (Figure 29). The following next two centuries witnessed the Golden Age of Valencian literature helped by the fast transformation of the city of Valencia into a cultural hub of the Mediterranean amidst the Catalan-Aragonese campaigns of expansion.

Jaume I is a very famous figure in the popular imagination for his military conquests and legends surrounding his exploits (Burns, 1973; Belenguer Cebrià, 2008). A depiction of the famous Battle of the Puig can be found at the Victoria and Albert Museum in London as part of Saint George altarpiece. The king is also very important for historiography since he produced two documents crucial for the study of the Medieval period in the Crown of Aragon: the *Llibre dels feits* (Book of the Deeds) and the *Llibre del Repartiment* (Book of the Distribution) (Lewis, 1968; Burns, 1973, 1975; Glick, 1970, 1977; Cabanes Pecourt, 1977; Burns, 1991).

The Book of the Deeds is an autobiographical chronicle of the king and the most relevant events that unfolded during the conquest of Mallorca and Valencia. The aim of Jaume I with this chronicle was very clear, as the king himself is quoted saying in the book: *"E per tal que los hòmens coneguessen, quan hauríem passada aquesta vida mortal, ço que nós hauríem fet, e per dar eximpli a tots los altres hòmens del món" (So mankind can acknowledge, when I have left this mortal life, what I have accomplished, and to give example to all other men in the world)* (Soldevila et al., 2007). Such a personal document helps us gain insight into what the city of Valencia meant for the king during the siege of 1238. He had already conquered major settlemets (Morella, Burriana, Almenara, Nules, Jubayla/El Puig...) north of Valencia, and when confronted by Zayyan to abandon the siege in exchange of generous tribute Jaume I replied: *"Nós som venguts a hora e a punt que podem haver València, e així haurem la gallina e puis los pollets" (I am here now so I can take Valencia, and like this I shall have the hen and all the chicks)* This quote reflects how important the city was at this time, because the king expects that a domino effect will precipitate more places to surrender when the city is taken. It is even more relevant in the light of the fact that many of his vassals an allies preferred the city to remain under Islamic control for their own benefit (Belenguer Cebrià, 2008). In the terms of surrender of Valencia on the 28th of September 1238, Jaume I guaranteed twenty days of safe passage to the Islamic controlled south for the muslim population (Torres Balbás, 1951) that wished to abandon the city with everything they could carry with them. This was also extended to Zayyan who left the city on the 8th of October, with an agreement for a truce of seven years. Meanwhile, the border was set on the river Xuquer. Jaume I entered the city on the 9th of October, however, unrest amongst the Christian noblemen who had been prevented from sacking the city, attacked the fleeing Muslim refugees and the the King had to intervene (Belenguer Cebrià, 2008).

The other important document left by Jaume I - the Book of the Distribution - is a compilation of legal documents with meticulously recorded donations of properties and promises of land by Jaume I at the conclusion of the Valencian campaing. The *Repartimiento* was a system of repopulation used in Valencia for the land taken from the 13th century. It was an ordely manner to distribute land, farms, houses, etc... based on social rank and merit of the help granted to Jaume I during his campaigns (Cabanes Pecourt, 1977). Based on these books, we know that given the relatively pacific nature of the conquest the majority of the Muslim rural population remained in

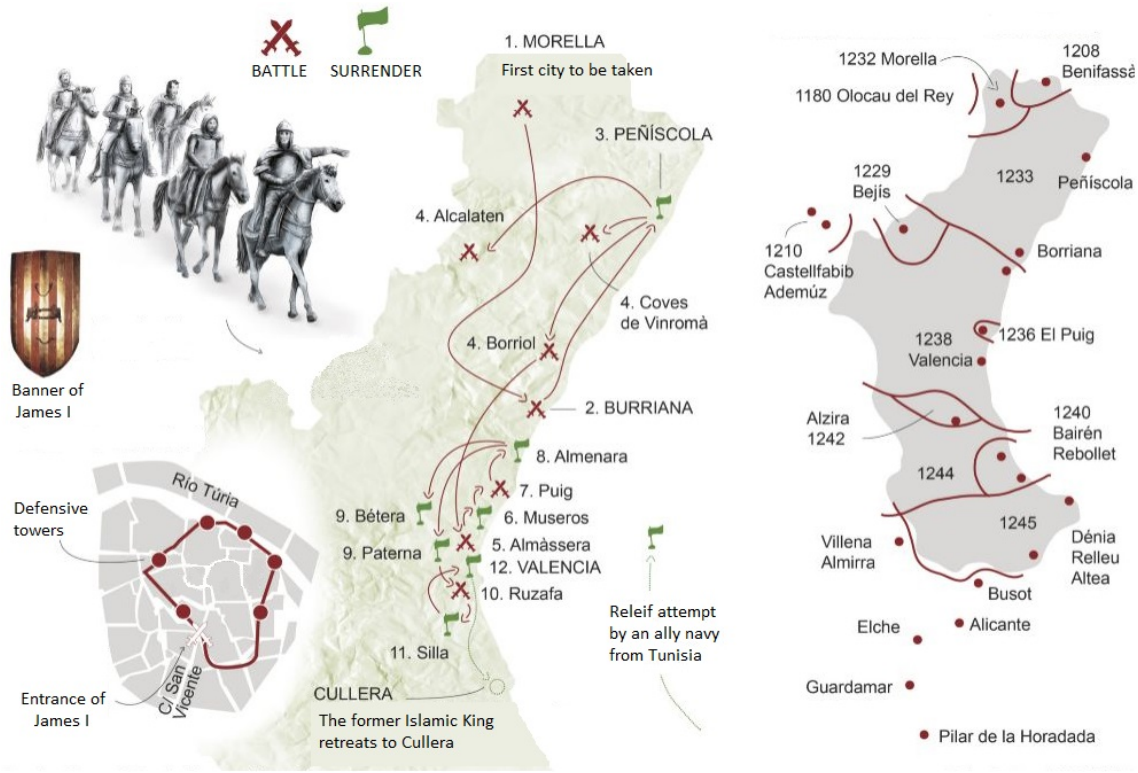


Figure 29: Itineray and dates of the conquest campaign of the kingdom of Valencia by Jaume I. Illustration originally published by El Mundo newspaper. The artist is Pedro Jiménez (@Pedro-Jimnez). Image has been modified totranslate and adapt the text with permission

place (Burns, 1973, 1975), with only urban sites being depopulated of natives and repopulated by Aragonese and Catalan folk. The dispensations also led to the fragmentation of the land into smaller states but also developed a rural middle class and the prosperity of the agricultural lifestyle in the region until the 20th century. Looking at the *Llibre del Repartiment* (Cabanes Pecourt, 1977) we can also conclude that the population turnover in Valencia city, and the genetic makeup, must have changed at least some centuries earlier than in the rural world, where the muslim population carried on, albeit in less favourable conditions. These Muslims natives that remained became to be known as *Mudejares, Sarrains* or *Moros*. Originally *Mudejar* was the Castilian term employed in the Crown of Castilla (the other two were more common in Aragon) but it has replaced the other two since Castilian imposed over modern Spanish speech.

Three religious communities, Muslims, Christians and Jews, coexisted through the next century in the city (Figure 30), but despite the promises made to the Muslim community by Jaume I following the conquest, segregation of both Muslims and Jews was set in place in the following decades. Christians designated quarters in cities where the Muslim community could live (*Morería*) and another for the Jewish minority (*Judería*). Theoretically, the Jewish community was under direct protection of the king since it was considered as the king's own private property. In the rural world there was no presence of Jewish communities because they could not develop their main economic activity, in practice Jews acted bankers since money lending was prohibited by Christian dogmas. Instead Jews concentrated in cities like Valencia and Sagunto. In Valencia, an important Jewish quarter existed until the late 14th Century when it ceased to exist after a violent and devastating pogrom on the 9th of July 1391, following a countrywide anti-Judaism revolt. As a consequence of this historically recorded pogrom, the Jewish community in the city disappeared. In the same century, the city endured a recorded episode of the Black Death plague in 1348 CE (Rubio Vela, 1980; Ruiz-de Loizaga, 2009).

One might ask why, if religion was so central to society at the time, Muslim *Mudejares* were

allowed to stay (Torres Balbás, 1951) and keep their traditions in a fairly openly, at least initially. The answer is a simple one, the *Mudejar* population avoided prosecution from the Inquisition initially because it was in the best interest of the rural nobility - and the Church - to maintain sufficient labouring population. As mentioned above, *Mudejar* numbers might had dwindled in urban settlements, but they were still a majority in the rural population working the fields of the states that the new Christian lords had inherited from the expelled Islamic rulers. It was not viable to vacate the lands without risking economic collapse (Burns, 1973, 1975).

The views about the Muslim community that remained were conflicting: valued for their work force by the lords; viewed by the Church as evangelizing targets as the ultimate goal of the Valencian Crusade had been; but also, resented by the Christian common folk for what they represented and the few rights they had retained (Burns, 1973, 1975).

On paper Jaume I had signed covenants of friendship for the protection of the remaining Muslims, but word was betrayed and the reality turned out to be very different. Coexistence was troubled by cultural and religious segregation, and became the seed for future mistrust. We have already mentioned that cities developed an urban layout that isolated the Muslim and Jewish ghettos. Meanwhile, in the rural hinterland there was no need to create such neighbourhoods of confinement since the entire town was usually either entirely Muslim, or Christian. The lord lived in the castle separated from the villagers (Burns, 1973, 1975).

After the *Repartimiento* many rural land lords devoted themselves to try convert their new Muslim serfs. One of these many stories is Raimundo de Morelló; first Christian lord of the villages of Algar and Arguines, near Valencia, a farmer from Morella (Castellón) that helped Jaume I in his Valencian campaign. He had his origins in an important *Mozarabe* family, the Christians that had lived under Islamic rule as a minority before the conquest of Jaume I. He was not a nobleman, hence he could not be awarded with a title but was granted stewardship of the lands and the people of the village of Algar. He was religious to the fault and member of the Order of Mercy, not surprisingly given his family background, and so he probably was imbued with the intention of baptizing his new serfs. He ordered the construction of a small parish near a farm where a foundling image of the Virgin Mary was encountered. Miraculous findings of a Virgins statues, such as this one, were is not unheard of in the region because they were fabrications destined to act as Christian propaganda. He also founded a hospital for peregrines and the poor in his domains managed by the Order of Mercy. In his will he legated everything to the Order of Mercy and stated his desired to be buried in the church he had built (Recio, 2011). Judging by later events in the 16th century, it t appears that the efforts of Morelló and other such attempts were largely unsuccessful.

The resentment among the Christian lower classes mounted through the centuries in all Iberian kingdoms. This resulted in a decree of forced conversion for *Mudejares* of Castilla in 1502 CE. Muslims in the kingdom of Castilla were given the opportunity to choose between conversion or exile. From this moment the Muslims that had been living as *Mudejares*, became converts known as *Moriscos*. In the kingdom of Aragon these events were delayed by some years until the *Revuelta de las Germanías* (Revolt of the Brotherhoods; 1519-1523 CE). The *Germanías* in Valencia were an armed social conflict between the burgeoise and Christian urban guilds of the cities against the rural world of Christian lords and Muslims serfs, the *Mudejares*. It was especially difficult to suffocate because the guilds had acquired the right to arm themselves to fight off barbary pirates in the previous centuries (Díaz Borrás, 1993).

The mob violece during the *Germanías* sparked forceful mass conversions of Muslims. These conversions were later validated with the same formula as in Castile: conversion or exile (1525-1526 CE). This marks the end of the *Mudejares* in Valencia, and the start of the *Moriscos*. Althought there is discontinuity in the terms used, the reader should not forget that *Mudejar* and *Morisco* refer to the same group of people that remained culturally distinct. Many individuals experienced the transition during their lifetime. In practice, rural lords and the *Aljamas* (self-governing Muslim community living under Christian rule) offered payments to go back to the

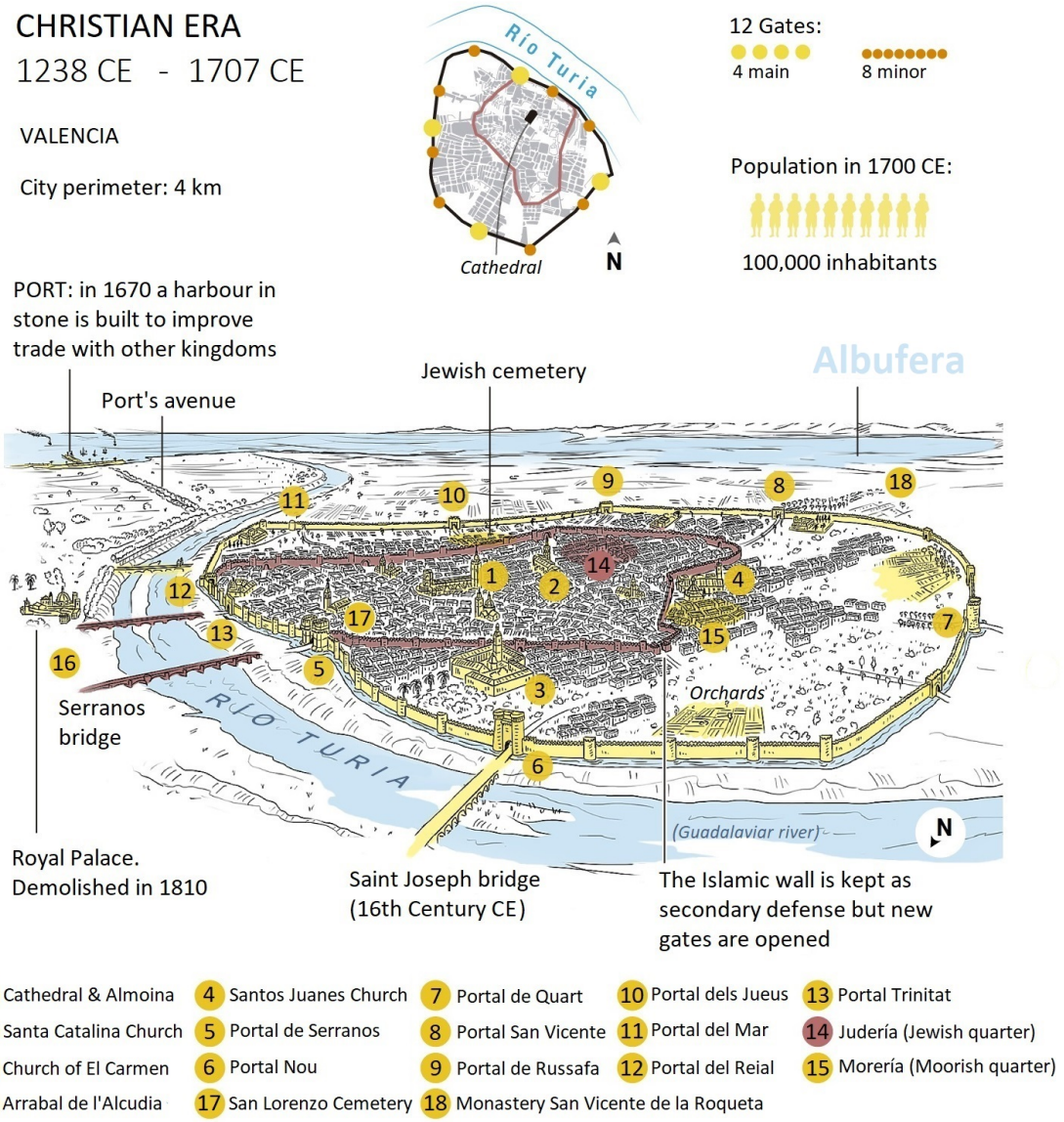


Figure 30: Infography depicting a reconstruction of Valencia in the Christian period. *Portal* means gate. Numbers 5, 6 10, 11 and 13 would translate as Gate of the Mountains, New Gate, Gate of the Jews, Gate of the Sea and Trinity Gate for example. The Jewry (number 14) ceased to exist in the 14th CE after a pogrom on July 1391. Illustration originally published in the article *La Valencia por desenterrar* in *El Mundo* newspaper on the 28th of October 2018. The artist is Pedro Jiménez (@PedroJimnez), material available at *infografia-pedrojimenez.blogspot.com*. Image has been modified to translate and adapt the text with permission.

status quo prior to the revolt. The rural lords managed to maintain a cheap labour force and prevented the intervention of the Inquisition which resulted in no expulsions, but the wheels were in motion towards a point of no return (Figure 31).

In the Mediterranean fringe of Iberia the *Morisco* population remained abundant and culturally differentiated, whereas in Castilian hinterland was diluted by integration to a greater degree. Conflicts and *Morisco* revolts endured during the 16th century (e.g. War of the Alpujarras in 1568-1571 CE) because their liberties continued to be trimmed.

The escalation of events led to the promulgation in 1609 CE, by Philip III of Spain, of a decree that declared expulsion of the *Moriscos* from Valencia, and later the rest of Spain (Halperin Donghi, 1980). The *Moriscos*, who made up one third of the Valencian population, were order to pack their things and head towards the ports of Valencian ports where they would be embarked in ships with destiny in North Africa. The fate these people met, banished from their ancestral homeland, was terrible. Headed to the great unknown, some families were thrown to the sea, whereas, the ones that arrived to North Africa were assaulted and mugged by the locals, who despite sharing the same religion regarded them as foreigner. This events are represented in several art pieces of the time. We also know that some Valencian *Moriscos* refused to go quiet and headed to the mountains of Castellón where they were later massacred by the Christians (Recio, 2011). It seems that the decree succeeded, culminating a centuries long cultural genocide and effectively erasing the last remnants of the Islamic past in the peninsula. Nevertheless, surviving documents from the time indicate that some young *Moriscos* were allowed to stay, for reasons not clear, with Christian families and employed as servants (Recio, 2011).

Following the events of 1609 CE, the indigenous population living in what is nowadays Castellón, Valencia, Alicante, Murcia and parts of Andalucía (Almeria and Granada) suffered a severe demographic decrease (Halperin Donghi, 1980; Bycroft et al., 2018). Historical documentation indicates that many towns that were fully composed by Moriscos ended up depopulated, and new settlers had to be brought from other territories (Aragon, Catalonia and Navarre) and from Valencia itself (Recio, 2011) to avoid economic and demographic collapse (Halperin Donghi, 1980). This was especially dramatic in the Valencian region since it harboured the majority of the Moriscos in the Crown of Aragon. In the kingdom of Valencia the Moriscos were systematically targeted and greatly suffered the consequences of the decree. This is in opposition to the Crown of Castilla where they had diluted and integrated for longer and more effectively. Many surnames currently common in the Valencian region (e.g., Navarro, Zaragozá, Catalá) are geographically structured and reflect the provenance of the post Medieval Christian re-settlers. It is traditionally assumed that the hinterland was mostly repopulated by non-Catalan speaking Aragonese folk, and the coast by Catalonian folk (Catalan speakers). A divide that still echoes faint linguistic boundaries today. A recent study (Bycroft et al., 2018) has identified genomic signals derived from these late and post-Medieval events of repopulation. It is debated whether the original population of the modern Valencian region that first contacted with incoming settlers arriving during the Islamic period was largely reduced or even replaced by the 17th century. Therefore, DNA from archaeological samples from this period provides an important tool to understand the demographic dynamics of the Islamic period in East Iberia, and clarify why seven centuries of Arab and North African presence in the Peninsula left such a modest genetic mark in the present-day population.

The mass conversion aimed to transform the divided colonial Valencian into a unified Christian nation to maintain the stratified social structure, but the illusion did not last long. The *Moriscos* were only Christians in name, and they never let go of their traditions and true beliefs which in the end was the ultimate cause of the expulsion in 1609 CE. However, the remedy turned out to be more painful because even two centuries later hundreds of forgotten villages and ruins were still scattered across the Valencian landscape (Halperin Donghi, 1980; Simon, 1996) (Figure 31).

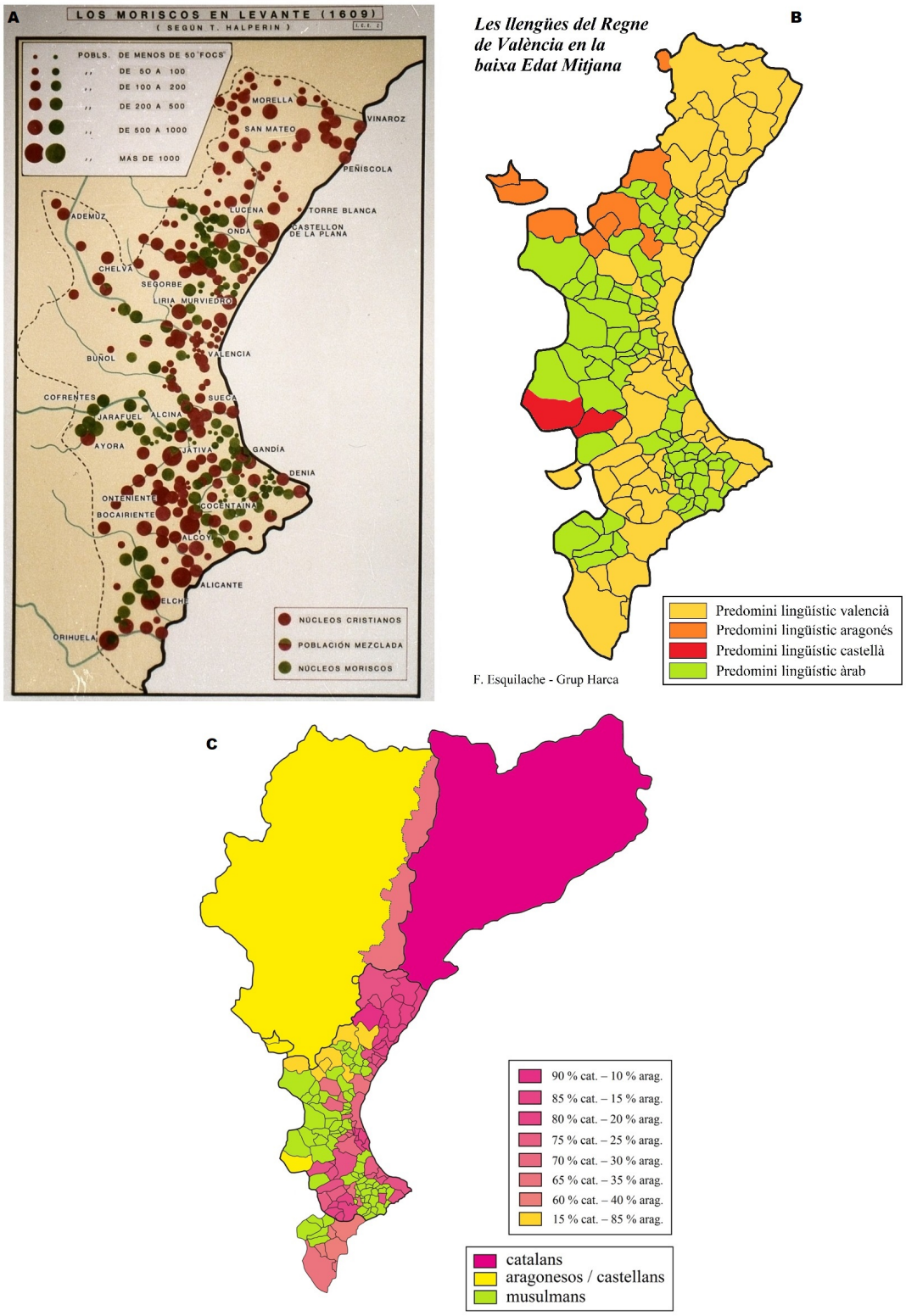


Figure 31: (Map A) Christians (red) and Moriscos (green) in towns of Valencia in the 17th century before expulsion. From Halperin Donghi (1980). (Map B) languages of the Kingdom of Valencia in the Late Middle Ages. Yellow denotes zones where Valencian/Catalan language predominates, orange where Aragonese dominates, red where Castilian predominates and green indicates majority of Arab speakers. Sourced from Ferran Esquilache (Grup Harca). (Map C) Catalan and Aragonese colonization of the new Kingdom of Valencia (1238 to 1425 CE). Pink-Yellow gradient indicates majority of Catalan or Aragonese settlers respectively, green indicates majority of local Muslim folk. Sourced from Vicent Baydal (Grup Harca).

Slavery in the Mediterranean Middle Ages

In parallel to the internal struggles of the Valencian mixed Christian and Muslim society in the Late and Post Medieval period, another social storm started to brew in the 14th century: the trans-Mediterranean slave trade, a greatly overlooked episode (Phillips, 1985). In the century following the conquest, the city of Valencia integrated in the Crown of Aragon which was immersed in an expansion across the Mediterranean. From the late 13th century, this Catalan-Aragonese expansion resulted in the conquest of Sardinia, Sicily, the kingdom of Naples in South Italy, and the Duchy of Athens in Greece, but also enclaves in the Magrheb. In the same manner, the conquest of America brought riches for Castilla, the Aragonese crown entered a period economic bonanza, thanks in part to the constant influx of slaves (Graullera Sanz, 1978; Simon, 1996) (Figure 32).

The Christian kingdoms had regained control of the Strait of Gibraltar in the 14th century, even before the fall of Granada in 1492 CE, and the military superiority in land started to push their North African rivals to the sea as the only scenario to compete. By the decade of the 1470s, we start to see evidence of Islamic pirates in the Mediterranean with bases in the Barbary Coast of North Africa (Díaz Borrás, 1993). These pirates usually launched raids in the Mediterranean coasts of Iberia to capture slaves that were later sold in slave markets in North Africa, but they rarely engaged in naval battle. Miguel de Cervantes, author of the famous novel *Don Quijote de la Mancha*, found himself prisoner in one of these prisons for captives in his youth. By the 16th century, however, the activity of Berber pirates was so intense that Christians had been cornered in land (Díaz Borrás, 1993) in a sort of primitive form of maritime guerrilla war. This phenomenon still resonates today in the use of a famous Spanish expression. In reality, the formation of charity organizations that payed sums of money to redeem Christian captives had heightened the problem over the years. Of course, piracy was not a one-way phenomenon, Christians also jumped at the opportunity to raid the North African coast in search of slaves that were later sold in the European slave market (Graullera Sanz, 1978) (Figure 32).

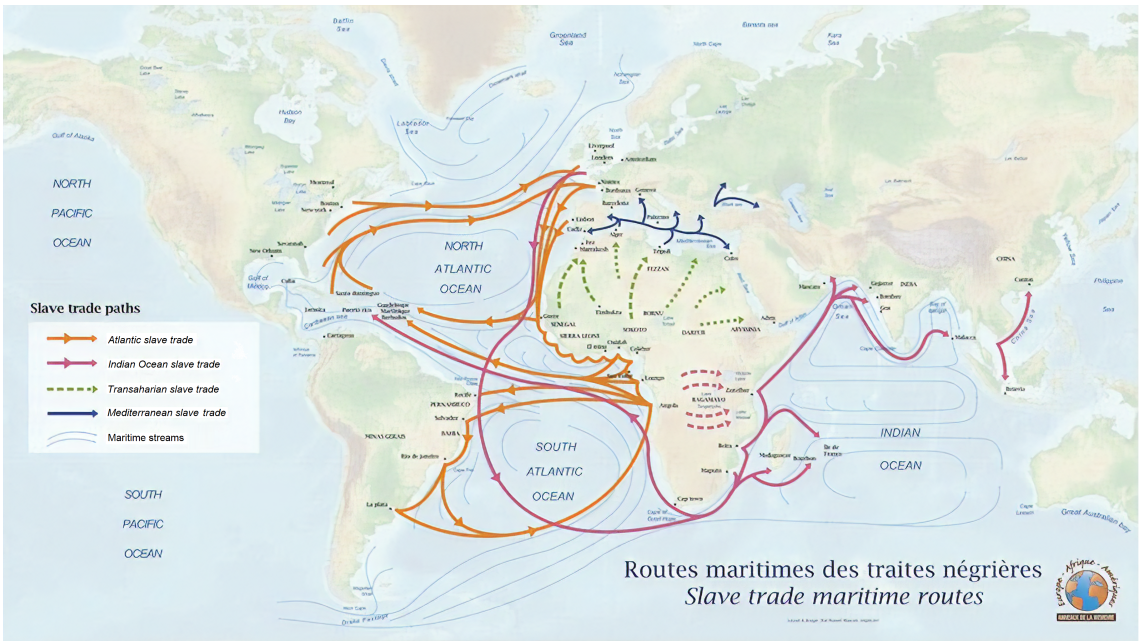


Figure 32: Extent of slave trade routes across the world during historical times. Sourced: Whitney Plantation.

The slave trade phenomenon in the Mediterranean Middle Ages was widespread and with deep roots (Phillips, 1985), and there were two models that supported it. The first, we have already gone through, was the provition of slaves obtained in pirate raids (*corso*). This type was more common in places like Portugal, Naples, Sicily, Castile, Catalonia, Majorca and Valencia. In this type of

enslavement the sex ratio was more balanced, and the diversity among the slaves was high, and so it was the range of tasks they were employed to do. The slaves from North Africa were not only Mediterranean, there were also sub-Saharan slaves. For example the Barca Mountains in Cyrennaica were an important point of redistribution since it was the end point of a trans-Saharan slave caravan route. The other model was more common in the North of Italy and the origin of the slaves was not derived directly from piracy (*trata*). The archetypical slave in the Medieval city states of Genova, Venice and Florence was a white female dedicated to domestic service.

In the 15th century Valencia started to become an important slave trade centre in the Western Mediterranean (Cortés-Sánchez et al., 2019; Graullera Sanz, 1978; Marzal Palacios, 2002, 2006), so important that the academic Charles Verlinden states that it had a significant role in trafficking of black slaves from Africa. However, Valencian slavery had a distinct feature: not all slaves were foreign, some were local Muslims that for different reasons had lost their freedom. The reasons for a local Muslim to become a slave included punishment for having participated in the revolts of the 13th century and later or being captured in other conflicts between Iberian kingdoms. However, the losing of freedom could also come from petty crimes such as theft and adultery; or religious offences; or violations of the movement restrictions like illegal emigration outside their area of confinement. Muslim slaves, local or international, feature during the whole Medieval period, but sub-Saharan slaves peak during the late 15th century, and are rare before and after this moment judging by the records available (Cortés López, 1989).

Slavery in Medieval Valencia is an obscure topic, rarely discussed, but recovered documentation by researchers for the period between 1375 CE and 1425 CE helps to aprehend the magnitude of the phenomenon. A total of 1,275 sales of slaves between individuals were registered before the *Baile General* of Valencia. Their procedence was recorded as follows: 597 Orientals (Tatars Tatars, Russians, etc... consituted the majority, around 46%), 502 Saracens (39%), 63 Balcanics (5%), 47 Sub-Saharans (4%), 26 Sards (2%), 13 Turks (1%), 5 Guanches (1%), and 22 were unknown (2%). Some examples of such transactions include: in 1407 CE Pedro de Viladarant, trader from Teruel based in Valencia sold a female Tatar slave; in 1414 CE an Aragonese Jew named Gostantí based in Sagunto sold a female Saracen slave. The list goes on, and it even includes members of the Church. The amount of Orientals or Slavs might seem surprisingly high but lest we forget that the etymological origin of the word *slave* derives from the ethnonym term *Slavic*. This is because this group was enslaved in mass during the Middle Ages by other nations. Valencia seems to be no exception. Caffa and Tana in the Black Sea were major enclaves for redistribution of Oriental slaves to the Mediterranean. Extreme poverty among Tatars, the biggest group, and war often led parents to sell children into slavery (Marzal Palacios, 2002).

Regardless of their origins, all slaves were seem to have been baptized and given a Christian name to replace the original. For instance, surviving documents detail the birthplace and names of African slaves that arrived to Medieval Valencia. Slaves were probably buried in Christian burial grounds. However, it is known that at least some slaves were branded in the cheeks with the letter S and a nail symbol or the word Sclau (both mean slave in Catalan). They could lead fairly independent lives or carry permanent shackles (Figure 33).

It is clear that the enslavement of indigenous Muslims and foreigners in the kingdom of Valencia was legislated and institutionalized, making the region not only a sink for international slaves but also a source of slaves. The institutionalization of slavery can be generalized to the rest of the Medieval powers in the Mediterranean, Christian or Muslim. This internationalization of the Mediterranean world explains the development of Sabir, a *lingua franca* used by Italian, Spanish, Berber, and other traders and sailors to communicate across the Mediterranean. Today is lost, and we know little about it but it must have been sufficiently important at the time for even Molière to know about Sabir and refer to it in his play *Le Burgeois gentilhomme* of 1670 CE. Perhaps, the character of Othello, a Moorish officer in the Venetian navy, in Shakespeare homonym work is the perfect reflection of the Mediterranean globalization of the Middle Ages.

To conclude this chapter, I would like to highlight something that sometimes can go unnoticed

Figure 33: Depiction of North African or Black sub-Saharan slaves in tile paintings from Valencia in the 18th century.

when we study this episode from an academic point of view: the human tragedy and trauma that enslavement meant, and still represents, for millions of people over the centuries. Even Medieval slavers themselves noticed and recorded a behaviour in their victims that they called *malaltia de mal de sentiment* (sickness of the bad emotion). There are two accounts of this sickness that often led the slaves to take their lives or jump into the sea: a 20 year-old African girl, named Beatriz, who in 1507 no one wanted to buy because she was fragile and sickly; and in 1510 five individuals brought from Benin that jumped off the ship (Cortés Alonso, 1972). Although devoid of any empathy, this clearly is an acknowledgement of the trauma captives endured when taken away from their homeland forever.

1.6 Prediction of Phenotypic Traits from Genomic Data

The genetic mechanisms behind physical features or phenotypes in any organism are complex because they involve several genes, and more often than not only a few associated functional variants are known (Figure 34). However, the characterization of multitude of human genetic variants in genome-wide association studies (GWAS) from the last decade has opened the door to the possibility of making inferences about phenotypes from DNA sequences (Claes et al., 2014; Adhikari and et al, 2016*a*,*b*; Claes et al., 2018; Adhikari et al., 2019) (Figure 36). GWAS have revolutionized the medicine and the search for the causes of genetic diseases, but in order to discover phenotype-associated variants large datasets consisting of a control and susceptible groups are needed and for analyses. The identification of these risk alleles are identified with GWAS allows to establish links between particular SNPs and the trait or susceptibility to certain conditions. However, the correlation identified it not so much about a SNP but a region in linkage disequilibirum (LD) that harbours the causal gene. This is why the functional correlation between a variant and the phenotypic effect is not always direct (Cole et al., 2017; Sero et al., 2019).

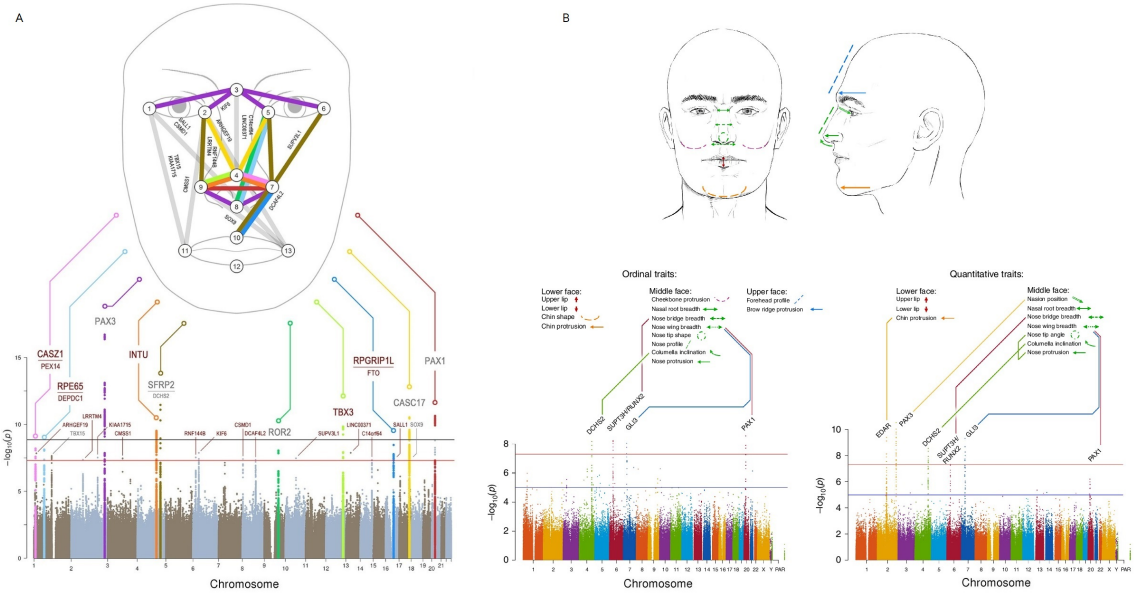


Figure 34: Genes related to facial variation (taken from Xiong et al. (2019) (A) and Adhikari and et al (2016*a*) (B).

Excluding genetic diseases, which is what GWAS are most intensively used for, some of the classic phenotypic traits in humans that have attracted the most attention of researchers are: persistence of lactase enzymes and the ability to digest lactose in adulthood, blood types, genetic height, pigmentation of eyes, hair and skin. Occasionally the traits are relatively simple and have been studied for decades, but sometimes the traits remain uncannily complex and we are only starting to scratch the surface of their synergistic molecular base. Notwithstanding, a genetic basis is not the only mechanism playing a role, environment is also important and its effect is not easy to quantify.

Among individuals of European ancestry, lactase persistence into adulthood (Figure 35) is a relatively simple trait ruled by the *LCT-MCM6* gene cluster (Liebert et al., 2017), other metabolic phenotypes can also be relatively well associated can be to single nucleotide polymorphism (Haber et al., 2016; Mathieson and Mathieson, 2018). However, because it is so simple, and the advantage acquired by the ability to digest milk as adults was advantageous in certain societies has caused this trait to evolve independently in different parts of the world at different times. The ability has evolve independently in Europe and Western Africa at least, meaning that mutations in different loci are responsible for lactase persistence. Therefore, independent genetic markers can be associated to lactase persistence (Mathieson et al., 2015), it is a convergent phenotype.

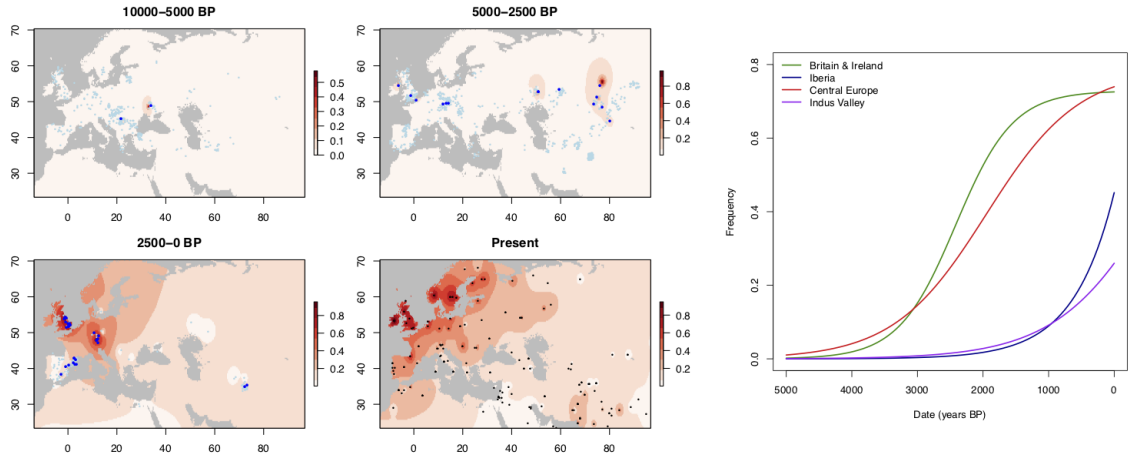


Figure 35: Evolution of the prevalence of lactase persistence among European populations through time. Taken from Ian Mathieson post on the spread of the European lactase persistence allele based on the frequency of SNP known as rs4988235. (http://mathii.github.io/2019/10/12/the-spread-of-the-european-lactase-persistence-allele)

Human blood type is another phenotype that although not so simple, it is at least, very well understood. Blood types classification in the ABO system is determined by the expression of particular proteins on the surface of red blood cells. The expression of such proteis does not change since it is coded in the genomic information of each person. There are two types of proteins that can be expressed or not, an depending on how the alleles combine, individuals can have genotypes AA, A0, BB, B0, 00, or AB. The expression of proteins is dominant and the no-expression (0) is recessive. For example, functionally two individuals with genotypes AA and A0, phenotypically both are type A. The original discovery of various types of blood groups in humans was a monumental step forward in medicine because it paved the way for safer blood transfusions mainly. Karl Landsteiner's characterized the ABO for the first time early in the 20th century, however, it was not fully complete until some forty years later when the Rhesus system was discovered. Although the ABO-Rh combination is the most commonly used blood classification today for humans, many other blood systems have been developed since. The ABO system is a good candidate for prediction of blood groups from DNA because there are a handful of known SNP markers in Chromosome 9 (positions 136132908, 136131650, 136131414 in Hg19 build of the human reference genome) that combined yield very accurate classifications. One of the SNPs, is an indel that determines the ability express proteins in the surface of blood cells and therefore is directly related to alleles of blood type 0. The genetic prediction using SNPs match accurately the results of serological tests that have been the standard procedure to identify ABO and Rhesus types for decades (Cassidy et al., 2016).

The cases of height and pigmentation are much more complex because it is well known that these have a polygenic genetic base and are highly continuous traits. Polygenic traits are those in which many genes intervene to build the phenotype associated. Perhaps because height and pigmentation are some of the most challenging phenotypes to reconstruct for the study of past populations, it why research on this topic attracts so much attention (Mathieson et al., 2015; Martiniano et al., 2017).

It is natural to feel curious about what an ancient individuals could have been like, and even to know if our looks have changed over time at all. In this regard, it is very likely that the work by Claes et al. (2014) represents the most ambitious attempt at trying to unravel the genetic basis of a complex trait from DNA. This work aimed to develop a model that allowed reconstruction of the most iconic aesthetic feature of a human individual: its face. However, despite the face being the main vehicle to convey information like emotions, intentions, health, sex, and age (Claes et al.,

2014), until very recently for example, there was no scientific understanding of how the width of the nose is determined.

In Claes et al. (2014) and Lippert et al. (2017) the modelling is able to reconstruct faces from DNA by taking into account facial variation, sex, genomic ancestry, genotypes, and the independent effects of particular alleles related to craniofacial features. They identified twenty genes implicated in facial variation, however, since then even more variants have been found and the models could be improved (Xiong et al., 2019; White et al., 2020) (Figure 36).

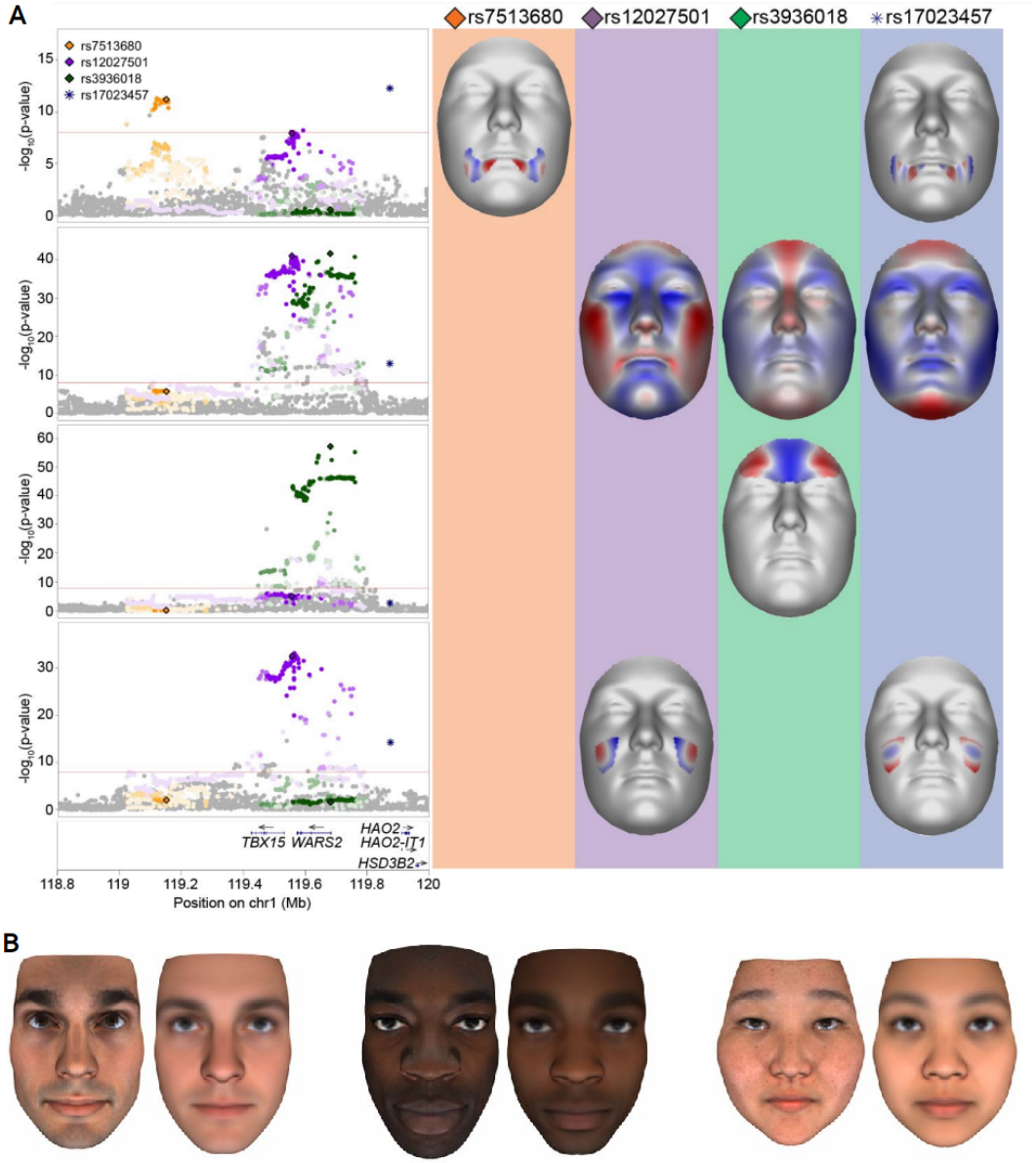


Figure 36: A) Facial effects of various SNPs associated to the TBX15-WARS2 locus regions (White et al., 2020). B) faces from DNA Examples of real faces (left) and reconstructed faces (right) from DNA and whole-genome sequencing data (Lippert et al., 2017)

As a cautionary note it is important to mention that GWAS have been traditionally heavily biased towards individuals of European origin. The most recent works regarding genetic variation related to facial features have made efforts to make use of multi-ethnic datasets but Europeans are still overepresented (Walsh et al., 2013; Xiong et al., 2019). Several recent GWAS studies have uncovered the relevance of genetic loci linked to facial phenotypes (Liu et al., 2012; Paternoster

et al., 2012; Adhikari and et al, 2016*a*; Pickrell et al., 2016; Cole et al., 2017; Claes et al., 2018; Cha et al., 2018; Evans, 2018; Li et al., 2019), but these well established genetic findings which include variants nearby or within the following genes: *CACNA2D3, DCHS2, EPHB3, HOXD, PAX1, PAX3, PKDCC, SOX9, SUPT3H*, and *TBX15*, have been identified mostly in Europeans. This is an important bias to keep in mind because some variants can be population specific for varying reasons. In other populations a different variant might have arisen evolved causing identical effects. It is crucial not to forget that SNP-to-trait associations that are true for Europeans might not necessarily be true for other natives groups from Africa, Asia or America.

2 Objectives

The aims of this project evolved during the time of my four years completing my doctoral thesis but they are connected through the use of aDNA to study the past and present genetic landscape of Iberia, and how to further apply new forensic approaches to aDNA. Originally I was in charge of generating a Spanish dataset consisting of more than 1000 complete mitochondrial genomes for phylogeographic purposes. For this, I had available a set of samples whose hipervariable region I and II had already been sequenced (Barral-Arca et al., 2016).

The second and main goal was to study the Bell Beaker phenomenon by sequencing ancient genomes from this period in Iberia. However, not having suitable ancient samples to work with from the beginning and the difficulty to obtain samples from collaborators associated to this particular material culture, and the publication of Olalde et al. (2018) about the Bell Beaker phenomenon made me decide to re-focus the topic of the ancient samples. I decided to downscale and only focus on the Mediterranean part of Iberia, what we call the Spanish Levant or the Valencian area. I also decided that it was better to start doing the sampling myself and get to know the collaborating archaeologists and Museum directors personally to improve confidence in the project and communicate better our research. Instead of focusing on one period I decided to sample over an extensive temporal transect for the region.

A third topic of interest was also developed over the first year of my thesis, regarding the possibility of phenotype prediction using genetic data, and so it was added to the final aims.

The following are the final aims that were developed over the first year of project while I was doing the sampling:

- To characterize the modern mitochondrial composition and diversity of the Spanish population as a whole and investigate whether regional differences may exist, by sequencing 1000 mitogenomes as planned originally.

- To sequence ancient genomes from Mediterranean Iberia through the Late Paleolithic to the Metal Ages, and from Roman times to the Late Medieval period. By studying genetic affinities, testing admixture and incorporating archaeological information, I hoped to reconstruct the population changes that occurred during prehistory and the Middle Ages in rural and urban settlements of the Valencian region in Spain.

- To investigate an alternative approach to current methods for predicting phenotypic traits using genomic data and machine learning approaches. More specifically, the main interest was to investigate how accurate predictions for traits such as skin pigmentation and blood types in high coverage ancient and modern genomes are.

3 Chapter I: 1000 Mitogenomes from Modern Spain

3.1 Introduction

The study of uniparental markers, Y chromosome and especially mitochondrial DNA (mtDNA), has been widely exploited over the years to investigate past migrations and other demographic processes in humans and other animals. From the moment the first mtDNA haplogroups (A, B, C, and D) were defined by Torroni et al. (1993) in Amerindians, the popularity of mtDNA became even greater than the Y chromosome because it was better understood. Mitochondrial diversity ever since has been successfully used to identify structure in populations within Eurasia and between different continental regions. In practice, this has become the basis of the traditional phylogeographic approach (Richards et al., 1998; Torroni et al., 2006).

However, since the advent of the second-generation sequencing revolution, whole genome data has displaced uniparental markers to a secondary role due to the amount of information that can be extracted from millions of genome-wide SNPs. Maternal and paternal lineages only inform us of a fraction of the ancestry. Nevertheless, research with Y-chromosome and mitochondrial DNA lineages is still useful to make inferences about population dynamics in the past, even more so in understudied species. What is more, uniparental markers are not only useful in evolutionary genetics of past and present populations, but they still have practical utilities in forensic and biomedical applications (Amorim et al., 2019).

At the onset of the molecular era in the 1980s and 1990s, when the earliest protein and genetic data started accumulating (Pereira et al., 2009), it became apparent that European gene diversity was modest compared to Asia, and very low compared to Africa. European populations are therefore very similar to each other in the broader picture of worldwide diversity. However, Europe has benefited from intense research about its past and that has built knowledge that has allowed researchers to reconstruct the transitional events between the Mesolithic and Neolithic, and between the Neolithic and the Metal Ages that forged its genetic pool (Haak et al., 2015; Cassidy et al., 2016; Olalde et al., 2018). From a combined genomic and mitochondrial perspective, the Neolithic is the demographic transition with a greater impact on the genetic diversity (Richards et al., 2000, 2016). Today we know that Neolithic migrants from the Near East added to the pool of all major European mtDNA haplogroups (HV, H, K, J, T, X) to the previous indigenous ones (e.g. U5a and U5b) (Torroni et al., 1996; Richards, 2003; Bramanti et al., 2009; Brandt et al., 2013).

Despite understanding the broader picture of the three-wave migration scenario, we still lack detailed ideas or models that explain in detail how the European mitochondrial diversity came to be, and what regional mechanisms best explain how mitochondrial frequencies shifted in one direction or another. For example, it is currently not known how H haplogroups became so frequent throughout Europe and Iberia (Hervella et al., 2015; Hernández et al., 2017; Olalde et al., 2019) by the end of the Copper Age when it was not so prominent amongst earlier Neolithic groups. The reason behind this explosive spread remains unanswered despite genomic and archaeological knowledge about the period. Although Iberia has been pointed as a possible source of H to the rest of Europe (Achilli et al., 2004), it is yet to be substantiated.

Usually only isolated ethnic groups present stark differences with the broader population in a given country, and such geographically isolated groups in Iberia have been under the focus of different studies (Maragato, Basque, Pasiegos, Romani) (Sánchez-Velasco et al., 1999; Larruga et al., 2001; Maca-Meyer, Sánchez-Velasco, Flores, Larruga, González, Oterino and Leyva-Cobián, 2003; Martínez-Cruz et al., 2012). Among the Basques for example, the frequency of H haplogroups is 5-10% higher than the Spanish and European average but other than that they have similar haplogroup composition (Cardoso et al., 2011, 2013). The Saami are an actual European outlier because haplogroup V reaches frequencies 40% whereas in any other known human population the frequency of V is low (below 10%). In some groups of Saami, V5 together with a sub-lineage of U5b constitute more than 80% of the mitochondrial diversity (Tambets et al., 2004).

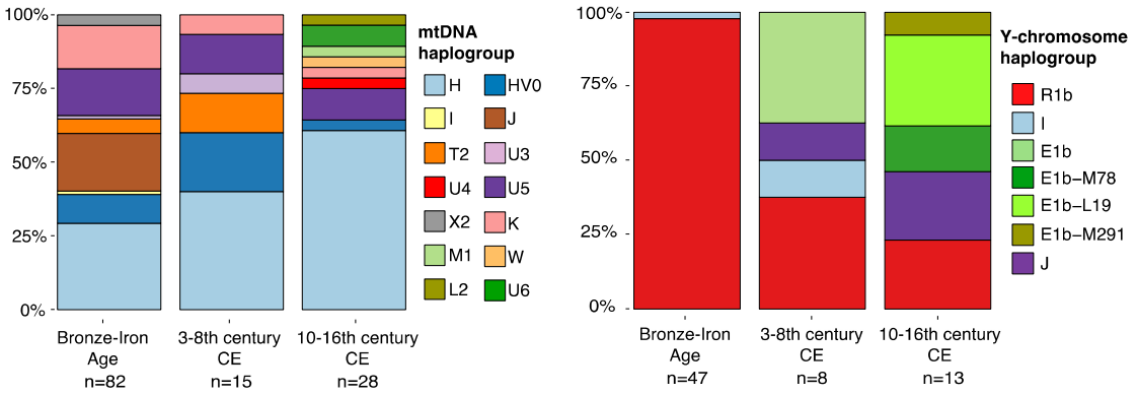


Figure 37: Evolution through late prehistorical times to historical times of the frequencies of uniparental markers in Iberia. The two right-hand columns for both the mtDNA and Y-chromosome represent Southeast Iberia and the left column is overall Iberia. Taken from supplementary material in Olalde et al. (2019).

In Iberia, the Basque and Spanish populations have been the subject of several of studies on mtDNA variation because the Franco-Cantabrian refuge hypothesis suggested that mtDNA diversity from human populations survived the Last Glacial Maximum there (Torroni et al., 1998). This is the reason why a majority of studies of Spain have targeted either particular haplogroups or populations in this northern region. Individual research papers have traditionally focused on interrogating geographically specific sub-haplogroups of the mitochondrial phylogenetic tree. In particular, many hypervariable region sequences (e.g. in Basques, Cantabrians and Asturians) have been published but far fewer modern full sequences of maternally inherited mitochondrial genomes outside the Basques have been sequenced (Cardoso et al., 2011, 2013; Pardiñas et al., 2012). This has resulted in a lack of of datasets of full mitogenomes built from a population genetics point of view.

For the case of the Iberian Peninsula, this issue has not been adressed yet but Barral-Arca et al. (2016) recently contributed numerous HVS sequences. Coupled with with extensive territorial sampling, over 3000 individual mtDNA control regions were generated.The sampling of the individuals covered all regions of the Iberian Peninsula, including for the first time important southern regions like Andalusia. However, even this relatively recent effort (Barral-Arca et al., 2016) has been rendered obsolete rather rapidly since the ability to sequence full mitochondrial genomes has now become both faster and much less expensive. The main drawback to HVS-I data is that not having full sequences limits both the resolution that can be achieved to classify haplogroups and the precision of molecular-clock coalescence time estimates, thereby limiting the inferences that can be made from partial datasets. The difficulty to classify in detail the haplogroups is especially relevant for H, because without mutations outside the control region, it is not possible to make an accurate determination to which sub-haplogroup of H, out of more than one hundred, a sample belongs (Soares et al., 2009).

Sample size is the other limiting factor for the discovery of differences in haplotype frequencies between groups. However, this was not a problem in the Barral-Arca et al. (2016) dataset. Given the large number of samples available from modern Iberian individuals an opportunity presented itself to fully sequence the mitochondrial genomes from Barral-Arca et al. (2016) and reanalyse the data to investigate patterns of geographic structure in much greater detail. This allows us to provide an accurate characterization of the mitochondrial diversity of the Spanish population in the 21st century.

3.2 Materials and Methods

3.2.1 Sampling

We selected a sample size of 1023 Spanish individuals for mitogenome sequencing from the 3024 individuals available in the Barral-Arca et al. (2016) dataset. The individuals of this dataset have confirmed pedigrees of Spanish origins since the parents and the four grand-parents are of Spanish origin. I used the provenance of the maternal grand-mother to assign a geographic origin to each mtDNA genome. The samples come from all 50 Spanish provinces and were collected by Antonio Salas team at the through an associated hospital. The DNA from the majority of the Spanish samples was already extracted and was provided DNA by Professor Antonio Salas, a collaborator from University of Santiago de Compostela (Spain). Another seven extra Spanish and Portuguese samples were extracted from buccal swabs of myself and other colleagues in our lab in Huddersfield for other purposes but were included in the dataset. We extracted these using the PureLink® Genomic DNA Mini Kit (Thermo Fisher Scientific). Finally, 103 extra Portuguese extracts were sent by Dr Teresa Rito (University of Minho, Portugal) and by Professor Antonio Brehm (University of Madeira, Portugal) at a later stage and also incorporated into the project. However, the sampling from Portugal is geographically biased towards the north, which represents more than than 80% of the Portuguese set.

The 1023 Spanish mitogenomes were amplified and analysed in cooperation with Marina Silva, a fellow PhD student in the group who was also in charge of the additional 103 Portuguese sequences. The re-sampling was designed to cover all regions of Spain but can be considered as a random sample of the modern Spanish population since it was not guided by HVS the pre-existing information from Barral-Arca et al. (2016).

In addition to the newly generated 1023 modern mitogenomes, I gathered all the ancient mitochondrial genomes from all periods in Iberia published to date (n = 379) and and built a comparative dataset with them (Günther et al., 2015; Mathieson et al., 2015; Olalde et al., 2015; González-Fortes et al., 2017; Lipson et al., 2017; Martiniano et al., 2017; Olalde et al., 2018; Fregel et al., 2018; Valdiosera et al., 2018; Olalde et al., 2019; González-Fortes et al., 2019).

3.2.2 Processing

The original long-range PCR protocol was tested and validated to work with half the volumes of reagents required to perform the PCR amplification. In order to reduce costs the current working volume is 25μL, instead of 50μL. I amplified the DNA extracts with PCR dividing the mitochondrial genome in two fragments of similar lengths (Table 3). PCR conditions were optimized for the specifications of the GoTaq® Long PCR Master Mix Kit (Promega), which includes a hot-start DNA polymerase (GoTaq® Hot Start Polymerase, Promega Corporation) in combination with a thermostable proofreading polymerase, that allows amplification of up to 30kb of DNA. PCR reaction contained 0.5 μL of template DNA, 11 μL of nuclease-free water, 1X GoTaq® Long PCR Master Mix (Promega), and 0.2 μM of each primer (final volume 25 μL). The PCR program consisted on an initial denaturation step of 2 minutes at 94°C, followed by 30 cycles (denaturation at 94°C for 30 seconds, primer annealing at 55°C for 30 seconds, and extension for 9 minutes at 65°C) and final step of 10 minutes at 72°C (Brandini et al., 2018) (Tables 1, 2 and 3).

The amplified PCR products were confirmed to be present with an electrophoresis agarose get at 1% (Cleaver Scientific Agarose, stained with Midori Green), then visualized in UV light with InGenius3 and using GeneSys 1.2.5.0 software, both from Syngene.

I performed the PCR product purification following instructions in the Wizard SV Gel and PCR Clean-Up System (Promega) protocol. Later I quantified the DNA product with Qubit 3.0 Fluorometer (ThermoFisher Scientific), using the Qubit® dsDNA HS Assay Kit (volume of DNA sample = 1 µL). The purified samples were then diluted to 1 ng/µL and both fragments pooled together in the same well, for a final volume of 40 µL per sample.

The 96-well plates (with 40 µL each of complete amplified mtDNA sequence for the 1023 individuals) were sent to the Earlham Institute (Norwich). Library preparation step was carried out at Earlham Institute using an optimised protocol based on Nextera® DNA Library Prep Kit (llumina, Inc). Libraries were then pooled and sequenced with an Illumina MiSeq machine targeting 200x coverage. I received the raw sequencing results as FASTQ files already demultiplexed (paired-end and fragment sizes of approximate 150 bp).

PCR Mix Reagents	**Volumes (μL)**
Nuclease-free water	11
GoTaq Long PCR Master Mix (Promega), 2X	5
Forward Primer (10 pmol/mL)	0.5
Reverse Primer (10 pmol/mL)	0.5
Template DNA	0.5

Table 1: Proportions of the PCR mix.

Stage	**Temperature**	**Time**
Denaturation	94°C	2 minutes
Denaturation	94°C	30 seconds
Annealing	55°C	30 seconds
Annealing	65°C	9 minutes
Extension	72°C	10 minutes

Table 2: PCR conditions for each cycle. These conditions are repeated for a total of 30 cycles.

Fragment	**Primer (position)**	**Sequence (5'-3')**	**Fragment size**
F1	5871for	GCTTCACTCAGCCATTTTACCT	7959 bp
F1	13829rev	AGTCCTAGGAAAGTGACAGCGA	7959 bp
F2	13477for	GCAGGAATACCTTTCCTCACAG	9438 bp
F2	6345rev	AGATGGTTAGGTCTACGGAGGC	9438 bp

Table 3: Summary information about the primers (forward & reverse) used in the PCR reaction to amplify both fragments for the complete mitogenomes.

3.2.3 Bioinformatic analysis

I used EAGER (efficient ancient genome reconstruction) pipeline (Peltzer et al., 2016) for many initial steps. I then aligned the reads to rCRS with BWA–MEM, which is optimized for long Illumina reads, and posteriorly identified PCR duplicates with DeDup (included in EAGER pipeline). I performed quality control of the alignment with QualiMap (v.2.2.1).

3.2.4 Data Processing

The raw data were pipelined into EAGER (efficient ancient genome reconstruction) (Peltzer et al., 2016), designed as tool to deal with big genomic sets of NGS data. I evaluated the FASTQ files with FastQC (Andrews et al., 2012) and removed adapters with AdapterRemoval (v.2.2) (Schubert et al., 2016), merged paired read files, and duplicated reads marked with Picard tools.

The mitochondrial reference genome of choice was the rCRS (Bandelt et al., 2013) and I made the alignment to the reference using the algorithm MEM from the Burrows-Wheeler Aligner (BWA) as implemented in EAGER. The quality of the final BAM files was checked with Qualimap, as implemented in EAGER by default.

I called mutations using GATK HaplotypeCaller and VCF files were created containing the individual haplotypes. The haplotypes were used to determnine the haplogroup classification through the online tool of Haplogrep2.0. For the heteroplasmies, a threshold of 30% was set, meaning that anything below the threshold is not considered as a mutation. This is an arbitrary threshold generally accepted as reliable and agreed by our group. However heteroplasmies were called for each sample and recorded in separate files, later added to the samples final haplotypes specifying their status as heteroplasmies.

3.2.5 Diversity Indices

To calculate the different diversity measurements for nucleotides, haplotypes and haplogroups, first I created frequency tables using R (with package sjPlot). Then, I imported the tables into PAST4 where I performed the various analyses (Shannon H, Simpson 1-D, Renyi and SHE) as instructed in the manual of PAST4 (Hammer et al., 2001).

The Simpson index is defined as 1-D, where D is dominance. D ranges from zero (when all taxa/categories are equally represented) to 1 (one taxon/category dominates the diversity completely). Simpson 1–D measures the *evenness* of the community from 0 to 1.

Shannon entropy is a diversity index that takes into account the number of individuals and taxa/categories. It can range from zero in communities with a single taxon to higher values when many taxa with few individuals are present.

SHE analysis computes logarithmic species abundance (ln S), Shannon index (H) and logarithmic evenness (ln E = H – ln S) for the first sample. Then the second sample is added to the first, and the process continues. If the samples do not come from a homogeneous population but instead a gradient or structured population, jumps in the curve can be used to infer differences.

The Renyi index relies upon a parameter alpha. When alpha = 0, the function gives the total species number. When alpha = 1 it represents the Shannon index and when alpha = 2 it behaves like the Simpson index.

3.2.6 Density Maps

The two-dimensional kernel density maps were made using the kde2d function from the R package MASS, then plotted on a map of Spain with ggplot and ggmap packages. We had 53 locations as sampling points of all the 1023 samples that were used as interpolating points for each haplogroup. The level contours do not represent absolute frequency in the dataset but relative occurrence of each individual lineage based on their presence in the 53 locations. Furthermore, the Sankey

and Chord diagrams were also plotted using the same data and geographic information using R (circlize and plotly packages).

3.2.7 Founder Analysis

For the construction of haplogroups phylogenies I used a maximum parsimony approach using mtPhyl software (http://eltsov.org/mtphyl.aspx), and made corrections to match with PhyloTree, the reference tree.

I used the founder analysis in house software and the phylogenies converted to XML format to estimate the timing of migrations. This method assumes a source and a sink population and tries to identify possible candidates for mitochondrial founder lineages. There are two criteria of stringency (*f1* and *f2*) which stipulate that that founder lineages have at least one (*f1*) or two (*f2*) derived branches in the source population.

Age of the migration was estimated with the *rho* statistic and an approximated linear mutation rate (2651 years for a mutation to happen) for the full mtDNA length due to constraints of the method with time-dependent rates. The scenario analysed here was: Near East and East Mediterranean into Iberia.

3.3 Results

A total of 1023 mitochondrial genomes from present-day individuals were sequenced in an attempt to provide the scientific community with a comprehensive dataset describing the mitochondrial diversity of Spain. The success rate for the total amount of mitochondrial genomes sent for sequencing was 98% success. Making this the biggest effort to characterize the Spanish mitochondrial genetic pool at a maximum resolution to date.

For the sake of simplicity, Table 4 presents how haplogroups and paragroups are distributed in modern Spain according to frequency and geography. These sequences have been so far included in the calculations to date the age of some important nodes of the mitochondrial phylogeny. The mitogenomes have been also used to build a large dataset for founder analysis in order to detect prehistoric migrations to and from Iberia and within Europe more widely when merged with other datasets, as shown below. A breakdown of frequencies into more specific sub-haplogroups and provinces is also available further below.

REGIONS	CATEGORY	TOTAL (n)	H1	H3	H5	H*	HV*	I*	J*	K*	L*	T*	U*	U5	U6	V*	X*	Other*
North	***Average***	***143***	***24.5***	***4.9***	***3.5***	***16.1***	***1.4***	***4.2***	***8.4***	***7***	***1.4***	***7.7***	***2.8***	***7***	***0.7***	***3.5***	***3.5***	***3.5***
Asturias	North	36	19.4	2.8	8.3	11.2	2.8	2.8	2.8	11.1	2.8	13.9	2.8	11.1	2.8	5.6	0	0
Basque Country	North	9	22.2	0	11.1	11.1	0	0	11.1	0	0	22.2	0	11.1	0	11.1	0	0
Cantabria	North	27	40.7	3.7	0	18.5	0	7.4	3.7	3.7	0	3.7	0	7.4	0	0	7.4	3.7
Galicia	North	44	18.2	6.8	2.3	18.2	2.3	2.3	15.9	4.5	2.3	2.3	6.8	2.3	0	2.3	4.6	9.1
La Rioja	North	5	20	0	0	20	20	0	0	0	0	20	0	0	0	0	0	20
Navarra	North	12	25	8.3	0	8.3	0	16.7	8.3	0	0	8.3	0	16.7	0	8.3	0	0
Central	***Average***	**396**	**20.7**	**7.3**	**3.5**	**16.5**	**2.2**	**1.6**	**10.1**	**6.3**	**1.1**	**10.1**	**4.8**	**7.6**	**0.8**	**3.8**	**2.5**	**2.4**
Aragon	Central	45	20	8.9	2.2	13.2	0	0	4.4	13.3	2.2	11.1	0	6.7	2.2	6.6	6.6	2.2
Castilla La Mancha	Central	98	26.5	9.2	4.1	16.1	2	3	10.2	4.1	2	12.2	4	4.1	0	1	0	1
Castilla y Leon	Central	168	22	5.4	4.2	15.6	1.8	1.2	10.1	7.8	0.6	7.1	5.4	8.9	1.2	3.6	1.8	3.6
Extremadura	Central	53	18.9	5.7	1.9	17.1	0	1.9	9.4	5.7	0	15.1	5.7	5.7	0	5.7	7.6	0
Madrid	Central	42	7.1	11.9	2.4	21.5	4.8	0	16.7	4.8	0	7.2	7.2	11.9	0	2.4	2.4	0
East	***Average***	***183***	***18.6***	***4.9***	***3.3***	***18.1***	***5.5***	***2.1***	***8.8***	***8.7***	***0***	***10.4***	***6***	***6.6***	***1.1***	***2.6***	***1.1***	***1.5***
Baleares	East	8	25	0	0	37.5	0	0	0	12.5	0	0	12.5	12.5	0	0	0	0
Catalunya	East	121	19.8	5.8	3.3	18.9	6.7	2.5	9.1	9.9	0	10.7	4.2	4.1	1.7	0.8	0.8	1.6
Valencia	East	54	14.8	3.7	3.7	15.1	3.7	1.9	9.3	5.6	0	11.2	9.3	11.1	0	7.5	1.9	1.9
South	***Average***	**289**	**18.3**	**6.6**	**3.5**	**16.3**	**2.6**	**0.9**	**7.2**	**8.3**	**3.4**	**7.6**	**4.4**	**7.3**	**3.1**	**3.6**	**2.4**	**3**
Andalucia	South	224	16.1	5.4	3.1	18.7	3	0.8	5.4	9.3	3	8	3.4	8.5	3.6	4.3	3.1	3
Murcia	South	65	26.2	10.8	4.6	9	1.5	1.5	13.8	4.6	4.6	6.2	7.7	3.1	1.5	1.5	0	3
West	***Average***	*96*	*19.8*	*7.3*	*1*	*15.4*	*3*	*1*	*6.3*	*9.4*	*3.1*	*14.6*	*4.1*	*4.2*	*3.1*	*3*	*1*	*3*
Portugal	West	96	19.8	7.3	1	15.4	3	1	6.3	9.4	3.1	14.6	4.1	4.2	3.1	3	1	3
Total	***Average***	**1107**	**20.1**	**6.4**	**3.3**	**17.2**	**2.9**	**1.8**	**8.6**	**7.6**	**1.8**	**9.6**	**4.7**	**7**	**1.6**	**3.6**	**2.3**	**2.8**

Table 4: Frequency of mitochondrial haplogroups and paragroups* per region in Spain. Haplogroup and paragroup categories do not refer to their position as ancestral root in the mitochondrial tree, instead they represent samples whose extant sub-haplogroups a some type of L(xM'N) (i.e. haplogroup L minus the non-African M and N haplogroups: L encompasses the whole modern human mtDNA tree), HV*(xH'V), M(xG'D), N(xA'X'R), R(xHV'JT'U).

Figure 38 shows how the mitochondrial major haplogroups are distributed across Spanish regions divided into North, South East and Central (see Figure 39). In Figure 38 it becomes clear that haplogroup H dominates the Spanish mitochondrial diversity in all regions, accounting for almost 50% of the total haplogroups, followed distantly by U (13%), J (9%), T (9%) and K (6%). Other haplogroups like HV*(xH'V), I, L(xM'N), M, N(xA'X'R'I), V, W and X occur at a frequency lower than 5%. Another fact that can also be appreciated is the great homogeneity of frequencies across all regions.

Diversity indices like Shannon H and Simpson 1-D reveal more intimate details about provinces and major areas of modern Spain. The maps in Figure 39 show the diversity indices calculated using all mitochondrial haplotypes found in the 1023 Spanish samples sequenced. When using both provinces and larger regions, the haplotype diversity appears consistently lower in the Northern region of Spain, whereas the Central and Southern region harbour slightly higher haplotype diversity. This is also true when using nucleotide diversity instead, although when using haplogroups frequencies no differences can be appreciated (Figures S1, S2, S3, S4, S5 in Supplement I).

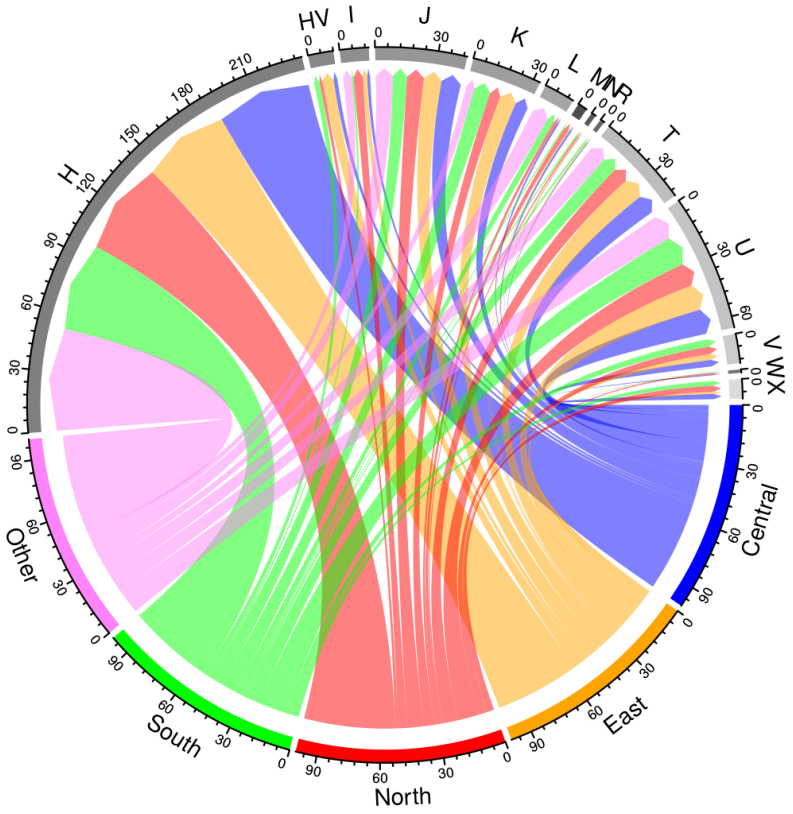


Figure 38: Proportions of haplogroups in each region of Spain and how the contribute to the common genetic mitochondrial pool.

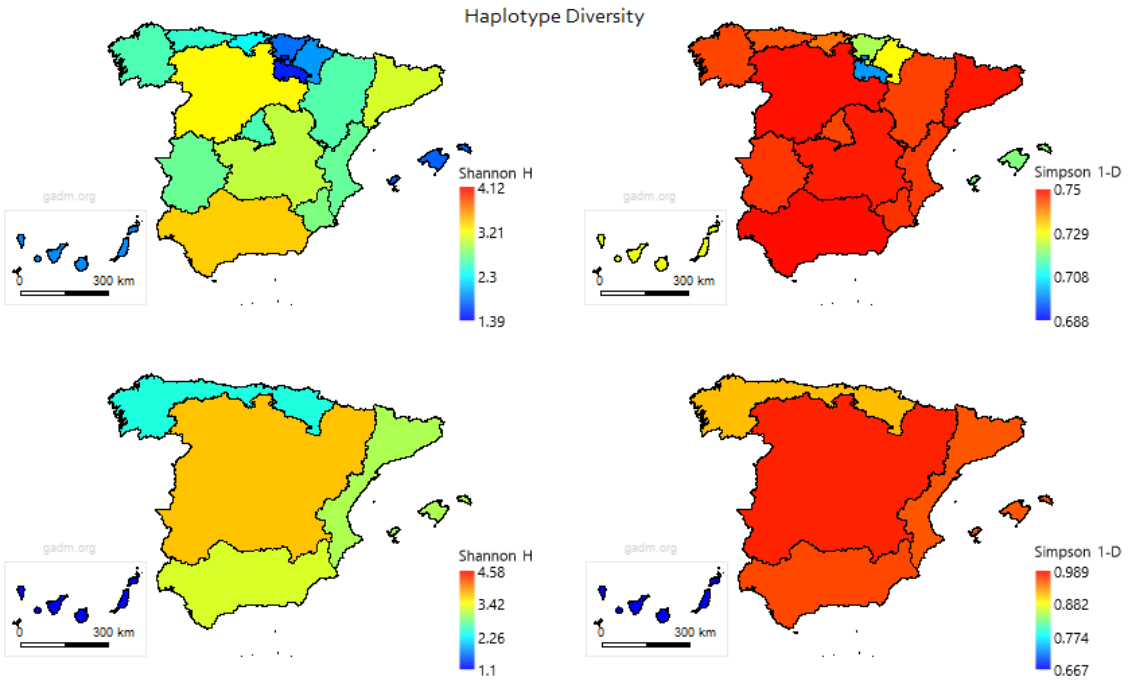


Figure 39: Haplotype diversity in Spain, as measured with Shannon Index (H) and Simpson Index (1-D), for a division according to Autonomous Regions and a broader geographic grouping.

3.3.1 Density Maps and Modern mtDNA Diversity in Spain

The distribution and density of major haplogroups (H, HV(xH'V), J, I, K, T, U, V and X) in Spain appears to be fairly homogeneous and does not reveal population structure, however this could be due to the lack of sufficient sample size of specific sub-haplogroups (Figure 41).

A breakdown of H into its main sub-haplogroups (H1, H2, H3, H4, H5, H6, H7, H9, H10, H11, H13 and H20) shows possible evidence of a west-east divide pattern (Figure 41). However, this could well be due to a sample size issue because some of these sub-haplogroups, with the exception of H1, H3 and H5, are not very common.

Haplogoup H was found to have 47 sub-haplogroups in this dataset out of the total of around one hundred sub-branches found worldwide. H1 represents one in every five H lineages in the dataset, accounting for 20% of the total frequency, making it the most common haplogroup in Spain. In a distant second and third position there are H3 and H5 with 6.5% and 3.3% frequencies respectively. After that, H2, H4 and H6 are the only other H lineages to have a prevalence greater than one percent in Spain. The remainder of the composition of the HV branch (types of HV sub-haplogroups, plus other V sub-haplogroups), of which H is the major sub-clade, only accounts for 2.5% of the population.

It was not possible to break down other haplogroups into their sub-clades to make maps because the count of each sub-haplogroup was most of the times very low (few occurrences). Such sample size would produce maps representing an artefact in the density distribution of a lineage. That is actually the case of H9 and H10, which are very scarce in the dataset and only present in one or two interpolating locations.

However, it possible to detail the frequencies of the branches of the important haplogroups in Spain as a whole. For examples for J, I found that the prevalence of J1 is 6.6% and for J2 is 2%. In the case of haplogroup K, K1 represents 6.2% and K2 1.4%. For T, T1 is 2.2% and T2 is 7.4%. Finally, U5b is the most common within U with a 5.4% but there are also other lineages, all in frequencies higher than 1% such as U5a (2.1%), U2 (1.2%), U3 (1.2%), U4 (1.3%) and U6 (1.6%).

The behaviour of haplogroup X could be explained by sample size, since it is found across Spain but is rare. Most likely differences between main regions do not exist. Intra-exploration of X shows that X2b accounts for 50% of X lineages in Spain and occurs at a frequency of 1.2%. In this dataset its distribution peaks in the southwest. I also characterized other rare X samples from other sub-branches like three X1c and another three X3a.

The only two major haplogroups, albeit ones that are low in frequency, that display a clear geographic patterns are U6a (1.6%) and paragroup L* (including L1b, L2a, L2c, L3b, L3d, L3e, L4b: 1.8%) (Figure 40), whose phylogeographic origins are in Africa. The L* label here indicates all the diversity of this umbrella paragroup L. This is only for practical purposes since individual L haplogroups are also very scarce in the dataset. From a phylogenetic point of view, lineages like L1, L2, L3 are actually highly divergent and diversified. The patterns reflected in Figure 40 (where I also included the Portuguese data) show that L and U6a lineages in the south are more commonly found among modern inhabitants in the Spanish south.

U6a peaks in the southwest, in the regions of Seville and Algarve, and a decreasing gradient is observable as the latitude increases, or in other words, U6a becomes rare when the distance from North Africa increases. On the other hand, U6a is virtually absent in the eastern part of the Iberian peninsula.

L haplogroups as a whole also peak in the south of Spain. However the centre of gravity does not overlap that of U6a since for L* it appears to be skewed towards the SE instead. In the rest of the Iberian peninsula L haplogroups are even less common than U6a, although there are two patches in the territory shared between southern Galicia with northern Portugal, and in southern Asturias.

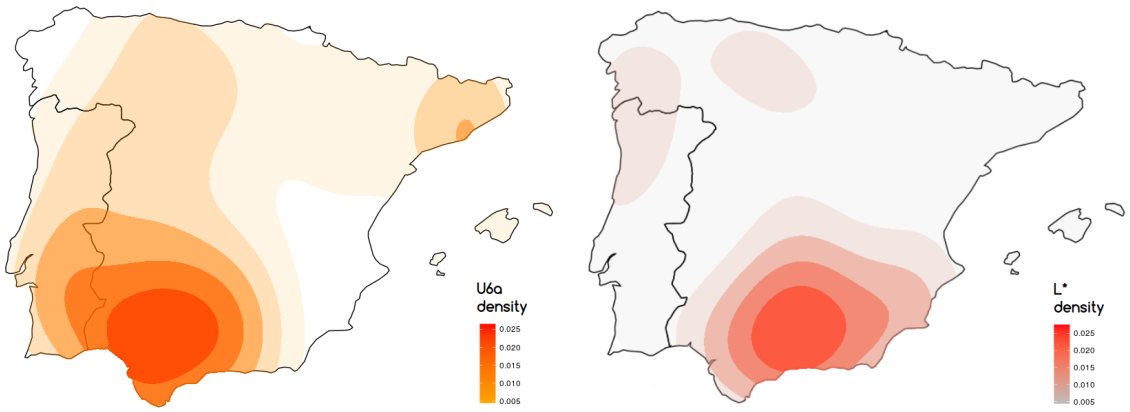


Figure 40: Two-dimensional kernel density estimations in a grid for minor and rare haplogroups.

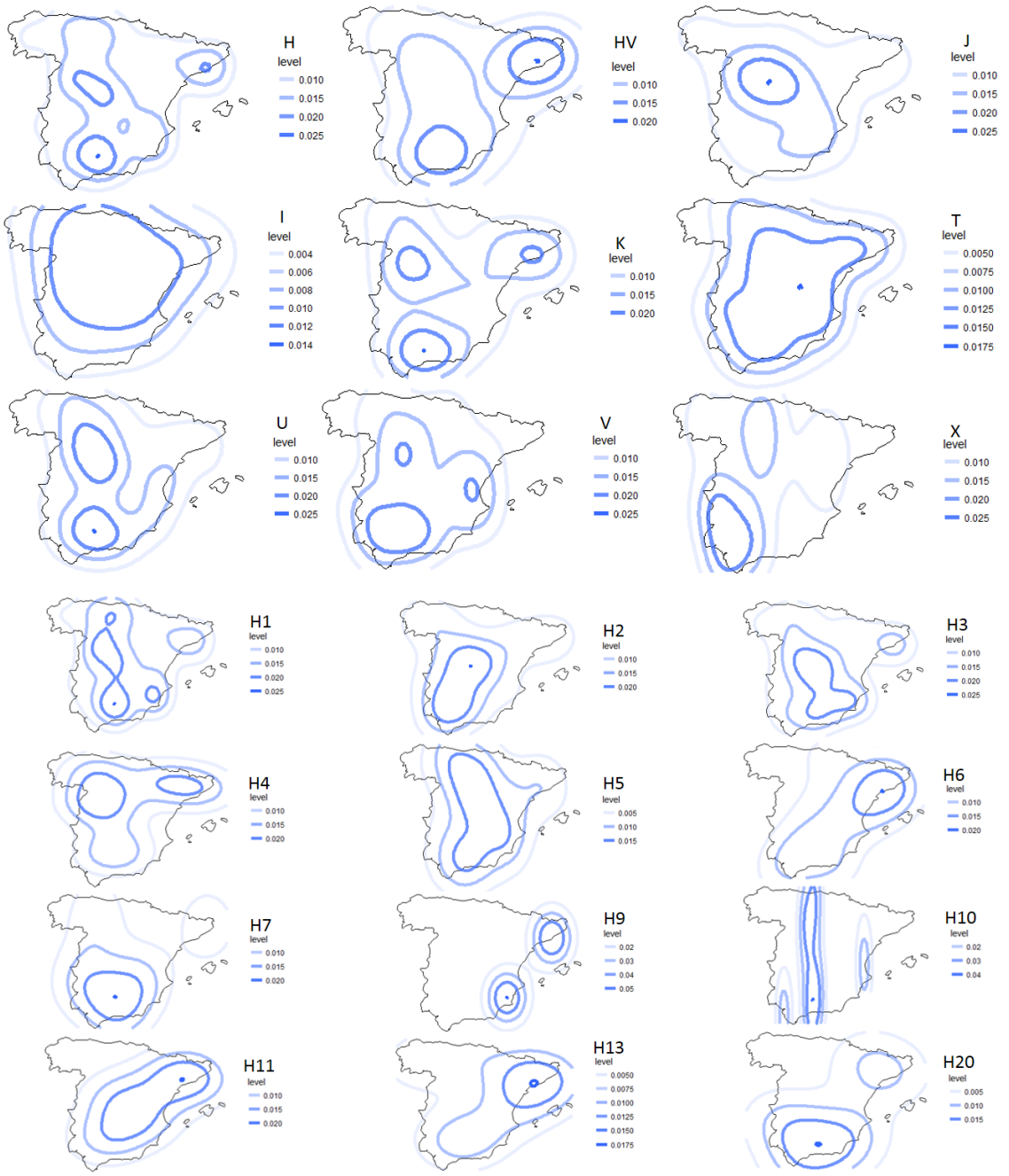


Figure 41: Two-dimensional kernel density estimations in a grid for all major mitochondrial haplogroups in Spain.

3.3.2 The Roots of Mitochondrial Diversity in Spain

Figure 41 represents the apparition over time of every single one of the 379 ancient mitochondrial genomes from Iberia published to date (Günther et al., 2015; Mathieson et al., 2015; Olalde et al., 2015; González-Fortes et al., 2017; Lipson et al., 2017; Martiniano et al., 2017; Olalde et al., 2018; Fregel et al., 2018; Valdiosera et al., 2018; Olalde et al., 2019; González-Fortes et al., 2019), plus 20 new mitogenomes reported in this thesis. The earliest haplogroups to appear in the genetic fossil record are U lineages (U5b more specifically) which are those typically found among the native hunter-gatherer inhabitants of Iberia.

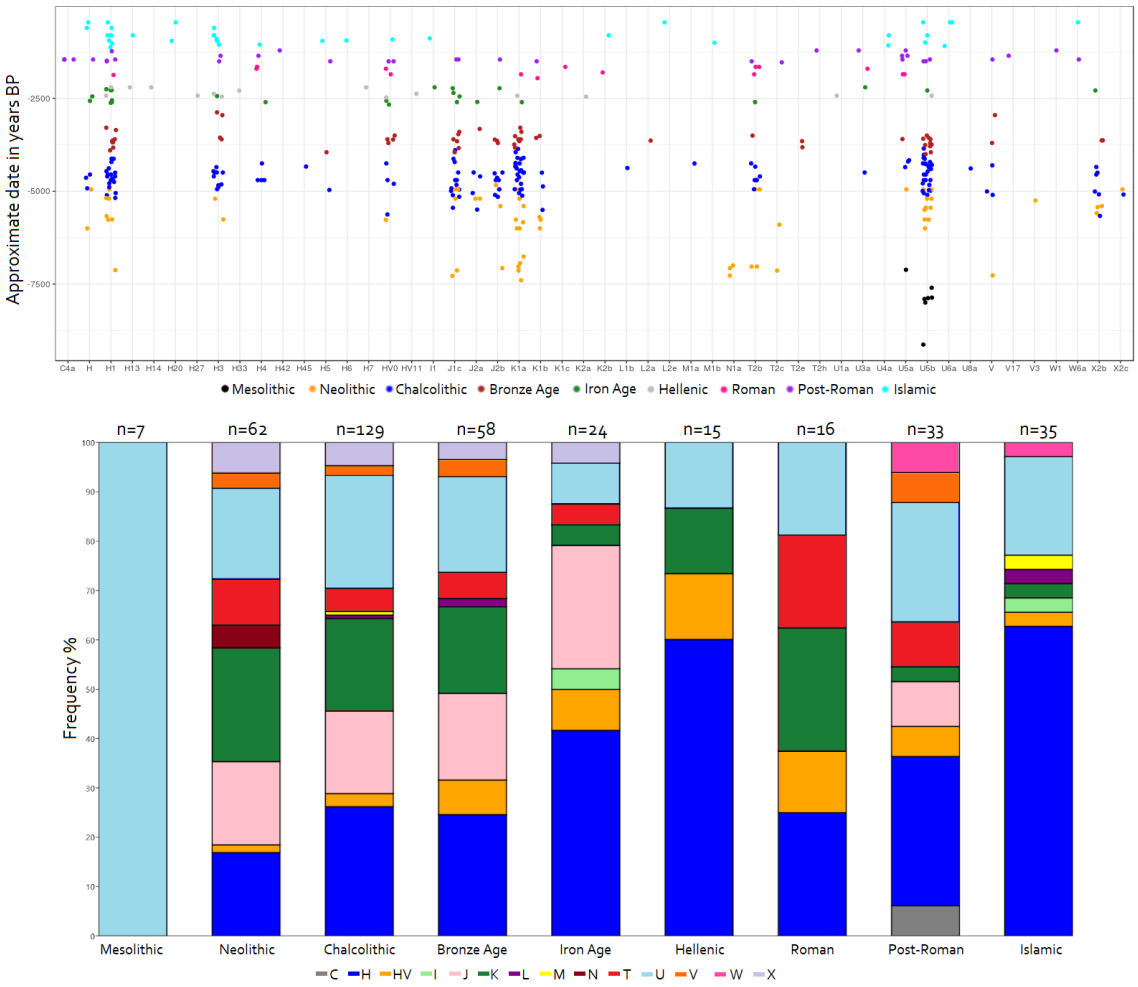


Figure 42: Timeline with 379 ancient mtDNA haplogrups from Iberia (top). Stacked bars with the relative frequency of each haplogroup per period (bottom).

In the depiction of mtDNA haplogroups segmented by period (Figure 42) U lineages (represented by six mesolithic U5b and one U5a mtDNAs) experience sharp drop in frequency because they go from being the only among hunter-gatherer samples to contribute only 10% to the newly established Neolithic society (11 out of 62 Neolithic mitogenomes are U5b, and there is only one U5a). However, evidence of U5b or U5a in Neolithic individuals does not manifest until the middle Neolithic, around 5500 years BP. This represents a hiatus in the datatset of almost 2000 years since the latest observed U5b haplogroup found in a hunter-gatherer individual from Iberia (Figure 42).

Among the early pioneer Neolithic individuals from Iberia around 7500 years before present, haplogroups K, J, T and N (specifically N1a, which used to be very common among individuals of the LBK culture (Haak et al., 2005)) appear for the first time. In Figure 42 there is evidence of

early presence (circa 7500 BP) of H and V lineages but neither seems to have consolidated at that time. Interestingly, mitochondrial N lineages are part of the early Neolithic package that arrive to Iberia only to disappear completely from the record after the pioneer colonization. Contrary to the trend of haplogroup N1a, haplogroup H (mainly H1 and H3 but not H5 although it is the third most common H lineage in modern Spain) goes from very minor to an explosion in density and frequency around 5000 years ago during the Copper Age (Figure 42). This is also the Megalithic period and the time of the Bell Beaker phenomenon.

Much like the results in Szécsényi-Nagy et al. (2017) working with ancient HVS data, the results here also show a quick rise of haplogroup H in the Neolithic-Chalcolithic transition in the whole peninsula. However, the increase is not so dramatic in the southwest where H appears already consolidated during the Neolithic according to Szécsényi-Nagy et al. (2017). This also fits the results of the six new Neolithic/Chalcolithic genomes from the east sequenced in this thesis because I found one H1 and three H3 haplogroups (see table in Supplentary Material for details on the haplotypes). This might suggest a mid to late Neolithic maritime introduction of H1 and H3 in east-southeast Iberia from where it expanded across the peninsula later on. Furthermore, N1a also disappears from the Szécsényi-Nagy et al. (2017) dataset in the Chalcolithic.

Haplogroup X is another Neolithic maternal lineage that today is a rare but wide-spread haplogroup in Europe. Fernandes et al. (2012) argue a Late Glacial spread but this contrasts with aDNA evidence. It is a basal N lineage and appeared in West Asia some 30-20 thousand years ago but only seen amongst Europeans when the Neolithic farmers arrived (Mathieson et al., 2015). The global distribution of X is interesting because it can be found in west Eurasia and Native Americans but is absent in East Asia, raising questions about how it could have spread (Richards et al., 2000; Reidla et al., 2003; Fernandes et al., 2012). In the Iberian aDNA record only appears by the mid-late Neolithic, coinciding with the explosion of H, although the founder analysis that will be explored later points to an early Neolithic introduction around 7200 years BP.

X2b is the most common clade among modern Spaniards but it is also widespread throughout Europe, consistent with general a Neolithic dispersal from the Near East. X2b consolidates by the Chalcolithic in Iberia but its frequency wanes in the following millennia. Other sub-haplogroups like X1c and X3a that are rare but well established in modern Spain do not appear in the ancient dataset.

Other haplogroups that appear as early as H in the record are K, J and T, but their behaviour through time differs from that of H1 and H3. All three haplogroups are introduced in Iberia during the Neolithic in frequencies between 20 % and 5% (Figure 42).

K1a arrived with the earliest Neolithic pioneers to Iberia and is the only haplogroup together with U5b that remains a constant relevant haplogroup since the early Neolithic and during later periods until the present day (Figure 42). K1b follows a similar path albeit it is found in much lower frequencies. Nonetheless, by the Iron Age, K haplogroups will suffer a great reduction in relative abundance, the apparent recovery in Roman times might be an efect of sample size or bias due to migrants from the eastern Mediterranean.

J lineages J1c, J2a and J2b, only become common in the early Chalcolithic. Although there is evidence of two early Neolithic J1c sequences neither of the three haplogroups appears during the rest of the Neolithic period in the records of this dataset Figure 42).

T2b and T2c first occurrence in Iberia is in the early Neolithic although they also do not seem to have been particularly relevant during the period. Later on, only T2b consolidates and becomes constant in the record while T2c completely disappears. It also appears that there was a contribution of T2e lineages in the early Bronze Age, perhaps linked to migrants with steppe or Yamnaya-related ancestry, although other known Bronze Age mitochondrial expansions of T1a, I3 and I4 lineages did not reach Iberia at that time (Figure 42).

Unfortunately, it appears that there is a gap in the record around 6250-6500 years ago and the Neolithic transect is not as well documented as the transition to the later Copper Age period (Figure 42). On the other hand, the transition to the early Bronze Age seems more complete.

There is another gap during the Bronze Age because funerary practices shift towards cremation at around 3000 years BP 42. Unlike the case of the paternal uniparental markers and appearance of R1b, there is no signal for the introduction, enhancement or decline of any mitochondrial macrohaplogroup associated to the arrival of the Indo-European influence to Iberia between 4500 and 4000 years ago.

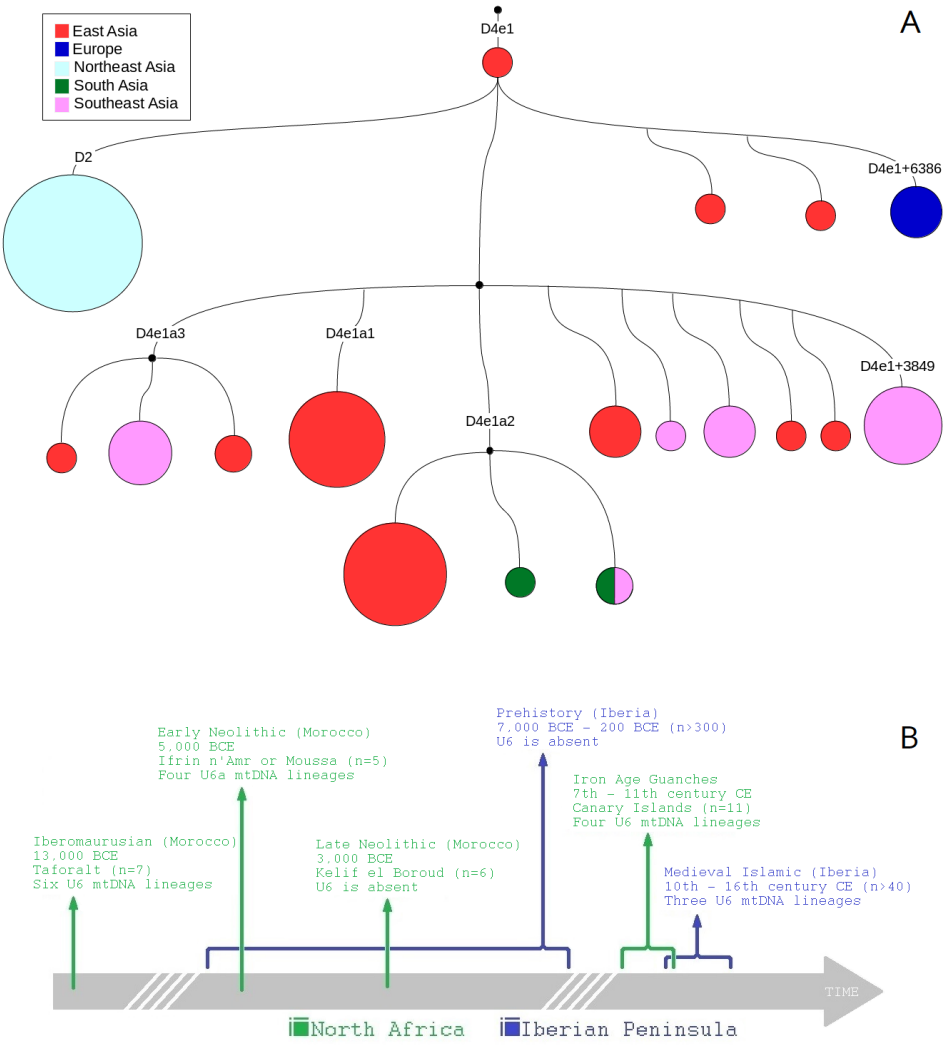


Figure 43: Phylogeographic network-tree of D4e1+6386 branch (A). Timeline of recorded evidence of mitochondrial lineages of U6 found in Iberia and North Africa (B).

From the Roman period onwards, exotic mitochondrial lineages start to show up. For example, two C haplogroups found in two related individuals of Germanic origin found in Catalonia (Olalde et al., 2019). Haplogroup D of East Asian origin, found in one of the ancient samples sequenced here (D4e1+6386), also appears in a Roman sample. Closest evidence of D4 is in prehistoric Central Asia (e.g. Xiong-nu). Other haplogroups like W and I also start to make their first appearance in the record in post-Roman times (Figures 42, 43A).

From the moment of the Islamic conquest, haplogroups belonging to L and U6a acquire a presence in Spain. U6a is a particularly good marker to measure novel Berber contribution since it had never been present in the Iberian peninsula until the Middle Ages despite being being extremely common in North Africa (Figures 40, 43B).

The founder analysis (Figure 44) of the mitochondrial phylogenetic tree of Spain mirrors much of what the timeline of the fossil record in Figure 42 shows. According to this founder analysis that

took the Near East and the eastern Mediterranean as source and Iberia as sink, the vast majority of mitochondrial lineages in Spain trace their origin in time to the Neolithic. Around 80-85% of mtDNA lineages have their coalescent origin between 5000 and 10,000 years before present; that is, the duration of the extended Neolithic period in Europe.

The height of the curve of the founder analysis reveals (with both *f1* and *f2* approaches) one clear peak matching the accepted dates for the earliest evidence of Neolithic colonization of Iberia around 7500 years (Figure 44), mtDNA evidence of the time of the pioneer Neolithic colonization.

The older lineages correspond mainly to U5b since it was carried in Europe among the survivors of the last Ice Age. The founder analysis indicates the coalescent time for a sub-haplogroup in Iberia, like U5b1c, is closer to 12,000 years BP. However the founder analysis also suggest pre-Neolithic coalescent times for specific haplogroups like H, K1b2 and J1c2e among few others, a results that does not find any backing in the aDNA record. Also, the founder analysis does not reflect the fact that U5a is recorded in Mesolithic Iberia albeit in very low frequency.

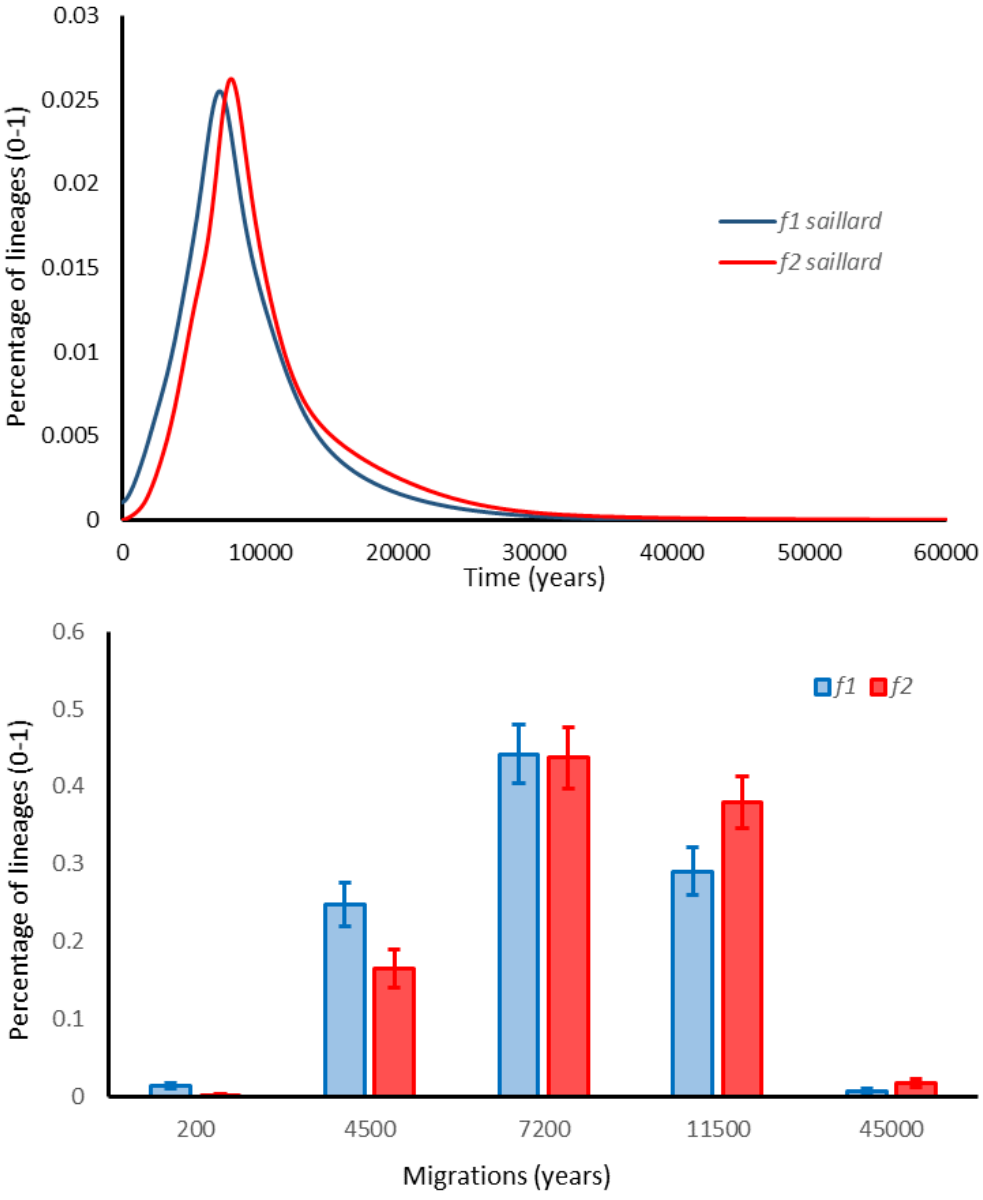


Figure 44: Founder Analysis with the Near East and Eastern Mediterranean as source and Iberia as sink. The guide tree used to build this figure contained publicly available samples (every European haplogroup) from the relevant regions.

3.4 Discussion

Possibly the most striking result I obtained here is the consistent evidence suggesting that the mitochondrial diversity both in the shape of haplotype and nucleotide diversity is lower in the northern part of the Iberian Peninsula. Although this is not entirely new result, since it has been reported before (Barral-Arca et al., 2016), it is the first time to my knowledge that has been calculated with such a big dataset of complete mitochondrial genomes.

From a mitochondrial perspective, Iberia as a whole appears to be a composition of maternal haplogroups that for the major part arrived some 7000 years ago, with the exception of U5a and U5b which were already present since the Ice Age. All the major mtDNA haplogroups that are important today in Spain appeared for the first time together with the archaeological evidence of arrival of agriculture to Iberia. These haplogroups (H, K, J, T) kept prominence and have never disappeared from the record and were also already present by the time of another important male-dominated population reshaping occurred with the advent of the Metal Ages. The still unexplained rise to prominence of haplogroups H1 and H3 in the Iberian Chalcolithic is remarkable. Although already noted in previous works (Szécsényi-Nagy et al., 2017), it has not been possible to link this phenomenon to any archaeological culture, event or expansion after it became evident that the Bell Beakers were not apparently responsible for it (Olalde et al., 2018). It remains a scenario to be tested the idea that H1 and H3 hitch-hiked with the expansion of the Megalithic culture, especially now that it is known it involved some long-distance familial connections, in Ireland at least (Cassidy et al., 2020). Haplogroup X is another lineage introduced in the Neolithic fow which I reconstructed its global phylogeny. Interestingly, when looking at the tree phylogeny, both X1c and X3a appear as Mediterranean-exclusive clades dominated by Spanish individuals. However, looking at the phylogeny of X2b which is the most important sub-haplogroup, Spanish individuals do not form any Iberian specific sub-clades.

Nevertheless, the Spanish maternal diversity has remained broadly uninterrupted for thousands of years since the Neolithic. This is supported by the founder analysis since not significant post-Neolithic contributions are observable.

It is true however, that the mitochondrial profiles of the Hellenic, Roman and post-Roman periods presented here might be in contradiction to this claim but this is likely due to the small samples sizes of these periods in the datasets. Presence of direct foreign highly mobile migrants from places like Greece, Italy an North Africa from non-randomly chosen archaeological contexts as shown in Olalde et al. (2019) and Antonio et al. (2019) affects frequency proportions too.

As mentioned above, the arrival of Indo-European related male lineages at the start of the Bronze Age does not seem to have had an effect, at least in a significant way, in the mitochondrial composition of the peninsula. This observation is very much in line with the idea that the Steppe-related migration into Europe was a male mediated migration (Haak et al., 2015). The conclusion that Indo-European steppe-related migration did not involved significant number of mtDNA differentiated females is backed by the fossil record of mtDNA haplogroups.

Whatever is the scenario, even with the current enhanced samples sizes it not possible to make meaningful mitochondrial diversity measurements for the area division of the peninsula into Central, North, East and South. Furthermore, this period is probably the key to explain why H rose to such high prevalence in Iberia and elsewhere in Europe. In ancient Bell Beakers from Central Europe the frequency is also around 50% and this is interesting because H samples prior to the Chalcolithic there are at much lower frequencies (<30%) (Brandt et al., 2013; Brotherton et al., 2013; Szécsényi-Nagy et al., 2017). However, it is still not clear to me how the current distribution of H across Europe came to be. Although Iberia becomes more and more the main suspect to be a putative source, especially haplogoups H1 and H3 as suggested in the past (Achilli et al., 2004; Pereira et al., 2005; Soares et al., 2010), haplogroup H1 dominates modern Spanish mitochondrial diversity along with H3 and both appear to start rising by the onset of the Chalcolithic.

The mounting evidence that European hunter-gatherers only carried U5a and U5b lineages has, however, discredited the idea that although H most likely arose in Southwest Asia (Richards et al.,

2000; Roostalu et al., 2007) it could have arrived before or around the Last Glacial Maximum, survived in Southwest Europe and then re-expanded (Achilli et al., 2004; Zheng et al., 2012). This was an hypothesis initially backed by HVS evidence for presence of H in pre-Neolithic Iberia (e.g. El Miron in Hervella et al. (2014)) but in fact this was result of contamination.

As I shall discuss below, later historical events in the peninsula have probably blurred the picture complicating the interpretation with modern data. Evidence for this phenomenon is the sprouting of more very rare haplogroups, many of Asiatic origin as indicated by the phylogeograpy, when the Mediterranean had turned into nothing but a highway linking east and west.

Mitochondrial haplogroups along with recent evidence recovered from genomic data have informed us that the genetic landscape of Iberia in historical times underwent a reshaping to a considerable degree - perhaps not the whole territory since the north is fairly isolated but at least the east and the south since they are the regions most exposed to Mediterranean and North African influence.

The contribution of mitochondrial diversity by other Mediterranean and North African groups is hard to detect without a detailed research into haplotype and clades in the tree. However, the apparition of unseen haplogroups in historical times suggests exposure to foreign mitochondrial influences as suggested by Figure 42. This can be one of the drivers behind the higher haplotype and nucleotide diversity in the non-Northern regions of Iberia.

The distinctness in the Franco-Cantabrian could be, in part at least, related to this isolation from historical events since the Iron Age. The survival of Basque language after Latin implantation in Iberia is a support for this idea. A traditional and persisting smaller and more fragmented population structure until the Industrial Revolution is also likely responsible to the rise of mitochondrial haplogroup H by sheer force of drift to an ever higher frequency among the Basques for example. By Ockham's principle, this saves the need to invoke a more remote explanation of this phenomenon, at least until ancient data becomes available from the heartland of the Basque Country (Cardoso et al., 2011, 2013).

There are several haplogroups that can be used to exemplify the mitochondrial enrichment of Iberia in Roman and Islamic historical times. One such haplogroup is U6a whose southern distribution is very pronounced. The U6a lineage traces its origin to North Africa without a doubt (Maca-Meyer, González, Pestano, Flores, Larruga and Cabrera, 2003; Secher et al., 2014; van de Loosdrecht et al., 2018). Whatsmore, it reaches very high frequencies (sometimes over 80%) in Berber populations native to the North African desertic hinterland, which is striking when compared to any other ethnic group. To date, no evidence for presence of U6 in Iberia has been found during prehistory or antiquity among over 400 published samples. In fact, the first ancient U6a haplogroup to be found in Spain is a medieval sample from an 11th century CE Islamic necropolis from Segorbe reported here (Barrachina Ibáñez, 2005), as well as some later ones from Olalde et al. (2019). This is very likely an indication that U6a was introduced in Spain during the Islamic conquest by the Berber soldiers and settlers coming from North Africa. The modern peak occurs in the SW of Spain, an area that suffered less from the deportations of Moriscos (converted muslims, formerly called Mudejares) in 1609, compared to the SE and the east of Spain. The reason for this is that the east of Spain harboured the majority of Moriscos of the Crown of Aragon, where the Moriscos represented a third of the population and kept their own communities and religious traditions. However, they were almost completely expelled and the area repopulated by Christians from the north, consistent with the lack of U6a in both those areas.

The other example that mirrors U6a among rare haplogroups found in Spain is the paragroup L(xM'N). I found a distinct pattern for L, and both L and U6 lineages have clear African origins. L(xM'N) lineages are mostly found in sub-Saharan Africa and therefore they represent markers of migrations from Africa to the Iberian peninsula. In our case we found that L(xM'N) haplogroups are predominatly restricted to the south of the peninsula, which indicates a movement across the strait but no further. The area roughly coincides with the restricted territory held by the last Islamic powers centered in Granada. Given what is observed here, Islamic kingdoms might be responsible

for the introduction of such haplogroups via the trans-Saharan slave trade, because at the decline of their rule it became difficult to capture Christian slaves. There was a significant degree of social mobility for Islamic slaves which leaves the door open for possible assimilation in the South. The northern patches of higher frequency in Spain can be related to the Pasiego folk, a marginalized historical etno-cultural group that also shows high prevalence of U6 amogst them (Maca-Meyer, González, Pestano, Flores, Larruga and Cabrera, 2003; Rodríguez-Varela et al., 2017).

My closing remark is that Iberia seems to have been a rather homogeneous entity from a mitochondrial perspective throughout prehistory regarding the distribution of the major haplogroups such as H1, H3, U5, K1, J1 and T2, but there are regional dynamics in the north, west east and south that have generated local patterns over time. Such is the case of L(xM'N) haplogroups and U6a in the south. In the north, the absence of further genomic or archaeological evidence seems to suggest that the Franco-Cantabrian relative uniqueness has probably been driven by drift in the last 3000 years due to isolation from historical processes in Antiquity and Medieval times, a topic long debated in the literature over decades (Bertranpetit and Cavalli-Sforza, 1991; Achilli et al., 2004; Cardoso et al., 2011; Behar and et al, 2012*b*; Cardoso et al., 2013; Valverde et al., n.d.). However, some questions and unknowns remain, such is the case of the mysterious expansion of mitochondrial haplogroups H1 and H3 for which I feel have not been able to present a definite answer in this work with the current data.

In any case, here I have presented a unique snapshot of the modern mitochondrial diversity of Spain that will provide a go-to reference resource for future researchers.

4 Chapter II: Ancient Genomes from Mediterranean Iberia

4.1 Introduction

The past of the Iberian Peninsula has been the subject of intensive research through ancient DNA in the last few years. The earliest publications focusing specifically on Iberia like Sánchez-Quinto et al. (2012); Olalde et al. (2014, 2015); Günther et al. (2015) provided the first insights into ancient genomes from the Mesolithic and early Neolithic individuals.

Later work with samples from different locations in Portugal or Spain produced more Neolithic genomes and expanded into the Metal Ages. These samples have been studied in Martiniano et al. (2017); Lipson et al. (2017); González-Fortes et al. (2017); Olalde et al. (2018); Fregel et al. (2018); Valdiosera et al. (2018); Zalloua et al. (2018) and González-Fortes et al. (2019).

These new genomes have helped to further expand the knowledge about the population turnover that occurred at the transition between the Copper and Bronze Age. Long-held ideas suggested by archaeology were finally started to be confirmed by genetics: migrants related to people from the Eastern Mediterranean came to Iberia through a maritime route.

By 2018, there were over one hundred genomes from Iberia available covering the Neolithic, Chalcolithic and early Bronze Age of Portugal and Spain. Martiniano et al. (2017); Olalde et al. (2018) were the first to focus on the second population turnover that befell the European continent; the advent of the steppe pastoralists. In the Middle Bronze Age samples from Martiniano et al. (2017) we see the first hints for the replacement of indigenous Y chromosome lineages by the R1b-M269 (Myres et al., 2011; Lucotte, 2015; Solé-Morata et al., 2017), whereas Olalde et al. (2018) argue for the male lineages and genomic shift at the larger scale in central Europe, in connection with the Bell Beaker phenomenon.

Now, after the publication of a massive temporal transect for Iberia there are over 400 ancient genomes from Spain and Portugal publicly available for all periods, from the Mesolithic to the Middle Ages (Olalde et al., 2019). However, the majority of the samples were produced by SNP capture microarray. In total, only six of the samples are Mesolithic and about 70 are Neolithic individuals. By far, the best covered period is the Chalcolithic with 150 samples. For the metal ages there are about 100 samples available spread across the Bronze Age to the Iron Age, with 60 and 40 individuals each, respectively. Among the 40 samples that are dated to the Iron Ages, fifteen are associated to non-local Hellenic culture. The Roman and Visigothic periods together only feature 20 ancient genomes. Finally, for the Islamic period 35 samples have been sequenced so far.

There are other individuals that died in distant lands identified as potentially Iberian in other papers. For instance, there are some probable Iberian individuals from the Roman Imperial era buried in Isola Sacra of Ostia, the port of Rome (Antonio et al., 2019), and some Medieval individuals from the Crusades period in Lebanon (Haber et al., 2019) are genetically consistent with being originally from Iberia. Obviously, this also has an echo in Spanish archaeology: some individuals excavated and sequenced in Spain do not look genetically indigenous. Findings like these have highlighted the existence of long-distance mobility, which could not be taken for granted before the advent of ancient DNA. This extreme mobility was possible even within the lifetime of a prehistoric individual of the Copper Age, as reflected in a Chalcolithic individual found in Camino de las Yeseras near Madrid (Olalde et al., 2019). This individual was evidently a North African who somehow voyaged to Iberia. Another example, although not actually discussed in the paper itself, are two of the Neolithic individuals from El Toro (southern Spain) analysed in Fregel et al. (2018). At least one of the two displays a very obvious North African genomic contribution which is not too surprising, since in the same paper it is showed that there was Iberian maternal genetic and genomic input into the Late Neolithic settlement of Kelif el Boroud, in Morocco. These interactions must have been very limited, though, because they do not persist or have a significant or lasting impact in the later Iberian population, at least any that has been detected so far.

Out of the total of 408 samples from all the works cited, a 67% (275) of them were sequenced

by target capture in Olalde et al. (2019) alone. Although this work expanded previous periods and painted a broad picture of the events that happened after the Bronze Age in Iberia at different key points in its history, the nature of the data produced also limits the types of analyses that can be done. This is because sequencing by capture only recovers information for a fraction (about 1 million variable positions) of the genome, in the best case scenario. Nevertheless, even with only this information a lot can, and has been, done. Note that, this work has produced the only genomic data currently available from individuals of the Iberian Iron Age, Roman Hispania, the Visigothic reign of Hispania, and the Islamic period of Al-Andalus, with the single exception of a Punic individual from Ibiza (Zalloua et al., 2018). The most recent samples date to around the 16th century in Granada, theoretically very shortly after the Christian conquest.These are the only ancient public genomes for these particular periods in Iberia, but although the sampling effort covers the majority of the Peninsula, it is very uneven and patchy across regions and periods. For example, all the Islamic samples are either from Andalusia or the east, and there are no contemporary samples from Christian territories Olalde et al. (2019).

As acknowledged by a recent review (Racimo, Sikora, Vander Linden, Schroeder and Lalueza-Fox, 2020), the fields of ancient DNA is shifting from large-scale studies that aim to cover continents or extensive regions, towards more local projects that aim to shed light on particular historical or more recent prehistorical events linked to various social and economical dynamics. A really good example of this, is the reconstruction of the genetic history of Rome from before its foundation until the Medieval age by Antonio et al. (2019). Another good example of small scale studies shedding light on intimate details of populations in the past, is the reconstruction of kinship relationships in a 5000 years old mass burial, associated to the Globular Amphora culture from Poland at the time between the end of the Neolithic and onset of the Bronze Age in Central Europe (Schroeder et al., 2019; Linderholm et al., 2020). This attention to local details is what the two biggest ancient DNA works on Iberia lack. That is the reason underlying the direction it was decided to follow in this project.

The focus was set on the eastern region of Iberia, currently referred to as the Valencian region. This region, like many other coastal landscapes in the Mediterranean world, has been a multicultural crossroads where swords have been crossed and goods traded. Not in vain, the destruction by Hannibal Barca of Saguntum, perhaps the most important city of the region in the 3rd Century BCE, ignited the Second Punic War. Historical records tell us that the aboriginal non-Indo-European speaking inhabitants interacted from very early on with Phoenician traders that established permanent settlements in Gadir, Malaca, Sexi, Akra Leuke and Ebussus. The Iron Age is a key moment, because popular imagination traditionally places there the formation of the Celtic, Iberian, Roman and Germanic identities. In Iberia, Celtic and Iberian speakers never identified as such and did not recognise any political or ethnic supra-organization beyond the tribe. What drove their linguistic differentiation is unknown, but Iberians culturally also show elements introduced from Indo-European traditions, especially in the funerary context and the importance of weapons.

What has been extracted from the first Iron Age genomes from Iberia is that the non-Indo-European speakers did not lack the genomic contribution introduced by steppe-related peoples (the Yamnaya are the alleged introducers of PIE (Bouckaert et al., 2012; Anthony, 2010; Haak et al., 2015; Anthony and Ringe, 2015)). Since the Bronze Age, all the genomes studied carry between 10-20% of steppe ancestry. What is more, since the beginning of the Iron Age, it seems that individuals display a small increase in the amount of this steppe ancestry, and the majority of these samples are labelled as Iberians, not Celts. Without more data it is impossible to say whether this is assimilation of distinct Bronze Age groups or an extra intake of migrants from Central Europe in the Iron Age. Archaeological evidence favours the second scenario. Archaeological evidence has attested the intrusion of the Urnfield Culture funerary tradition imported from continental Europe in north-east Iberia. Although pre-existing Late Bronze Age and Early Iron Age cultures also practised funerary cremations before.

In any case, it is early in the preceding Bronze Age when the wider European genetic pool consolidates. Despite that it is usually thought of as the moment when European root populations are forged, there is actually little genetic change in the Iron Age. We have to wait until the time of Imperial Rome to see more significant genetic transformations occurring in the Mediterranean populations, especially those under direct Roman rule (Olalde et al., 2019; Antonio et al., 2019). The genetic makeup of the peoples in the rich eastern provinces was exported across the sea, especially into cosmopolitan Italy but also into the northernmost remote fringes of the Empire such as Britain (Martiniano et al., 2016). The Near Eastern contribution was less in Iberia, but much like in Rome, for the first time there is evidence of foreigners with Greek related ancestry living and dying in the Greek colonies of Iberia such as Emporion (Olalde et al., 2019).

The data for the following Islamic period in Spain is mostly restricted to the South but it is sufficient to corroborate gene flow from North Africa and the introduction of exogenous uniparental markers typical of North Africa that were absent before Islamic times. Interestingly, however, the modern population from Andalusia harbors less North African ancestry than its Islamic medieval predecessor. This is a reflection of the fact that Bronze Age migrants from the steppe were not the architects of the last genetic mass transformation in Iberia (Haak et al., 2015; Martiniano et al., 2017; Olalde et al., 2018, 2019).

To address questions related to all these periods of human past in the eastern Mediterranean region of Spain, a sample set of over 250 samples was assembled with the collaboration of several museums and local authorities. The human samples collected for this project can be easily divided into two categories based on their geographical origin and time frame. The first group of the samples falls in the category of Genetic Prehistory, which includes samples from the Upper Paleolithic to the Iron Age, including the Late Mesolithic, all the Neolithic, the Copper Age, some Bronze Age individuals and the Iron Age. The second group of samples can be labeled as part of Genetic History, and includes several Roman sites, a couple of Visigothic sites, and almost a dozen of different Islamic, Christian Medieval and Post-Medieval archaeological sites.

Of the total sample size, I selected 55 samples from across all periods to build an initial scaffold for a future more complete temporal transect. Many answers can be provided to questions of interest to archaeologists questions using ancient DNA. However, retrieving ancient DNA from warm regions like the Mediterranean is not a certain outcome. The warm and humid climate typical of the Valencian region has a considerable negative impact in the survival of DNA in bones of ancient individuals. From previous works it has been estimated that the success rate is between 20% and 40%. The ultimate long-term goal will be to sequence almost everyone who ever lived and build a Genopedia for the Valencian region in order to have a detailed understanding of how the population dynamics have changed over the millennia, from caves to cities.

4.2 Materials

The required permissions to access the samples were obtained with the approval of the museums involved, the Servei de Cultura i Esport de Castelló, the Direcció General de Cultura i Patrimoni, the Conselleria de Educació, Investigació, Cultura i Esport de la Generalitat Valenciana, and the Servicio de Investigación Arqueológica Municipal del Ayuntamiento de Valencia (SIAM). Sampling was arranged to be carried out with the museum directors or archaeologists once permissions were granted. A more detailed explanation of the samples context is summarized in Table 5 at the end of this section,and the locations of the archaeological sites can be found in Figure 45.

The processing of an ancient sample itself can be arduous given all the necessary precautions and it requires two people working at the same time in the dedicated facility. Once the samples have been screened for the content of endogenous DNA only the best preserved samples were selected for further sequencing to maximise resources. In ancient samples a good outcome relies on the endogenous content of DNA.

Figure 45: Map of the Levantine coast in East Iberia, roughly corresponding to the provinces of Castellón and Valencia, where the bulk of the samples come from. Red dots indicate locations from where samples have been collected.

4.2.1 Prehistoric Samples

The prehistoric part focuses on the more mountainous province of Castellón (Valencian Community, Spain) which is where most caves and prehistoric sites in the region are found. The project will cover a time period spanning the Upper Paleolithic (12,200 BP) to the Iron Age (6th century BC) and try to construct a genetic scaffold for the area. The thorough sampling work has provided around 70 samples representing all the relevant periods of the prehistory of a territory three times bigger than the West Yorkshire county. Relevant ages include samples from the Paleolithic, Mesolithic, Neolithic, Chalcolithic or Bronze Age and Iron Age. Radiocarbon dates are only available for some of the samples but probably the only notable exception regarding the periods covered is the 3rd millennium BC.

Not much is known about the peoples that inhabited this region but the area has been continuously inhabited since the Neanderthal times. Based on the little genetic data from other ancient samples from the Spanish Mediterranean Basin (Cova Bonica, Barcelona) (Olalde et al., 2015) a degree of similarity with other parts of Iberia is expected since it is one of the first areas where Neolithic farmers must have settled in Spain. Regarding the Chalcolithic, there is not much archaeological evidence of the Bell Beaker phenomenon despite the abundance of human remains

in caves. Some samples from Camino del Molino (Murcia) have been studied (Olalde et al., 2018).With the Bronze Age (3rd millennium BC and early Iron Age) human remains become scarce, possibly due to changes in funerary rituals. During the Iron Age, the people of the region are known to have spoken and written a non-Indo-European language (Iberian) which has not been deciphered yet despite surviving until Roman Empire times. They also cremated their dead which makes it impossible to recover their ancient DNA. These combined factors have contributed to the mystery surrounding these Iron Age peoples.

I) Pre-Neolithic Archaeological sites: During the periods of the *Old Stone Age and Middle Stone Age* (Paleolithic and Mesolithic) human societies relied on hunter-gatherer strategies for subsistence.

- Cova Fosca (Upper Paleolithic): one almost complete skeleton buried in a cave with levantine style rock art, common in this inland mountainous region. The skull has been partially affected by a later fired in the cave where it was buried. It belonged to a short woman, about 145cm tall, dated to about 12,000BP. The site is located in Ares del Maestrat (Castellón).

- Cingle del Mas Nou (Mesolithic): open air ritualistic burial site with nine individuals in it. Three are adults (two males, one female), one is a 15yo kid. It dates from around 7000BP and is only half kilometer away from Cova Fosca. Only 2 adults and 1 teenager can be sampled. Site located in Ares del Maestrat (Castellón).

II) Neolithic Archaeological sites: This is the period of the *New Stone Age* and it is during this time that new human societies start to be characterized by the development and use of agriculture and pottery.

- Costamar (Early Neolithic): also known as Torre la Sal, based on the pottery is an Early Neolithic site. The site is located on the Mediterranean coast. There are 4 inhumations in round burials, and some are rich in grave goods and ornaments. All four individuals were sampled. Site located in Marina d'Or (Castellón).

- Cova Font de Codina (Mid-Late Neolithic): site located near Borriol (Castellón) which is in turn not too far from Costa Lloguera (Castellón) and not too far from the sea. It is dated to 3639-3384BC.

- Barranc de Beniteixir (Late Neolithic): together with Sanxo Llop and La Vital this forms a coastal mega site linked to the river Serpis. There is evidence of continuous inhabitation. Teeth of 3 individuals were collected. Site located in Piles (Valencia) dated to 2671-2599BC.

- Sanxo Llop (Late Neolithic): together with Beniteixir and La Vital forms a mega site near the coast. One individual was sampled from this site. Site located in Gandia (Valencia).

III) Chalcolithic and Bronze Age sites: This period is the beginning of the *Metal Ages* which is when human societies start to work first with copper and then bronze. Copper use appeared independently in at least, Iberia and the Steppe region. In Iberia the origin is tightly linked to tool making, but in the Steppe is also linked to ornament production. It is not until the introduction of the use of bronze in Iberia (via steppe-related migrants) that ornamentation starts to be made with these materials (Murillo-Barroso and Montero-Ruiz, 2012).

- La Vital (Chalcolithic Bell Beaker): site located by the mouth of the river Serpis. It was occupied during the Chalcolithic, Iron Age and the Middle Ages. It is a well studied Bell Beaker site with four burials with pottery and elements typical of the Bell Beaker package. A teeth from one Bell Beaker burial was sampled and dated to 2529BC. Site located in Gandia (Valencia).

- Sima del Pozo Cerdaña (Chalcolithic: the site is located in Pina de Montalgrao (Castellón). Five samples, some belong to children (Gabarda and Aguilella, 2016).

- Cova dels Diablets (Chalcolithic): site located in the Irta's mountain range near Alcala de Xivert (Castellón). One sample (petrous bone) (Aguilella Arzo et al., 1999).
- Abrigo I de las Peñas (Chalcolithic/Bronze Age): the site is found in the hinterland as part of a complex of small caves on the right bank of the river Palancia. Teeth from, at least, three individuals were taken. Site located in Navajas (Castellón).
- Cova Tossal de la Font (Chalcolithic/Bronze Age): site located inland in Vilafames (Castellón). One sample available (molar).
- Cova Costa Lloguera (Chalcolithic/Bronze Age): site near el Castellet in Castellón de la Plana (Castellón), now disappeared. Extraction of mtDNA HVS-I has been done in the past successfully (Oliver and Fernández, 2008). Two samples (molars).
- Covetes Barranc del Diable (Chalcolithic/Bronze Age): site near Oropesa del Mar in the coast and also near the Neolithic site of Costamar (Castellón). One sample (molar).
- Pla dels Avencs (Chalcolithic/Bronze Age): site located in Cabanes (Castellón) near the coast and the Neolithic site of Costamar. One sample (molar).
- Coveta Tossal de Ribalta (Chalcolithic/Bronze Age): Site near Castellón de la Plana. One sample available in the form of a phalanx.
- Cova dels Blaus (Chalcolithic/Bronze Age): Site located in Vall d'Uixo (Castellón). It is a sepulchral cave with some evidences of occupation since the upper paleolithic. There were 8 individuals buried in the cave from Chalcolithic times, then the cave was sealed with a wall of rocks. The cave is regarded as a dolmen-like structure. The bones present lesions compatible with having suffered from tuberculosis, which would be the oldest cases recorded in Iberia.
- Pla de Rius - Rambla Garrut (Chalcolithic): likely Eneolithic burial from Vall d'Uixo (Castellón). One sample available (petrous).
- Cova de la Iguala (Bronze Age): site in Alcudia de Veo (Castellón) in an inland area, classified as an Valencian early Broze Age. One sample (molar).
- Cova Mas d'Abad (Bronze Age): classified as early Bronze Age, it is an inland site located in Coves de Vinroma (Castellón). Two samples available (molars).

IV) Iron Age Archaeological sites: This period is characterized by the use of iron and marks the transition to *Classical Antiquity* dominated by societies with written history.

- La Escudilla and Los Cabañiles (Iron Age): sites from the VII and VI century BC, located in Zucaina (Castellón). It is a monumental complex with unique funerary traits and architecture. Activity here seems heavily linked to the funerary world and fluvial network. There is evidence of iron usage. Great numbers of newborns found buried in urn underneath the houses. There is also a tumular necropolis of arcs that stem from an original circular tomb for the adult individuals.
- Los Morrones (Iron Age): site located in Cortes de Arenoso (Castellón). It is a former fortified settlement on a small hill. There is evidence of occupation since the Bronze Age until Roman times, cereal farming, cattle and iron. It was abandoned abruptly or inhabitants were forced to leave. It had a tower, built in the 7th century BC.
- Coves Sant Josep (Iron Age): site located in Ares del Maestrat (Castellón). One infant buried, not cremated, in similar context to the others already described. Location near other sites sampled for other periods.

- Puig de la Misericordia (Iron Age): site located in Vinaros (Castellón). One infant buried, not cremated, in similar context to the others already described.
- Puig de la Nau (Iron Age): site located in Benicarlo (Castellón). One infant buried, not cremated, in similar context to the others already described.

4.2.2 Historical Samples

I) The project to reconstruct the urban genetic history of Valencia aims to cover a time frame spanning from the year 138BC to the 14th century AD. There are about 150 samples representing all the relevant periods of the city: Roman Republic, Roman Empire, Visigothic rule, Islamic rule and post-Islamic Medieval times. The samples theoretically include individuals that belonged to the ruling elite of the city for each period as well as common people.

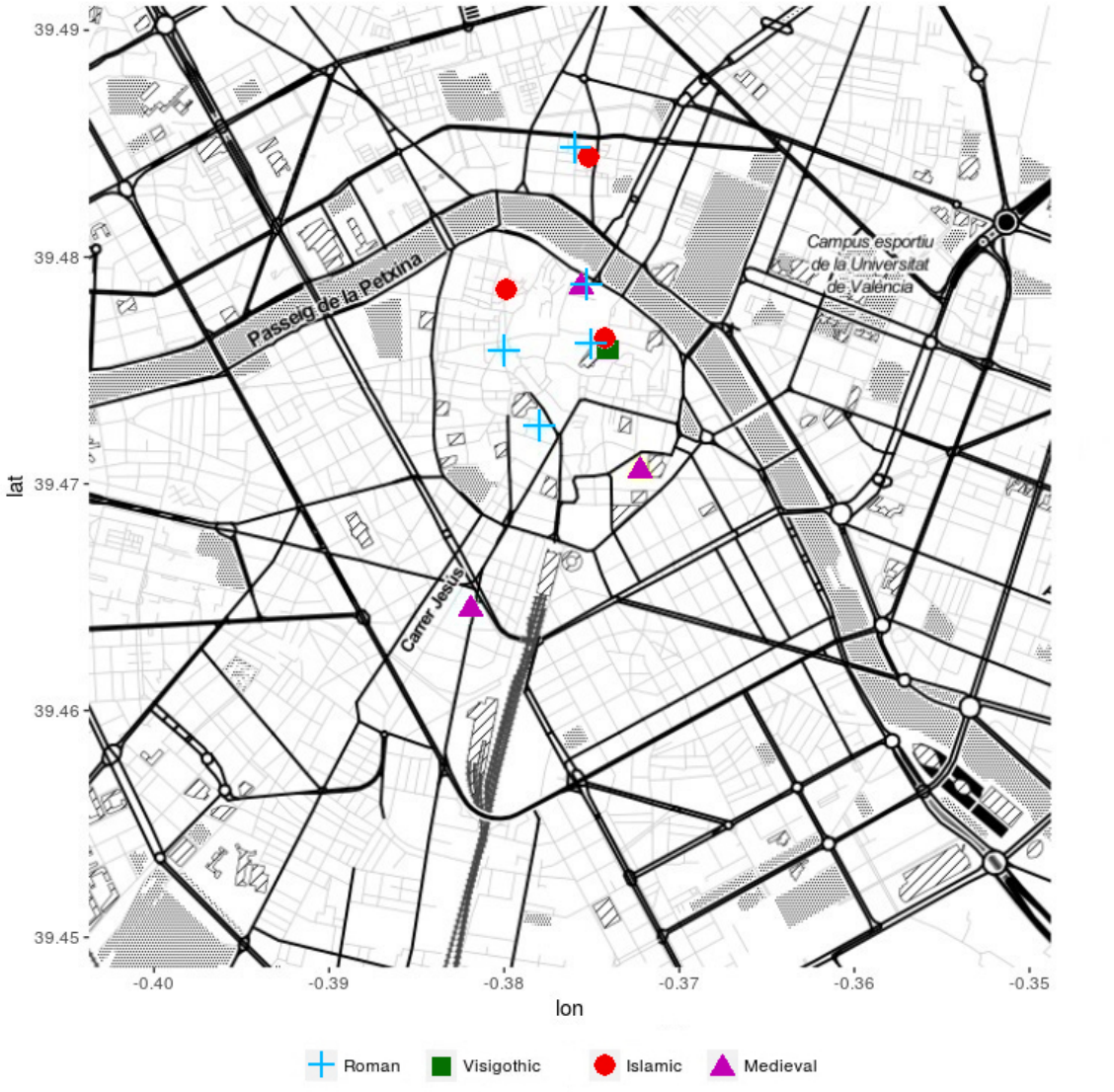


Figure 46: Sampling locations in Valencia city.

- Plaza de l'Almoina; the elites (Figure 46):

  - Roman Republic: six Sertorian legionnaires found in the area of the Forum of the Valentia of the late Roman Republican period. The bodies of the soldiers are thought to be victims of the destruction that happened 75 BCE in the context of the Sertorian War.
  - Roman Empire: Roman styled tombs with *tegulae* from the Late Empire.
  - Visigothic: tombs and family vaults with a style in clear break with the preceding Roman tradition, traditionally attributed to a new Germanic elite group but still associated to Christian cults.
  - Islamic Rauda: Necropolis associated with the royalty and other rulers of the city in the Islamic period, most likely from the post-Caliphate era. The tombs are simple, as the Quran indicates for the Islamic tradition, but traces of gold from the fabrics and some gravestones have been recovered.

- Necropolis Occidental de la C/Quart (Roman): this is commonly referred to by several names. This necropolis was the first one known to be used during the Republican period. There is heterogeneity in the style of the tombs, which indicates different social status and origin of the individuals. Seventeen individuals sampled (García-Prósper and Polo-Cerdá, 2016).

- Necropolis de la Boatella (Roman): this is the main necropolis of the city during the late Republican era discovered to date. It is also the most studied one. Six individuals from this site were sampled.

- Tumbas de La Saidia (Roman): the area was located north of the city. It is located on the north bank of the east-flowing river. The Via Augusta traversed the area (S-N). This area later became the arrabal de la Alcudia (Potter's Borough) during the Middle Ages. During the Roman Empire was an area favoured for burials, since many monumental and prestige tombs have been found in the area. Barely any other urban archaeological activity is recorded during this time. Five individuals were sampled, dating to the 1st-2nd century AD. The tombs were found in excavations at C/Pepita and C/Orihuela. One individual was found with a coin in the mouth as a payment for Charon.

- Monasterio de la Roqueta (Paleochristian): the site is linked to early Christians, and became a monastery that survives until nowadays. One individual buried in a lead coffin, identified as a paleochristian, was sampled.

- Necropolis del Arrabal de la Alcudia (Islamic): the name of the area derives from two Arabic words. The word *Kudya* means high ground or elevated land in a plain. *Arrabad* means a group of houses or suburb. The name can be loosely translated as high settlement or neighbourhood-on-the-hill. Such name makes sense given the location. The area is found extramural on the north bank on the river. Given that the river Turia is well known for causing devastating floods on a regular basis, the fact that the area is slightly more elevated than the surroundings could explain the evidence of settlement even before the foundation of the Roman colony. Interestingly an adjacent zone is called Marxalenes (a local name that recalls the former marshlands) and is known to have be regularly flooded until recent times, probably due to the influence of an old tributary stream. The Via Augusta crosses the heart of the neighbourhood as it exits the old walled town. During the Islamic period the area became a Quarter of Potters thanks to the abundant clay of the area. About fourty individuals have been sampled from this location and period (C/Sagunto49, C/Orihuela3-5 and C/Pepita27-29 excavations) (Machancoses López, 2015, 2016). During Roman times the location was not linked to urban activity and instead was sparsely used as sepulchral area.

- Necropolis de El Carmen or Bab al-Hanax (Islamic): this cemetery was located near the western gate of the city (*Bab* meaning door in Arabic), just outside the gate of the former Arabic wall, and hence the name. It is a big necropolis where thousands of burials within the Islamic rite have been found. It has been excavated in different campaigns and years, but probably not fully yet, since it is under urban land in the old town where buildings are not commonly demolished. Twenty individuals were sampled.

- Fosa de la Juheria (Late Medieval): seven individuals from the 14th century were recovered from a historically documented area dedicated for burials of the Jewish citizens. The site was known as Fossar dels Juheus and was adjacent to the eastern extramural side of the Christian wall. The Jewish Quarter was not far from the burial area but was instead located within the walled city. The individuals recovered were found in a common grave in a chaotic manner, indicating a need for a quick burial or a lack care for the bodies. These two scenarios are likely linked to two possibilities. One is that, the individuals were victims of an episode of the Black Death (*Yersina pestis*) in the city, recorded in historical accounts (1348 and 1401 AD). Another possibility is that the individuals were victims of a devastating pogrom, historically documented too, in the year 1391AD. This would explain why most of the skulls have deadly injuries caused by sword-like weapons.

- Cementerio de San Lorenzo (Late and Post-Medieval): Chrisitian cemetery inside the walled perimeter of Valencia. In use since at least the 15th century until the 19th century. In the 19th century there was a growing consensus that cemeteries inside cities were a health hazard and laws were promulgated to develop burial grounds outside urban areas. In 1841 following the confiscation of urban cemeteries, all the known and recent graves were moved to the new general cemetery of Valencia. Excavations in the 21st century found the older non-translocated medieval graves, spanning from the 14th century to the 17th, in the plots sold that became part of the urban network of the city after the land confiscation. The same area covered by the medieval cemetery was used in Late Roman times for burials too. I sampled 8 individuals from the 14-17th century period and two individuals from the Roman period.

II) Other Historical Samples: I also sampled individuals from locations of nearby neighbouring regions of Valencia to represent the rural world, including various periods:

- Sanxo Llop (Gandia): site in close proximity to La Vital, it is an area with evidence of inhabitation since prehistory. In recent year, an excavation revealed a burial ground dating to the Visigothic period. I sampled two individuals (adult and sub-adult) radiocarbon dated to the 6th-7th century CE. They were thrown into a round pit that had been used a dumpsite, and buried without regard for their final positions.

- Necropolis de La Union (Vall d'Uixo): another cemetery from the Visigothic period with over 40 inhumations. This site is peculiar in the sense that almost all the graves had several individuals buried one on top of the other. The archaeological interpretation is that the individuals in the same graves share some degree of kinship. They all appear to be victims derived from one violent event, and they must have died in a short period of time. Many individuals display signs of violence as revealed by the anthropological study. I sampled five individuals from this site.

- Islamic Maqbaras of Vall d'Uixo: these are a series of burial grounds discovered during excavations in the 1990s scattered across the town. They have been identified as the islamic cemeteries of the agricultural hamlets or farms (Alquería Benigafull, Alquería Benizahat, and Alquería Ceneja) that formed the primitive core of the Vall d'Uixo settlement. I sampled 10 individuals from four of these sites dated between the 11th and 14th century.

- Necropolis de la Plaza del Almudin (Segorbe): Islamic necropolis from the 11th-13th century period in the important walled medieval city of Segorbe. I samples five individuals from this site, including one individuals dubbed a giant by his remarkable height compared to the other forty individuals recovered. I also sampled various animals bones to help define a baseline for dietary isotope analysis (Barrachina Ibáñez, 2005).

Site	Samples	Location	Era	Site type & Age
Cova Fosca	1	Cortes de Arenoso (Castellón)	Upper Paleolithic	Cave (12000 years BP)
Cingle del Mas Nou	3	Cortes de Arenoso (Castellón)	Mesolithic	Inhumation (6000-5300 BC)
Costamar	4	Marina d'Or (Castellón)	Early Neolithic	Inhumation
Camino de Missena	1	Pobla del Duc (Castellón)	Early Neolithic	Inhumation (4797-4583 BC)
Barranc de Beniteixir	2	Piles (Valencia)	Late Neolithic	Inhumation (2671-2599 BC)
Sanxo Llop	4	Gandia (Valencia)	Late Neolithic	Inhumation
Arroyo Saladillo	2	Malaga (Andalucia)	Late Neolithic	Inhumation
Cova Font de Codina	1	Borriol (Castellón)	Mid-Late Neolithic	Cave (3639-3384 BC)
Abrigo I de las Peñas	3	Navajas (Castellón)	Chalcolithic	Cave
Cova dels Blaus	7	Vall d'Uixo (Castellón)	Chalcolithic	Cave
La Vital	1	Gandia (Valencia)	Chalcolithic	Inhumation (2529 BC)
Arenal de la Costa	1	Ontinyent (Castellón)	Chalcolithic	Inhumation
Pla de Rius - Rambla Garrut	1	Vall d'Uixo (Castellón)	Chalcolithic	Cave?
Sima del Pozo Cerdaña	5	Pina de Motalgrao (Castellón)	Chalcolithic	Cave (2615-2470 BC)
Costa Lloguera	2	Castellet (Castellón)	Chalcolithic/Bronze Age	Cave (1876-1643 BC)
Cova Barranc del Diable	2	Oropesa (Castellón)	Chalcolithic/Bronze Age	Cave (2894-2678 BC)
Cova del Diablets	1	Alcala de Xivert (Castellón)	Chalcolithic/Bronze Age	Cave (2889-2630 BC)
Pla dels Avencs	1	Cabanes (Castellón)	Chalcolithic/Bronze Age	Cave
Tossal de Ribalta	1	Castellón (Castellón)	Chalcolithic/Bronze Age	Cave
Tossal del la Font	1	Vilafames (Castellón)	Chalcolithic/Bronze Age	Cave (3341-3030 BC)
Cova del Mas d'Abad	2	Coves de Vinroma (Castellón)	Bronze Age	Cave (1746-1566 BC)
Cova l'Iguala	1	Alcudia de Veo (Castellón)	Bronze Age	Cave (2028-1884 BC)
Coves Sant Josep	1	Vall d'Uixo (Castellón)	Iron Age	Inhumation jar
La Escudilla	18	Zucaina (Castellón)	Iron Age	Inhumation jar
Los Cabañiles	8	Zucaina (Castellón)	Iron Age	Inhumation jar
Los Morrones	3	Cortes de Arenoso (Castellón)	Iron Age	*In situ*
Puig de la Misericordia	1	Vinaros (Castellón)	Iron Age	Inhumation jar
Puig de la Nau	1	Benicarlo (Castellón)	Iron Age	Inhumation jar
L'Almoina - Forum	6	Valencia (Valencia)	Roman Republic	*In situ* (75 BC)
C/Quart - Occidental	17	Valencia (Valencia)	Roman Republic	Necropolis
C/Orihuela - Norte	2	Valencia (Valencia)	Roman Empire	Tombs
C/Pepita - Norte	3	Valencia (Valencia)	Roman Empire	Tombs
L'Almoina - Necropolis	4	Valencia (Valencia)	Roman Empire	Necropolis
La Boatella	6	Valencia (Valencia)	Roman Empire	Necropolis
Monasteri S. V. La Roqueta	1	Valencia (Valencia)	Paleochristian	Tomb
Cementerio San Lorenzo	2	Valencia (Valencia)	Late Roman	Tomb
L'Almoina - Necropolis	21	Valencia (Valencia)	Visigothic	Necropolis
Necropolis La Union	5	Vall d'Uixo (Castellón)	Visigothic	Necropolis
C/Orihuela - Maqbara	30	Valencia (Valencia)	Islamic	Maqbara Urban
C/Sagunto - Maqbara	10	Valencia (Valencia)	Islamic	Maqbara Urban
El Carmen - Maqbara	21	Valencia (Valencia)	Islamic	Maqbara Urban
L'Almoina - Rawda	33	Valencia (Valencia)	Islamic	Maqbara Elite
Maqbaras de Alquerias	9	Vall d'Uixo (Castellón)	Islamic	Inhumations Rural
Plaza del Almudin	3	Segorbe (Castellón)	Islamic	Maqbara Urban/Rural
Fosa de la Juheria	7	Valencia (Valencia)	Late Medieval	Mass grave
Monasteri S. V. La Roqueta	2	Valencia (Valencia)	Medieval	Christian Urban
Cementerio San Lorenzo	10	Valencia (Valencia)	Late Medieval	Christian Urban

Table 5: Summary of the samples collected, listed in chronological order

4.3 Methods

All the following steps are included in the ANIBAL v2.6, which is a workflow I compiled for the Archaeogenetics Research Group at the University of Huddersfield, to perform standardized analyses to deal with ancient genomic data. The workflow consists of mostly public resources and software in addition to some in-house scripts that I developed.

Prior to sequencing with Illumina HiSeq4000 sequencers at Macrogen (South Korea), the samples were screened to estimate the endogenous human ancient DNA content. The screening was made with a MiSeq sequencer by staff (Valeria Mattiangeli) at Dan Bradley's Lab at Trinity College Dublin.

4.3.1 Processing of Ancient Bones

The Archaeogenetics Research Group at the University of Huddersfield runs a dedicated Ancient DNA Facility with four clean rooms to process archaeological samples. This laboratory is isolated from other molecular biology facilities since it is located in the Technology Building of the Queensgate Campus (Huddersfield). In every work session, I was outfitted with a full-body Tyvek suit, a hairnet and face mask and a double layer of gloves at all stages. The first room is used for UV radiation of bones and to put on the full-body suit. The second room had diverse post-UV exposure material and tools for cleaning tasks and re-filling consumables such as gloves and liquids. The other two rooms are only connected by a transition corridor and are dedicated to sample drilling and DNA extraction respectively. The protocol currently established that I followed required all materials to be regularly cleaned with LookOut® DNA Erase (Sigma-Aldrich), bleach and by exposure to UV-radiation after each session.

To reduce risk of bacterial and other contamination, the samples I selected to be processed by drilling had to be UV-radiated for a total of 60 minutes (30 minutes for each side) in the first room and then covered in tin foil with labels. This usually took a full morning if there were many samples, limited by the size of the UV machine. After the UV shower, I moved the samples to the drilling room. First step before drilling was to clean the surfaces of the petrous bones or molars by air-abrasion, with 29 μm aluminium oxide powder (OEA Labs) and a SWAM-Blaster® compressed air abrasive system (Crystal Mark).

I used a Micromotor System Maxima drill with a 22mm diameter diamond cutting edge. In the case of a tooth I separated root and crown to powder the roots for DNA extraction. In the case of petrous bone, I cut a wedge from the densest part (Pinhasi et al., 2015). In the case of other bones like phalanxes I kept the epiphysis. Once a piece was obtained, I introduced it into a dedicated machine where the sample is powdered and then stored in a tube and weighted, ready for the next step of DNA extraction in the last room.

Isotope analysis for further dietary analysis was conducted externally: carbon and nitrogen at the Research Laboratory for Archaeology (University of Oxford); and oxygen at BioArCh (University of York). I will spare the details of this step since it falls outside my area of technical expertise, but the details will be available in future publications when the data is released if the reader is interested.

4.3.2 Extraction of Ancient DNA

I followed a modified protocol extraction of DNA from ancient remains (Rohland and Hofreiter, 2007) following Yang et al. (1998) and MacHugh et al. (2000), and used Fisher reagents. In our facility, the extraction process lasts three days. The first step is preparing the extraction buffer (EB) that contains 20mM of Tris HCL, pH 8; 50mM of EDTA, pH 8, RNase and Proteinase free, and 0.5% of SDS (DNase, RNase and protease free, heated to a temperature of 37ºC). All components are clean and exposed to UV-light for 15 minutes before addition of 200μg of proteinase K.

Once ready, 1 mL of extraction buffer was added to the tube with the powder. The tubes with the mix of buffer and powder were incubated and rotated for approximately 24 hours at 37 ºC.

Following incubation, the tubes were centrifuged at 13,000rpm for 15 minutes to separate mineral particles from the supernatant. The supernatant was kept in a fridge as backup. To the remaining pellet I added 1 mL of new extraction buffer and vortexed the tubes to resuspend it. Then, the tubes have to go back to the rotating incubation for another 24 hours at 37ºC.

After the second incubation the tubes were centrifuged at 13,000rpm again for another 15 minutes. Then, the supernatant was transferred to 6mL Corning® Spin-X® UF Concentrator tubes. To these new tubes I added a quantity of 3mL of 10mM Tris HCL (pH 8) and then centrifuged them for 20 minutes at 2,500 rpm. The flow-through was discarded. The previous step was repeated, and another volume of 3mL of 10mM Tris HCL was added to the concentrator tubes followed by another centrifugation at 2,500rpm of 20-30 minutes.

The final volume to be kept is about 100μL and is retained above the filter of the column tube. This volume was transferred for purification to new silica columns (MinElute® PCR Purification Kit, commercialised by Qiagen) and purified following standard protocol by the manufacturer, plus addition of 0.05% Tween20 (0.03μL per sample) to 59.97μL per sample of EB Buffer to reduce absorption of DNA to plastics and keep viability of extraction in the long run.

The 100μL volume of DNA extracted was stored in 2 ML O-ring tubes (Molecular Bio Products) in a fridge at 4ºC. Finally, I took a small volume of each DNA extraction that I later quantified with a QubitTM 3.0 Fluorometer (ThermoFisher Scientific), using the Qubit® dsDNA HS Assay Kit (Invitrogen).

4.3.3 Library Preparation and Sequencing

The protocol from Meyer et al. (2012), with modifications introduced in Gamba et al. (2016); Cassidy et al. (2016), was followed to make the necessary library preparations for my samples. I ignored the step of DNA fragmentation because ancient DNA is already highly fragmented. In between all main stages (I, II, III, IV and V) of library preparation I performed clean-up steps using the MinElute PCR Purification Kit, according to manufacturer instructions, and adding Tween 0.05% to EB Buffer to obtain EBT Buffer again in the same way as in the extraction stage.

Samples that passed the minimum endogenous DNA threshold for further sequencing were UDG-treated and had an initial extra step as follows: addition of 5.0 μL of USER® enzyme (Uracil-Specific Excision Reagent by New England BioLabs®) to 16.5μL of DNA extract and incubated for 3 hours at 37ºC. USER is a mixture of uracil DNA glycosylase (UDG) and endonuclease VIII. UDG removes uracil residues derived from post-mortem damage characteristic of ancient DNA (Briggs et al., 2007; Lindahl, 1993, 1996), generating apyrimidinic sites, whereas the endonuclease VIII cleaves the molecule on those sites, by breaking the phosphodiester backbone at the 3' and 5' sides of the apyrimidinic site. This process breaks DNA into even smaller fragments that remain suitable for sequencing but also reducing damage patterns detected during sequencing (Briggs et al., 2009).

The library preparation protocol consisted of various stages which include blunt-end repair (I), adapter ligation (II), followed by an adapter fill-in reaction with Bst DNA polymerase (III). Indexing oligo sequences are added by amplification with IS4 primer (IV). Finally libraries to be sequenced together in the same lane are pooled together (V).

I) The blunt-end repair step is where overhanging ends are removed by T4 DNA polymerase. I merged the cleaned aDNA volume of 21.5μL from the USER-treatment, together with 3.5μL of NEBNext End Prep Enzyme Mix, 7μL of 10X of NEBNext End Repair Reaction Buffer (both included in the NEBNext® End Repair Module, New England BioLabs®), and 38μL of ddH2O (sterile ultra-pure water). The volumes reported here are for one sample; for bigger batches we multiplied by the number of samples to be done plus a margin of pipetting error. The final volume (for one sample) is 70.0μL that was later incubated at 25ºC for 15 minutes, followed by 5 minutes at 12ºC, and purified it with MinElute PCR Purification Kit.

II) The adapter mixes of P5 and P7 (20μM each) (by Sigma-Aldrich) were pooled together with 1μL of T4 DNA ligase I (5U/μL), 10μL of ddH2O (sterile ultra-pure water), 4μL of 10X

T4 DNA ligase buffer by Thermo Scientific (with prior warm up and vortexing to dissolve if precipitate is present), and 4μL 50% PEG-4000 (Thermo Scientific). The final volume is 20μL per sample and was pooled together with another 20 μL of eluate from the previous Step I, then it was incubated at 22ºC for 30 minutes. In this step, adapters were ligated by the activity of T4 DNA ligase catalysis of phosphodiester bonds between 5' and 3'-ends in dsDNA. PEG-4000 is necessary to ensure a successful blunt-end ligation of the adapter.

III) For the adapter fill-in step, I merged the 20μL of DNA from Step II with 13.5μL of ddH2O (sterile ultra-pure water), 4μL ThermoPol® Reaction Buffer 10X, 1μL of dNTP (10mM each), and 1.5μL of *Bst* DNA polymerase (Large Fragment, 8U/μL). This makes a total volume of 40 μL to be incubated for 30 minutes at 37ºC, followed by an extra 20 minutes at 80 ºC necessary to inactivate the *Bst* DNA polymerase and terminate the reaction. The rationale behind this step is that because P5 and P7 adapters do not have 5'-phosphates, single-end overhangs that appear have to be filled by the reaction with *Bst* DNA polymerase.

IV) The final amplification step is partly carried out outside the Ancient DNA facility. Library amplification reactions are prepared in the Ancient DNA lab, but then, the tubes are sealed and taken to our Modern DNA lab, where I set up out the PCR amplification reactions. The reaction ingredients need non-repeated barcoding indexes to be added to allow multiplexing. The reaction consists of 41 μL of Accuprime *Pfx* SuperMix (Thermo Scientific), 1μL of primer IS4 (10μM), 2μL of appropiate indexing oligo (both made by Sigma-Aldrich) plus 6µL of sample library from Step III. Total reaction volume is 50μL. The PCR reaction protocol consisted of an initial denaturation phase at 95ºC for 5 minutes, followed by 12 cycles of denaturation at 95ºC for 15 seconds, annealing at 60ºC for 30 seconds, extension at 68ºC for 30 seconds, and a final extension at 68ºC for 5 minutes.Finally, I performed another a final clean-up step with the MinElute PCR Purification Kit.

V) I measured the concentration of the libraries with a QubitTM 3.0 Fluorometer, using the Qubit® dsDNA HS Assay Kit. Following this, I checked the fragment size distribution with a Bioanalyzer (Agilent), using the Agilent High Sensitivity DNA Kit. Libraries are evaluated visually (generally there was one library per sample but with three PCR replicates each) and if deemed successful (majority of fragment lengths in the libraries with a peak at 150-200bp) they were finally pooled together for each lane, and sent to Macrogen (Seoul, South Korea) for whole genome next-generation sequencing (NGS) in Illumina HiSeq4000 lanes.

4.3.4 Processing Next-Generation Sequencing Data

First of all, the paired-end raw FASTQ files were evaluated using FastQC (Version 0.11.7 by Babraham Bioin- formatics group) to check for quality, over-represented sequences, etc... in the high throughput sequenced data. Results of the checks were reviewed visually with the HTMLs created.

Once it is clear the FASTQ files are good enough, it is necessary to proceed to remove adapters sequences. The sequence for the forward adapter was AGATCGGAAGAGCACACGTCTGAACT CCAGTCAC, the sequence for the reverse adapter is AGATCGGAAGAGCGTCGTGTAGGGAA AGAGTGT. For this step, either Cutadapt or AdapterRemoval can be used (base quality and a minimum read length cut-off at this stage is optional, but these filters are applied later). The primer sequences have to be provided in order for the software to perform the removal. However, because of the Burrows Wheeler Aligner sampe tool to merge paired-end FASTQ files did not work well and introduced noise in downstream analyses, Cutadapt and AdapterRemoval were replaced by LeeHom. LeeHom allowed us to merge paired-end FASTQ files without problem and to remove adapters in one step by making use of its algorithm. Also, in LeeHom, I activated the flag (–ancientdna) to specify the software is dealing with ancient DNA.

Example usage with the sequences of the forward and reverse adapters I used: leeHom -fq1 InputFile1 -fq2 InputFile2 -fqo ID –ancientdna -f ForwardAdapter -s ReverseAdapter –log ID_leeHom.log

Mapping NGS Data

A reference genome in Fasta format is required, in this workflow; I used the Human Reference Build 37 (hg19/GRCh37.p13). Before proceeding with the alignment of the FASTQ files, the reference fasta file was indexed using the Burrows Wheeler Aligner utility (bwa index reference.fa). A dictionary file also had to be created using Samtools (*samtools faidx reference.fa*). These two steps help the aligner algorithm to be more efficient and save time and memory.

Burrows Wheeler Aligner (BWA) is the software of choice for mapping the FASTQ file sequences against the reference genome (hg19). More specifically, the algorithm BWA-backtrack was chosen. The command "samse" and "aln" were used to produce SAM files. Specifications (-l, -n 0.01, -o 2) were used to disable the minimum seed length (-l option) and allow the aligning algorithm to be more flexible and map more reads increasing coverage. The base quality filtering option (-q option) is still optional at this point. For further processing steps raw BAM files are necessary, this is achieved using SAMtools to "sort" the SAMs and create BAMs.

Example of use for single end reads: *bwa aln -l (active) -n 0.01 -o 2 (-q 20) -t 6 reference.fasta file.fq > file.sai* and *bwa samse reference.fasta file.sai file.fq | samtools sort -o file.sorted.BWA.bam -O bam -T TemporalBWA.deleteme*

After the mapping process is complete, I used Qualimap to check the mean overall coverage, coverage per chromosome and mitochondrial DNA. In this manner, we can and evaluate statistics of the quality of the sample alignment such as: total number of reads, mapped reads, lengths of the reads, duplication rate, GC percentage, mean mapping quality, general error rate and other mismatches. With the number of mapped and total reads we can re-estimate the endogenous DNA of the sample.

Example usage: *qualimap bamqc -bam file.bam -outdir qualimap_results*

Trimming NGS Data

Removing duplicates is the first step after mapping the FASTQ files. The BAM file can be purged of duplicates using either the "rmdup" option from Samtools or with the "Dedup" option with Picard using the "MarkDuplicates" command. To keep consistency, the Samtools option was chosen (samtools rmdup -s file.bam file_rmdup.bam).

Once the duplicates were removed, the quality filters were applied. A minimum mapping quality of 20 was chosen because introduces less reference bias than the also commonly used threshold of 30. Minimum read lengths were set at 34 base pairs long too. These filters are the same as the ones used in other publication with ancient genomes sequenced using the shotgun technique, for example from the Trinity College Bradley group (Cassidy et al., 2016; Martiniano et al., 2017). This level of stringency is needed in order to avoid making false SNPs calls from reads mapped with low quality.

The last filters I applied were in order to avoid bias introduced by post-mortem deamination patterns (PMD) that ancient DNA typically displays at the ends of the reads. PMD patterns were evaluated visually by looking at the plots generated with mapDamage (v.2.0.7). I judged that trimming three bases pairs at the ends of each read was sufficient to remove the vast majority of the deamination damage, given that the sequences are already uracil–DNA–glycosylase (UDG) treated (Rohland et al., 2015). Three base pairs were soft-clipped at the ends of the reads by reducing their quality below the interval of confidence to avoid false SNP calling. This was performed using the *trimBam* (–clip) function in bamUtil.

A necessary step, because it was projected that downstream analyses with GATK were going to be performed, was to add read groups using Picard Tools to each BAM file before merging (*java -jar picard.jar AddOrReplaceReadGroups I=file.bam O=file_RG.bam RGID=Name RGLB=Library RGPL=Illumina RGSM=Sample RGPU=Unit*).

Finally, if more than one library existed for each sample, then all the resulting BAMs that had been treated independently until now, were merged in one file using the merging option in Picard.

In this instance, there were at least three libraries or BAM files per sample that had to be merged with different read groups to distinguish the three libraries within the merged final BAM.

SNP Calling from BAM Files

Due to the incomplete nature of ancient DNA and general low coverage of the data (<20X) there is always uncertainty as to whether an observed allele is in homozygous or heterozygous state, for the reference or alternative. This is the reason why the SNP calling in ancient DNA is made using a pseudo-haploid approach. This approach creates artificial homozygous-only genotypes by taking one allele at random at each position, and this helps reduce bias in SNP calling.

To call the variants we used two lists of SNPs. One is the 600k list used in the Human Origins dataset (Lazaridis et al., 2014). The other list was created *ad hoc* by myself by extracting 1 million mutations (transversions only) from the VCF files from the Simons Genome Diversity Project (SGDP).

Initially, SNPs were called on all ancient samples (I processed public ancient DNA data in the same way as the data newly generated here) using a combination of SAMtools *mpileup* and SequenceTools *pileupCaller* (with quality filters q20, Q20 enabled) but I later discovered that pileupCaller generated artificial affinities with other *pileupCaller* samples regardless the time period and geographic origin.

I moved on to a script that makes use of GATK Pileup instead of SAMtools *mpileup* to call SNPs and then takes one allele at random to create pseudo-diploid genotypes. This approach increases drift artificially but not in any particular direction, so it does not have an effect in the analysis about relationships between samples. The GATK Pileup tool in GATK making the SNP calls was set to require a minimum base and mapping quality of 20 in all reads mapped. All Heterozygous positions were later removed. Finally, the remaining SNPs were pruned for positions in linkage disequilibrium using PLINK as described in Cassidy et al. (2016). The final genotype calls for each ancient samples were finally merged with the public dataset also using PLINK.

For particular genotypes of SNPs of interest I also used GATK HaplotypeCaller (-T option) to produce VCF and gVCF files (e.g. mitochondrial or Y-chromosome haplotype and phenotypic traits). VCF is an output that will only include the variants positions of a given list or range specified. A gVCF file includes all positions of a given list, which is useful in some cases when the coverage is low.

I used the a whole genome data analysis toolset provided PLINK1.9 (Purcell et al., 2007) along with utilities like *convertf* and *mergeit* included in EIGENSOFT toolkit from the Harvard Reich Lab. The tools provided by these software programs were used to filter for things like minimum allele frequency and linkage disequilibrium, to convert between formats, and to merge files whenever necessary with the Human Origins dataset of modern populations (600k SNPs) (Lazaridis 2016) and the North African dataset from Barcelona Comas Group.

Classification of Uniparental Markers

I obtained the mitochondrial mutations using GATK v.3.7-0-gcfedb67 HaplotypeCaller (McKenna et al., 2010). I also checked the haplotypes manually on IGV v.2.3 (Thorvaldsdóttir et al., 2013) to confirm that no heteroplasmies were present, this is important since heteroplasmies can indicate contamination. All haplotypes were clean. Haplogroup classification was made with the online using HaploGrep 2.0 tool (Weissensteiner et al., 2016) which follows to the nomenclature in PhyloTree (Build 17, February 2016) (van Oven, 2015).

Initially, I performed Y-chromosome haplogroup classification manually by cross-checking VCF files of the Y-chromosome with the 2019 ISOGG list mutations (International Society of Genetic Genealogy). However, the final results were obtained using Yleaf (Ralf et al., 2018) which basically does the same but optimized. For samples whose Yleaf classification was doubtful I double checked them with pathPhynder, a tool developed by Rui Martiniano (ruidlpm at github)

that integrates ancient and modern variation. This approach does not rely on known variants only, but also takes into account all informative Y-chromosome markers in a given high-coverage dataset and in the ancient samples. The combined evaluation of derived and ancestral positions yields better resolved clade classification than Yleaf.

Data Authenticity and Kinship

Anti-contamination measures were in place while drilling, extracting aDNA and during library preparation in the Ancient DNA Facility as explained above. However, negative controls were also introduced in the chain leading to library preparation and sequencing with Illumina MiSeq (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland) in the form of air (empty Eppendorf opened for some time before sample drilling began) and water control (2ml of ddH2O shaken in the sample mixer used to powder bone, after bleaching and UV exposure). The DNA quantification and sequencing results of these controls in the first batch of samples screened showed that the levels of contamination by exogenous DNA were negligible to zero.

I further checked authenticity of the data by checking the patterns of post-mortem damage and DNA fragmentation in the GOG samples with MapDamage v.2.0.7 (Jónsson et al., 2013) and BamDamage (Malaspinas et al., 2014). An example of MapDamage for one of the libraries sequence is available in the Supplement II (Figure S29). All samples presented the typical misincorporation patterns of ancient DNA but in very low levels since they are USER treated. The damage pattern were almost identical to the example presented in the SM.

Another extra measure for authentication was to confirm that the mitochondrial haplotypes of each individual were consistent with one haplogroup only and no heteroplasmic positions existed. I also checked that no unexpected mutations in the mitochondrial haplotypes were potentially derived from my own mtDNA (H1e1a). For this purpose, my mitochondrial genome was sequenced and classified the year before starting the processing of bones in the Ancient DNA Facility. No heteroplasmies were found in the samples, and furthermore no sample belonged to haplogroup H1e1a.

I also used a genetic sex determination script by Skoglund et al. (2013), which is sometimes inconclusive when the sample is below a threshold of extreme low coverage or contaminated. This was not the case for any of the new GOG samples. The Ry values used to classify a sample as carrier of XX or XY were clearly differentiated into two groups, males and females. This serves as initial evidence that the DNA in each final merged BAM comes from one single individual, or at least from individuals of the same sex. Successful determination of genetic sex is already indicative of good quality data but also served to further validate authenticity of the ancient genomes.

To establish kinship relationships between samples I used READ (Relationship Estimation from Ancient DNA) (Kuhn et al., 2018). READ is a software designed to manage low-coverage pseudo-diploid data in EIGENSTRAT format. It tests systematically pairs of individuals in a given dataset and is able to classify the type of relatedness degree as non-related, second degree (e.g. grandparent-grandchild, half-siblings, uncle/aunt, nephew/niece), first degree (parent-offspring, siblings), and identical (twins or duplicated individual). All combinations of GOG samples were attempted regardless of location and time period to confirm no cross-contamination happened between samples or myself which would show up as unexpected artificial relatedness.

EIGENSOFT, AdmixTools and DATES

For Principal Component Analysis I used the EIGENSOFT tool *smartpca* (Patterson et al., 2006). The PCA dimensional reduction was carried out using a subset of modern populations from the 600k SNPs Human Origins dataset (Hellenthal et al., 2014; Lazaridis et al., 2014). The subset of populations included all individuals available from Europe, the Caucasus, Iran, the Near East, Arabia and North Africa. The genetic variation of 434 ancient genomes from Spain, Anatolia,

the Fertile Crescent and Morocco was later projected onto the PCA built with modern populations (option lsqproject: YES). Outliers were not removed from the calculations.

Formal tests to evaluate treeness and admixture were performed with D-statistics (Green et al., 2010), *f3*-statistics and *f4*-statistics (Reich et al., 2009; Peter, 2016) included in the AdmixTools package by the Harvard Reich Lab (Patterson et al., 2012). Default parameters were used in all *f3, f4* and *D* statistics. The *f3* tests were used both in the form of *outgroup-f3* and *admixture-f3* depending on the scenario to be investigated. For the case of *f4-ratio* test, *f4* values were outputted instead of *D* values.

I also used the software DATES (Distribution of Ancestry Tracts of Evolutionary Signals) as described in the supplementary material of Narasimhan et al. (2019). DATES is a method to estimate the time of admixture in ancient genomes using genotype data and linkage disequilibrium maps.

All the above analysis were performed out on the same dataset used in PCA and ADMIXTURE analysis.

ADMIXTURE

ADMIXTURE (Alexander et al., 2009) is a well-known software with a maximum likelihood algorithm that is able to infer proportions of a number of *K* ancestry components of individuals from a number of *N* genetic markers, typically SNP genotype datasets. I used the model-based clustering approach of ADMIXTURE to estimate ancestry components in the newly generated samples, together with the same modern and ancient populations from the PCA. Only ancient samples with at least 10,000 SNPs and with a minimum quality score threshold (mapping and base quality 20) covered were included. I also applied a filter for linkage disequilibrium in PLINK1.9. (text–indep-pairwise together with parameters 200, 25 and 0.4). This filter decreased the final number of SNPs used in ADMIXTURE to about 200k SNPs. I ran ADMIXTURE in supervised and unsupervised mode with the same dataset built with different SNP lists (HO transitions+transversion and SGDP transversions only) from K=2 to K=10 with the cross-validation flag activated (–cv). The results were very similar in all cases and datasets and modes. The lowest median CV error before a steep rise was for K=4; this result is different from the general value typically between K=9 and K=11 from other publications because ADMIXTURE was not run together the all the worldwide populations from the Human Origins dataset.

4.3.5 Datasets

In order to analyse our whole genome data in combination with the ancient genotypes and publicly available modern genotypes, I downloaded dataset of genomes available from the Reich lab https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dnadata-compiled-published-papers (v37, which includes the Human Origins dataset merged with several other ancient populations). I merged my ancient data with the dataset using the mergeit program available from the EIGENSOFT package v7.2.116. I filtered out SNPs in LD and triallelilic positions and I removed non west Eurasian populations from the file. The remaining final number of SNPs was 500k.

Additionally, I also extracted from the 1000 Genomes Project dataset all the positions denoting transversions and created a new SNP list, and after filtering for LD and triallelic sites, I used this dataset that included ancient and some modern populations to repeat all the analysis made with the Human Origins dataset. I also used the 1000 Genomes Project dataset to calculate Runs of Homozigosity (RoH), using transitions and transversions, in modern population to compare with the RoH patterns in my four imputed ancient genomes.

I also used the North African Affymetrix dataset by Arauna et al. (2017) to project and see the Islamic and Medieval ancient samples in a non-European context (S13). I further merged the Arauna dataset with the Human Origins dataset to complement the low number of native North

African populations in the Human Origin dataset. The resulting overlap of SNPs between the two dataset was only about 20k, not enough for f-statistic analysis but enough to at least perform a PCA.

4.3.6 Imputation of Ancient Genomes

For genotype imputation, I selected all GOG samples above 0.5X mean coverage. Due to limitations in computing resources and time at the University of Huddersfield, I only imputed Chromosome 22 from the samples above 0.5X, since it had to be done locally.

I called all variants present in the chromosome 22 panel from the 1000 Genomes Project dataset using GATK UnifiedGenotyper (McKenna et al., 2010). The BEAGLE 4.0 (Browning and Browning, 2007) was used to to phase and impute the missing genotypes making use of the the 1000Genomes phase 3 reference dataset (Auton et al., 2015). The final output was in the form of VCF files with genotypes for the 1,103,557 positions in chromosome 22 alone.

The information contained in the imputed VCF files was used to perform an analysis of runs of homozygosity (RoH). This analysis measures the length of stretches of DNA fragments that are homozygous. PLINK can be used for this purpose, but I did it with an in-house script that measures the distance between base positions in the VCF that have heterozygous genotypes. I did so both for SNPs that are transitions and transversions and also considering only transversion mutations.

4.3.7 Simulation of Hybrid Genomes

I developed a Python script that acts as a mating bot to emulate sexual reproduction when provided with two genotypes in EIGENSTART format. This was in order to generate artificial populations to be visualized in the PCA. The script simulates a hybrid genotype as if it were a first generation off-spring intermediate of the two different parental individuals chosen. The simulation follows Mendelian segregations rules for a case of one gene with two alleles (Figure 47) but does not take into account recombination.

For example, if both parents (P1 and P2 in Figure 47) are homozygous for the reference or alternative allele then the offspring (F1 in Figure 47) will carry two reference or alternate alleles respectively. If one parental is homozygous for allele A, and the other for allele B, then the offspring will always be assigned a heterozygous genotype. In situations where there could be more than one outcome genotype, a function assigned the genotype randomly based on the according probabilities. For a graphic explanation of the process and rule see Figure 47.

4.3.8 Multi-Sample Kombinator (MuSaK)

ADMIXTURE is a well known piece of software with a maximum likelihood algorithm that is able to infer a number of *K* ancestry proportions of individual from *M* a number of genetic markers, typically SNP genotype datasets. From ideas about admixture already introduced by Cavalli-Sforza and my personal observations drawn from PCA plots, I developed an interest in seeing if linear combination of the ancestry proportions of m individuals can recreate the ancestry profile of a target individual of population in a reliable way.

What follows hereafter is the mathematical explanation of the rationale for the linear combination of ADMIXTURE proportions to calculate admixture coeffcients for any number of parental populations. I named the script implemented to do this *MuSaK* (*Multi-Sample admixture Kombinator*) which provides a least squares answer to the question above.

Without loss of generality we assume that the target individual is P_N and illustrate the procedure with $m = 2$, so that P_1, P_2, will be the individuals whose ancestries we want to combine.

Let us denote the ancestry proportions of individual P_i as provided by ADMIXTURE by

$$P_i : \{C_i^{(1)}, C_i^{(2)}, \ldots, C_i^{(K)}\}, \qquad i = 1, \ldots, N. \tag{1}$$

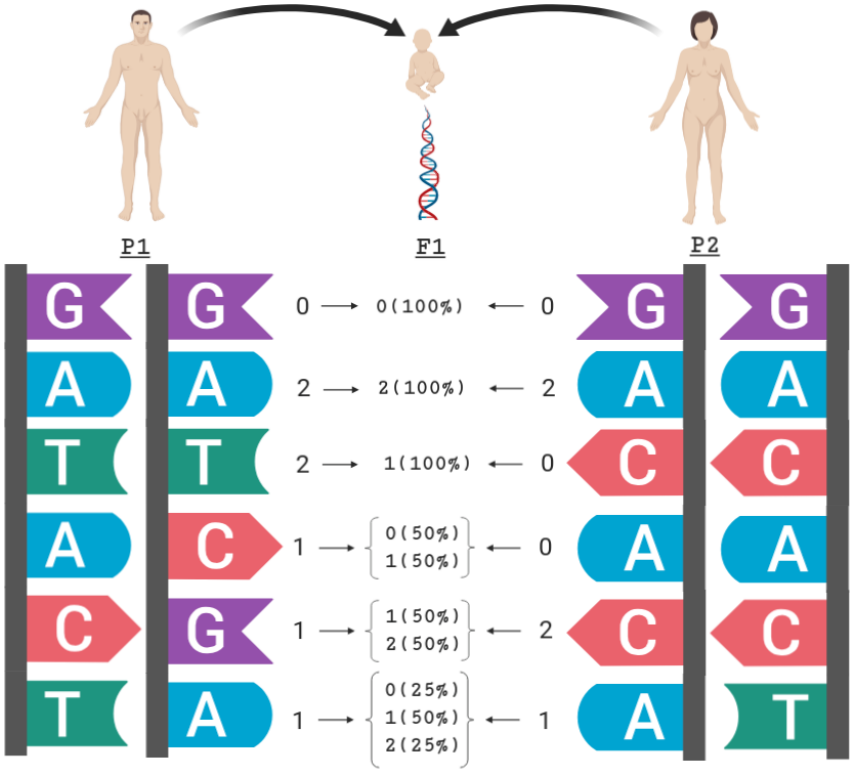


Figure 47: Parameters used to create the hybrids between two samples of choice based on the geno file genotypes. Where 0 indicates homozygous for the reference allele, 2 indicates homozygous for the alternative allele, and 1 indicates heterozygosity. Made with BioRender, own source.

Here, $0 \leq C_i^{(j)} \leq 1$, and $\sum_{j=1}^{K} C_i^{(j)} = 1$, for all individuals $j = 1, \ldots, N$. We want to find two ($m = 2$) real coefficients $0 \leq \alpha_j \leq 1$ such that the same linear combination holds for all ancestry proportions, namely

$$\begin{aligned} \alpha_1 C_1^{(1)} + \alpha_2 C_2^{(1)} &= C_N^{(1)} \\ \alpha_1 C_1^{(2)} + \alpha_2 C_2^{(2)} &= C_N^{(2)} \\ &\vdots \\ \alpha_1 C_1^{(K)} + \alpha_2 C_2^{(K)} &= C_N^{(K)} \end{aligned} \tag{2}$$

Thus, the coefficient α_j stands for the proportion assigned to individual P_j in the linear combination, with $j = 1, 2..$ We have to solve the algebraic linear system (2) with K equations and $m = 2$ unknowns: α_1, α_2. Since the number of ancestries is greater than the number of unknowns, $m < K$, the system is overdetermined and has not exact solution, as a rule. Moreover, if we sum up the K equations (2) we get the constraint $\alpha_1 + \alpha_2 = 1$, so that the system (2) becomes

$$\begin{aligned} \alpha_1 (C_1^{(1)} - C_2^{(1)}) &= C_N^{(1)} - C_2^{(1)} \\ \alpha_1 (C_1^{(2)} - C_2^{(2)}) &= C_N^{(2)} - C_2^{(2)} \\ &\vdots \\ \alpha_1 (C_1^{(K)} - C_2^{(K)}) &= C_N^{(K)} - C_2^{(K)} \end{aligned} \tag{3}$$

i.e., K equations and one unknown. We are then led to look for the least squares solution of the system. In this simple situation the solution may be expressed analytically

$$\alpha_1 = \frac{\sum_{i=1}^{K} (C_N^{(i)} - C_2^{(i)})(C_1^{(i)} - C_2^{(i)})}{\sum_{i=1}^{K} (C_1^{(i)} - C_2^{(i)})^2} \tag{4}$$

and $\alpha_2 = 1 - \alpha_1$. Remind that $C_1^{(i)}, C_2^{(i)}$, are numbers provided by ADMIXTURE proportions (K=9 in this particular case). Notice that the sums extend over the K ancestries.

Among the N individuals in the dataset, P_N is the target and so we are left with $N-1$ individuals. Therefore the number of pairs that can be studied is $(N-1)(N-2)/2$. We are interested in determining those pairs which give the lowest residual (RSS) R_2:

$$R_2 = \sum_{i=1}^{K} [\alpha_1 C_1^{(i)} + \alpha_2 C_2^{(i)} - C_N^{(i)}]^2 \tag{5}$$

where the subscript points out that we are combining pairs among the $N-1$ individuals.

The procedure generalizes to $m > 2$:

$$\begin{pmatrix} C_1^{(1)} & C_2^{(1)} & \cdots & C_m^{(1)} \\ C_1^{(2)} & C_2^{(2)}) & \cdots & C_m^{(1)} \\ \vdots & \vdots & & \vdots \\ C_1^{(K)} & C_2^{(K)}) & \cdots & C_m^{(1)} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} C_N^{(1)} \\ C_N^{(2)} \\ \vdots \\ C_N^{(K)} \end{pmatrix} \tag{6}$$

Using matrix notation, eq.(6) reads

$$C\vec{\alpha} = \vec{c} \tag{7}$$

where C is a matrix $N \times m$ and $\vec{\alpha}, \vec{c}$ are vectors of dimension N. The least squares solution of the system is

$$\vec{\alpha} = (C'C)^{-1} C'\vec{c} \tag{8}$$

provided the inverse $(C'C)^{-1}$ exists. The prime stands for matrix transpose and so $C'C$ is, in general, a square matrix and if its inverse exists then the least squares solution is found. Notice that the rhs in (8) is built up from the ancestry proportions provided by ADMIXTURE. The numerical task reduces merely to compute the matrix inverse. Eventually, to explore the best multiplet of size m among the $N-1$ individuals, we have to repeat the computation for the $(N-1)!/[m!(N-1-m)!]$ possible combinations of m individuals.

MuSaK code reads in the input file *MusaK-in.dat* with the set of ancestry proportions (as given by ADMIXTURE), the number of individuals and the size of the multiplet. The code repeats the computation for all possible combinations. The output is, for any given combination of m individuals, the percentage of every individual in the multiplet and the corresponding residual of that linear combination. The numerical values are sent to file *MusaK-out.dat*, altogether with the individual labels. Cases with negative percentages are discarded.

4.4 Results and Discussion

Here I present a summary with the most technical results of the sequencing, library preparation and processing of the successful newly generated shotgun ancient genomes (GOG samples, 11 females and 9 males, Table 6). I managed to build a solid scaffold for a temporal transect of the Valencian region, albeit small (16 GOG + 4 MS samples), stretching from the Late Neolithic (~5000 BCE) to the 17th century CE (Table 7). The samples with *MS* lab IDs were collected by me as part of my sampling effort, but were processed in the Ancient DNA facility by my colleague Marina Silva. I include them here but I do exclude them from some analysis due to low coverage.

ID	$Ntot$	$Ny+Nx$	Ny	Ry	SE	$95\%CI$	Sex
GOG05	7,554,603	215,751	19,283	0.0894	0.0006	0.0882-0.0906	XY
GOG06	7,370,287	361,147	2613	0.0072	0.0001	0.007-0.0075	XX
GOG11	35,282,406	1,705,131	10,900	0.0064	0.0001	0.0063-0.0065	XX
GOG20	19,609,928	1,017,593	5889	0.0058	0.0001	0.0056-0.0059	XX
GOG23	29,647,293	852,169	79,904	0.0938	0.0003	0.0931-0.0944	XY
GOG24	25,375,774	703,101	64,783	0.0921	0.0003	0.0915-0.0928	XY
GOG25	5,386,825	274,725	1687	0.0061	0.0001	0.0058-0.0064	XX
GOG26	101,717,219	5,055,942	26,885	0.0053	0.0	0.0053-0.0054	XX
GOG34	14,289,359	401,996	36,916	0.0918	0.0005	0.0909-0.0927	XY
GOG35	21,008,335	1,038,212	6477	0.0062	0.0001	0.0061-0.0064	XX
GOG38	4,856,202	137,715	13,173	0.0957	0.0008	0.0941-0.0972	XY
GOG50	87,759,042	4,400,676	25430	0.0058	0.0	0.0057-0.0058	XX
GOG56	9,200,463	462,558	2961	0.0064	0.0001	0.0062-0.0066	XX
GOG57	7,977,253	402,890	2739	0.0068	0.0001	0.0065-0.0071	XX
GOG59	9,504,636	270,760	25,628	0.0947	0.0006	0.0935-0.0958	XY
GOG60	8,326,646	241,977	23,111	0.0955	0.0006	0.0943-0.0967	XY

Table 6: Sex identification results of ancient human remains using DNA shotgun sequencing following the Ry index based on Skoglund et al. (2013). The ratio of number of chromosome X and Y sequences (Ny/Ny+Nx) in males and females follows different trends and is always higher in males than in females.

Endogenous DNA content of the 20 samples (10 petrous bones, 9 molars and 1 tarsus bone) that were selected for further sequencing after the screening stage ranged from 3% to 42%. However, out of the 279 samples collected (1 Upper Paleolithic, 3 Mesolithic, 5 Early Neolithic, 6 Late Neolithic, 27 Chalcolithic, 5 Bronze Age, 33 Iron Age, 41 Roman Period, 32 Visigothic Period, 109 Islamic Period, 17 Late and Post-Medieval) a total of 55 ancient samples were screened. The average endogenous content across the 55 samples was 6% (ranging between 0.01% and 42%), and only around a quarter of the samples had >10% endogenous DNA content. A minimum of 10% endogenous DNA was the lower threshold I considered originally to perform further sequencing on a sample, but this criterion was lowered for certain samples of particular archaeological interest. (Figure S8 in Supplement II).

Based on sequencing data previously produced by our lab, I was able to set up a method to predict mapped reads (which can be translated into coverage, 50-60 million mapped reads roughly correspond to 1X) that could be obtained for a sample taking into account endogenous content and the characteristics of the sequencing equipment (an Illumina HiSeq4000 lane produces 300 million reads). The prediction helped my sequencing strategy since I only had a limited budget (8 sequencing lanes). It worked out well because the correlation between endogenous content and coverage is fairly linear (Supplement II, Figure S9). However, since I processed my set of samples in two batches that were sequenced at different times, I encountered problems with the

adapter dimer in the first batch of libraries I prepared (the 11 samples from GOG05 to GOG35). In this first batch the adapter dimer percentage is higher than in the second batch (the 5 samples from GOG50 to GOG60), and for some samples is very high, up to 53% (Table 7). The amount of adapter dimer present in a library will decrease the coverage obtained for a sample because a lot of adapter is being sequenced in place of real DNA. This is the case for samples GOG05 and GOG23 whose coverage is much lower (about half) than what I originally predicted because the adapter dimer content in the libraries was 50%. This issue disappeared in the second batch.

The genomic coverage of the newly generated GOG samples ranged between 0.12X (GOG06, GOG38) and 2.34X (GOG26), and the mitochondrial coverages ranged from 9X to 182X. The genomic coverage of the MS samples is much lower (0.01-0.06X) because they were amongst the first ones to be sequenced by our research group and had no prior screening (Table 7).

In regards of the mitochondrial diversity, among the prehistoric samples I found haplogroups such as H1, H3, K1a, K1b and U5b, all typically found in either indigenous hunter-gatherer populations (U5b) or introduced in Iberia by the Neolithic migration wave (H and K). Haplogroup H is the most frequent in the Iberian peninsula today. The male lineages, two I2a and one G2a, are typically found in other Neolithic individuals from different populations across Europe.

The maternal lineage D4e1 of sample GOG50 (otherwise dubbed as Dafne) is unusual because it is virtually absent in past and present Europe. There are only two known modern public sequences from Central Europe, and another unpublished sequenced from Greece generated by our lab. Amongst ancient samples there is another D4 individual from Rome (R78) from about the same period (3rd century CE) (Antonio et al., 2019) and a few D4e1 prehistoric samples in Eneolithic Russia, Bronze Age Mongolia and Iron Age Kazakhstan (Jeong et al., 2018; de Barros Damgaard and et al, 2018*b*; Narasimhan et al., 2019).

The two samples from the Visigothic-Byzantine era carry mtDNA haplogroups found in Europe and the Near East, and the male individual is carrier of a Y chromosome haplogroup (R1b) that becomes ubiquitous after the Bronze Age in Iberia as well as the rest of Europe (Table 7).

In the Medieval samples I observe an increased diversity, although this could be helped by the fact that the sample size is bigger, both on the maternal and paternal side. Novel mtDNA types such as U6a1, L3d1 and R0a4 appear, an indication of incoming migrations from outside Iberia. U6a is overwhelmingly associated and common amongst native Maghrebi folk. L3d is a sub-Saharan lineage although it can be found sporadically in past and present Europeans without traces of African ancestry. There is at least one L3 mtDNA in a published Bronze Age Spanish sample (González-Fortes et al., 2019). Although R0a4 is found in modern Spain and it seems to be a subbranch that is most common in Iberia than elsewhere. There is no record of R0a type lineage found in Iberia before post-Roman times, and interestingly the roots of the R0a family tree trace back to the Arabian peninsula (Gandini et al., 2016). The paternal side also hints at the North African inflow of people to Iberia because of the proliferation of E1b lineages in the centuries around the Islamic conquest. E1b can be regarded as the male equivalent of mitochondrial haplogroup U6a because presents itself in Berber populations at high frequencies. Other haplogroups are also present, such as R1b, the most common male lineage nowadays in Spain, and the far less frequent J2a in the Islamic and post-Islamic samples (Table 7).

4.4.1 Kinship Analysis

I performed the kinship analysis test with READ on all my samples using all possible pairwise combinations. The output of the analysis provided by READ reveals that no samples from different sites or periods are related. The only relationship was a first degree of kinship between samples GOG34 and GOG35 from the same site which are dated by radiocarbon to the 6th-7th century CE (Figure 48). The two samples were found in a coastal settlement next to the river Serpis in a region and time characterized by the conflict between Visigoths and Byzantines in southern and eastern Iberia. Knowing GOG34 (male) and GOG35 (female) share a first-degree kinship, as well as considering their mitochondrial haplogroups ((HV+16311 and H2a1e1a respectively) I was able

Lab ID	Sample ID	Location	Period	Endogenous DNA	Adapter Dimer	Mapped Reads	Genome Coverage	Mitochondrial Coverage	Haplogroup mtDNA	Haplogroup ChrY	Genetic Sex	SNPs in HO 600k List	Kinship	Sample Type
GOG38	TF-12001	Tossal de la Font	Late Neolithic	10.00%	15.00%	4852709	0.12×	35.27×	K1a1	I2a1a	XY	61914	No	Molar
GOG05	A1P-Ind1	Abrigo I de las Peñas	Chalcolithic	22.00%	50.00%	7413953	0.14×	23.70×	H3	I2a1b2	XY	73538	No	Molar
GOG06	SPC09-Ind3	Sima del Pozo Cerdaña	Chalcolithic	10.00%	1.00%	6912972	0.12×	9.21×	K1b1a	-	XX	62699	No	Petrous
GOG11	SPC09-MD88	Sima del Pozo Cerdaña	Chalcolithic	30.00%	20.00%	33197859	0.53×	35.42×	U5b1	-	XX	226532	No	Molar
MS065	PdR-RdG	Vall d'Uixo	Chalcolithic	7.00%	12.00%	2577499	0.04×	3.19×	H3	-	XX	23912	No	Petrous
MS066	CD-Q1	Cova dels Diablets	Chalcolithic	3.00%	13.0%	905384	0.01×	2.08×	H1q	G2a	XY	9040	No	Molar
MS068	CI-00501	Cova Iguala	Early Bronze Age	6.00%	19.00%	1526493	0.03×	8.04×	H3	-	XX	16403	No	Molar
GOG50	2SABCIS-UE2600	Valencia	Late Roman	30.00%	5.00%	74393520	1.88×	156.27×	D4e1	-	XX	410970	No	Petrous
GOG34	SLlop50-A (adult)	Gandia	Visigothic-Byzantine	12.00%	15.00%	13653223	0.26×	19.47×	HV+16311	R1b1a1a2	XY	122004	Yes	Petrous
GOG35	SLlop50-B (kid)	Gandia	Visigothic-Byzantine	18.00%	16.00%	19893305	0.36×	40.63×	H2a1e1a	-	XX	164426	Yes	Petrous
MS060	Almu99-UE2298	Segorbe	Islamic	4.00%	17.00%	3519116	0.06×	19.96×	U6a1a1	E1b1b1b1	XY	37597	No	Molar
GOG20	Asuncion-UE7004	Vall d'Uixo	Islamic	15.00%	29.00%	19541513	0.47×	46.05×	J1c1b	-	XX	177882	No	Molar
GOG23	Peral10-2032	Vall d'Uixó	Islamic	35.00%	53.00%	29153340	0.63×	48.29×	HV*	E1b1b1b1	XY	230636	No	Petrous
GOG24	Peral10-2013	Vall d'Uixó	Islamic	18.00%	7.00%	24944366	0.51×	35.82×	U4a1d	J2a1b	XY	213353	No	Petrous
GOG25	Obon-Fosa19	Vall d'Uixó	Islamic	12.00%	13.00%	5405256	0.15×	12.81×	L3d1	-	XX	73257	No	Molar
GOG26	Obon-Fosa1	Vall d'Uixó	Islamic	40.00%	9.00%	100437963	2.34×	182.66×	H1	-	XX	434276	No	Petrous
GOG56	2SABCIS-UE2709	Valencia	Late Medieval	9.00%	6.00%	9135517	0.21×	14.12×	H+7720	-	XX	95546	No	Petrous
GOG57	2SABCIS-UE1789	Valencia	Late Medieval	7.00%	7.00%	7888239	0.17×	11.99×	R0a4	-	XX	80675	No	Petrous
GOG59	2SABCIS-UE1404	Valencia	Post Medieval	6.00%	2.00%	9508116	0.26×	18.76×	H5+152	E1b1b1b1	XY	127368	No	Tarsus
GOG60	2SABCIS-UE1822	Valencia	Post Medieval	6.00%	7.50%	8363273	0.23×	26.41×	K1a+195	R1b1a1a2	XY	107736	No	Molar

Table 7: Summary of genomic information for samples successfully sequenced with Illumina HiSeq4000. See Figure S8 in Supplement II for information about endogenous content of all 55 samples screened initially.

to conclude that the two samples correspond to a father and a daughter. This result is consistent with the archaeological information because the two samples were from the same site and buried together in a pit in strange positions (indicating that they were thrown unceremoniously into the dumpsite with their extremities tied). They both probably suffered from a violent or disease related death at the same time. In addition to that, it was also clear from examination of the skeletons that the male (GOG34) was an adult and the female (GOG35) was a infantile individual.

Another conclusion that can be extracted by looking at the rest of average pairwise P0 values is that combinations of prehistoric individuals (Neolithic and Chalcolithic) from distant sites and separated by some centuries tend to be borderline with the threshold for second degree relationship (P0 = 0.20). Whereas on the other hand, combination of individuals from the Islamic and Post-Islamic period sit at the end of the plot over the solid line (P0 = 0.22), even though many of them come from the same cemetery or settlement in Vall d'Uixó. This may be informative of the effective population size and/or degree of inbreeding of each population. The Chalcolithic population in eastern Iberia (Valencia) must have been relatively small and was more inbred, thus explaining why two individuals that should not be related *a priori* appear consistently borderline with a second-degree kinship scenario. The opposite applies for the Islamic Valencian population, because it was probably both larger and also enriched with genetic diversity by admixture as discussed below.

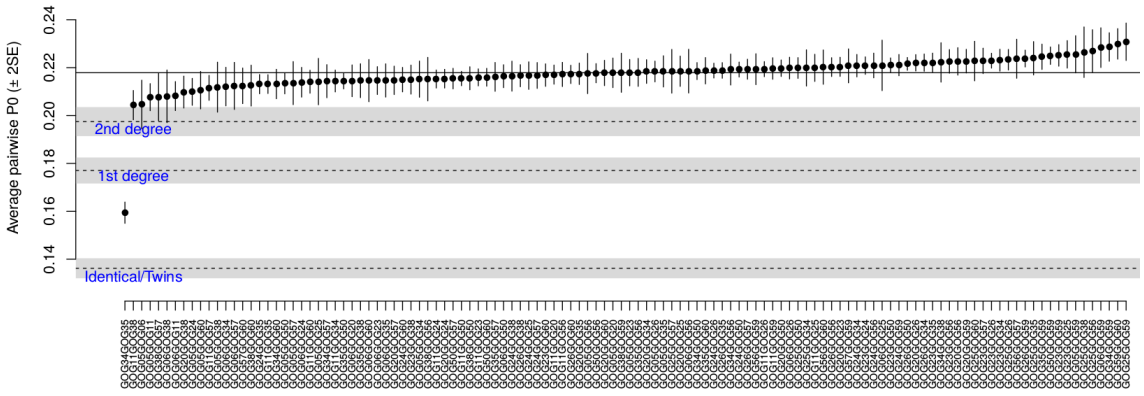


Figure 48: Results of the pairwise combisnations obtained with READ kinship analysis for the new GOG samples.

4.4.2 Principal Component Analysis

I performed several different Principal Component Analysis runs with four different datasets (Human Origins 620k SNPs, Human Origins 1240k SNPs, North African Affy 6.0 data (Arauna et al., 2017), and the Simon Genomes Diversity Project (SGDP) and also combining some of the datasets.

The first PCA plot (Figure 49) sets the ancient Valencian samples in the context of Human Origins dataset populations around the Mediterranean sea and its neighbouring regions. In the PCA plot, ancestral hunter-gatherer populations from Europe (Western Hunter-Gatherers) and North Africa (Iberomaurusians) occupy the top left corner and the bottom left corner respectively, whereas hunter-gatherers from the Near East (Natufians) and the extended Persian Plateau (Caucasus Hunter-Gatherers) are in the right side of the plot. These four groups form an irregular quadrilateral and in varying degrees they are the source of the ancestries that make modern European populations.

The position of the modern populations in the PCA resembles their distribution across the map of West Eurasia, with the white space in the middle being a vague representation of the Mediterranean sea. There is population structure mirroring geography within the modern coloured clusters of Europe, Caucasus & Persia, the Near East & Arbia and North Africa if one pays close attention and bears a geographical map in mind. For this instance I placed a three letter label next to some key populations (Finnish, Basque, Spanish, Sardinian, Sicilian, Italian, Greek, Turkish, Iran, Syrian, Yemeni, Saudi, Egyptian, Libyan, Tunisian, Algerian, Moroccan, Mozabite), to help the reader orientate while interpreting the plot.

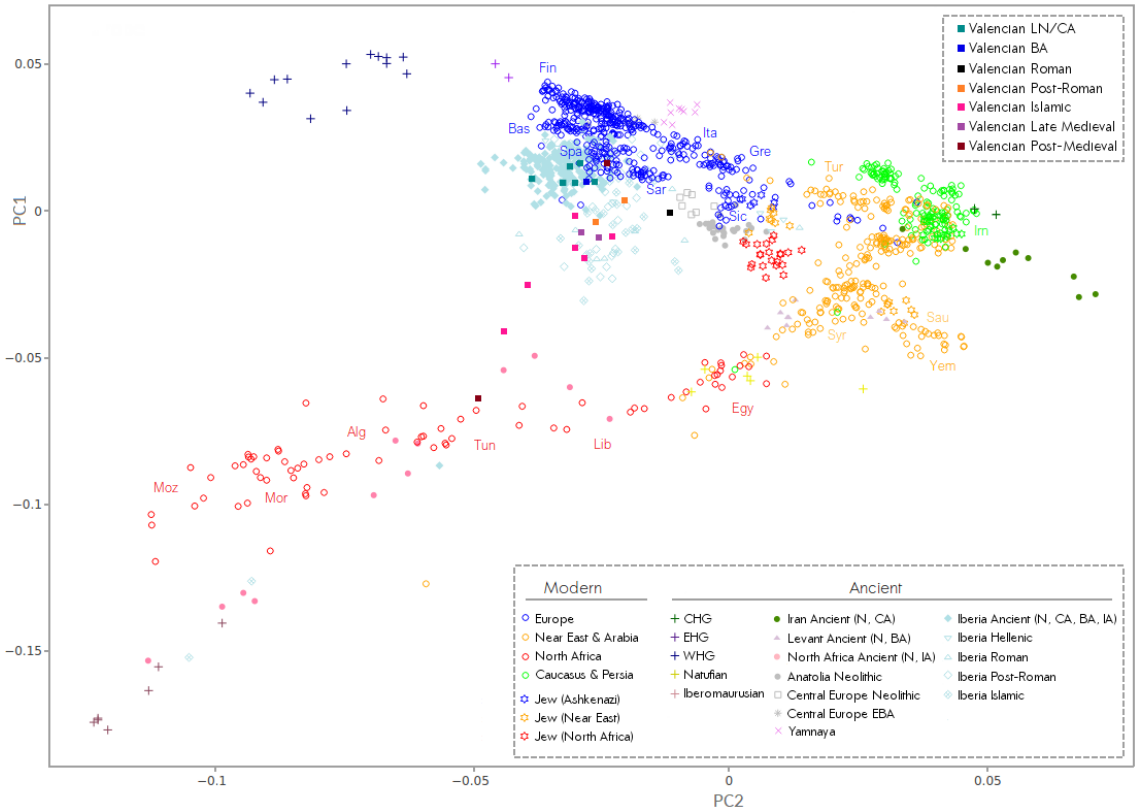


Figure 49: PCA with modern populations from Europe, Caucasus, Iran, the Near East, Arabia and North Africa. The GOG Valencian samples from all period are projected on top using *smartpca* along with various other published ancient populations.

The Valencian Neolithic (N) and Chalcolithic (CA) samples (which include GOG05, GOG06, GOG11, GOG38, MS065 and MS066) are in the same colour because it is well known from previous studies that for the case of the Iberian peninsula both periods are represented by roughly

the same population (Figure 49). All the Late Neolithic (LN) and Copper Age (CA) Valencian samples were recovered from excavations in caves scattered across the territory (mostly caves used for collective inhumations) which is the typical funerary tradition until the advent of the Bronze Age. In other words, there are no major cultural or genomic changes from the Neolithic to the Chalcolithic in Iberia. In fact, the position of the Valencian LN and CA samples that I have sequenced are in agreement with this fact. The Valencian LN/CA samples overlap in the PCA with the cluster of prehistoric Iberian samples that includes most published Neolithic and Copper Age genomes. This cluster is slightly shifted towards the left of the cluster of modern Spanish individuals and just under the cluster of modern Basques.

In this plot (Figure 49), the only Valencian Bronze Age (MS067) genome in my dataset also overlaps with the Iberian prehistoric cluster which includes Bronze Age (BA) and Iron Age (IA) genomes. The position of MS067 on the right half of the prehistoric cluster confirms its affinity to BA and IA Iberian individuals because it is where the majority concentrate, fully overlapping with the modern Spanish cluster. The LN and CA Iberian samples on the other hand, concentrate on the left half of the prehistoric Iberian cluster in this plot. However, this is a very subtle detail that cannot be seen clearly in this figure, but it can be better appreciated in Figure 50 and in the supplement with greater detail (Figure S12).

No Iron Age samples passed the screening test for my dataset, and the next available sample is from the city of Valencia in the Roman period (GOG50). Although there are no radiocarbon dates available, the stratigraphy of the excavation is clear and the archaeological dates places the sample in the Late Roman Imperial period (around the 3rd or 4th century CE). The current location where this burial was found is a public square but the area was an intramural Christian cemetery in the Middle Ages. However, in Roman times it was extra-mural but very close or adjacent to the city limits. Other evidences of Roman Imperial activity and archaeological remains have been found in the same area, including other burials one of which failed the screening test (GOG49 only had 1.87% endogenous DNA content).

The position of the Roman sample GOG50 is unusual because it sits almost alone in a space on the plot not occupied by any modern population, although still within the sphere of influence of the Mediterranean populations of the European cluster in blue (Figure 49). The black square is in between Sardinian and Spanish to the left and Sicilians and Maltese to the right but slightly shifted towards the bottom attracted in the direction of the Egyptian section from the North African cline.

Following the timeline chronologically, the post-Roman Valencian period is represented by two samples (GOG34 and GOG35), who are father and daughter as explained above. These two samples are also slightly shifted away from the prehistorical Iberian cluster as well as the modern Spanish cluster and towards the North African populations in a space with no modern individuals nearby (Figure 49). However, this area is occupied by several other public ancient samples from the post-Roman Visigothic era. This space is actually occupied by one modern population that is not well represented in the Human Origins dataset: Canary Islanders. There are only two Canary Islanders in the HO, which are the two blue open dots closest to GOG34 and the other Islamic-Medieval samples. The dataset from Arauna et al. (2017) contains more Canary Islanders along with the North Africans as it can be seen in the Medieval PCA from below (Figure 51), and that is the space shared with GOG34. Canary Islanders occupy this area because they have more North African ancestry than the average (ranging between 0-10%) (Botigué et al., 2013; Bycroft et al., 2018) of the general Spanish population and therefore appear closer to North Africa. This admixture signal is likely the genetic legacy left by the aboriginal inhabitants of the islands, the Guanches, who came from the coasts of Morocco at some point during the Iron Age.

Since we know the relationship between the samples an extra conclusion can be drawn from their position in the PCA. The distance between the locations of father and daughter in the plot is relatively large for first-degree of relatedness, more so than between some of the other non-related samples from the Islamic period that carry high diversity. Although the samples are projected, the coverage and number of SNPs is relatively high so it cannot be blamed on missing data. The

position of the father (GOG34) is the orange square closer to the other Islamic and Late Medieval individuals, and the daughter (GOG35) is the one closer to the Sardinian cluster. If we follow the logical conclusion that the offspring is the result of an intermediate contribution of two parental sources, then we can deduce that the genomic location of the mother in this plot – if the data existed – would be more or less at an equal distance as the father from the offspring but in the opposite direction. This means that the mother was most likely an individual of Italian or Sardinian origin (Figure 49).

The group of six samples (GOG20, GOG23, GOG24, GOG25, GOG26, MS060) from the 11th-13th centuries of the Islamic period are very heterogeneous. Four of them form a cluster (GOG20, GOG24, GOG25, GOG26) around the space where one of the post-Roman samples sits (Figure 49). In this area many other published Islamic individuals from south Spain (Olalde et al., 2019) are also present, along with some from the pre-Islamic period as mentioned above. This genetic space appears to consolidate as an Africanized Iberian cluster in the Islamic era, likely the results of a constant stream of genetic flow from North Africa since the times of the Roman Empire that grew with the Islamic conquest of Iberia. The cultural and genetic identity of these four samples can be associated with the Muladis, a social, and perhaps also ethnic, group of Al-Andalus.

The other two Islamic samples (GOG23, MS060) are just south of the Africanized Iberian-Islamic cluster at a distance similar to the one between modern Spain and the Islamic cluster (Figure 49). These two samples are two males and both trend even more in the direction of western Maghrebi populations, suggesting that they carry greater Northwestern African ancestries and are possibly the result of recent admixture (first or second generation). The position of MS060 is intermediate between Spain and Algeria–Morocco, which might mean this male individual is a half-and-half mixture of individuals from those two different populations. However, both his uniparental markers are of North African origin which indicates the scenario might be more complex. The individual GOG23 is beyond the position of MS060 and even further along this genetic bridge between Iberia and North Africa, which indicates a Northwestern African contribution to his genome greater than 50%. The paternal lineage of GOG23 is also of clear North African origin.

The two Late Medieval samples from the 14th and 15th century CE come from a Medieval Christian cemetery in the city of Valencia (*Cementerio de San Lorenzo*) and belong to a period where the city had been in Christian hands since 1238 CE. However, these two samples still appear clearly within the ancient Africanized Iberian-Islamic cluster (Figure 49). This means that even after one or two centuries since the Christian conquest of an important city like Valencia and a series of events such as: the disappearance of its former Islamic elite; confiscation and redistribution of buildings among the new Christian elites; expulsion of many native inhabitants and families; repopulation with Aragonese, Catalan, French and some other sporadic Mediterranean nationalities, the genetic make-up of the land had not changed much compared with the late Islamic individuals (11th-13th centuries). Much like the Islamic individuals probably represent the Muladi population of Spain, these two samples can be linked to the Morisco identity which refers to the Muladi people who converted to Christianity. That could explain the genetic continuity despite the massive political changes experienced at that time.

The remaining two samples are also from the same cemetery of San Lorenzo in the city of Valencia, but are post-Medieval since the burial ground was in use for a long time. The dates of these two samples (GOG59, GOG60) are 16th and 17th century CE respectively. Individual GOG59 comes from a burial where only the lower body was found and is archaeologically dated to the 16th century. What was interesting about this burial is that he had an iron shackle around his right ankle, interpreted as a sign that he was a slave. Although rare, this is not strange in light of the known records of slave trading in Medieval Valencia. The origin of the trafficked slaves was diverse but, judging by the position of individual GOG59 in this PCA (Figure 49) it is clear that he was North African in origin. The other post-Medieval samples dates to the 17th century CE and it is the first and only sample out the total of 20 from all periods that fully clusters within the cloud

of dots made up of modern Spanish genomes (Figure 49). There are no other published genomes from this post-Medieval period in Spain, but what the position of this sample reflects in the PCA is a genetic shift linked to a well-known historical event that happened in 1609 CE; the expulsion of the Moriscos from the kingdom of Valencia.

Prehistoric PCA plot

In this PCA plot (Figure 50) the focus is on the genetic transformation that happened in Iberia at the onset of the Bronze Age. The Iberian WHG, Anatolian Neolithic and Yamnaya clusters represent the three main poles of ancestry that make up all European populations. The prehistoric Iberian samples in light blue can be seen organized between two clouds of dots. The upper cloud, closer to the Yamnaya, consists of all Bronze Age and Iron Age genomes, whereas the lower cluster is made up of all the Neolithic and the majority of the Chalcolithic individuals in the dataset. Some samples from Copper Age both associated and non-associated to the Bell Beaker phenomenon appear in the Bronze Age cluster indicating the time of the beginning of the transformation. There is a clear separation between the pre-Bronze and post-Bronze clusters with CA samples acting as link. This reflects the arrival to Iberia of peoples from Central Europe carrying Steppe or Yamnaya-related ancestry that admixed with locals.

The Valencian samples are not an exception to this event of great genetic change. As the figure shows all the Valencian LN/CA individuals belong to the pre-Bronze cluster indicating that they do not carry Steppe-related ancestry (Figure 50). However, the only Bronze Age sample that I had (MS068), already appears in the post-Bronze Iberian cluster, which is the one more similar to the modern Spanish population. Radiocarbon dates indicate an age of around 4000 years old (2028-1884 cal. BC) for MS068, this allows to infer that by the horizon of 2000 BCE steppe ancestry had already arrived to Iberian Levantine territory. For more detail about each prehistoric Iberian dot see supplementary information (Figure S12).

Medieval PCA plot

This PCA plot (Figure 51) is a subset extracted from the figure in the supplementary material (Figures S14 and S15) generated combining the Human Origins Project and the North Africa Affymetrix 6.0 datasets. Here are represented three relevant groups of populations that act as poles of ancestry for Europe, the Near East and North Africa. In this PCA all the Medieval and post-Medieval GOG samples were projected using *smartpca*. Furthermore, the genotypes of 25 hybrids simulated combining five random Spanish and five randomly chosen North African (Guanches) genotypes using the NMG script were also projected.

As expected the position of the hybrids is intermediate between the European and Northwestern African cluster. Sample MS060 overlaps with the hybrid cluster which is further confirmation of an ancestry intermediate between those in Iberia and the Maghreb.

The Islamic individual GOG23 also appears closer to the Maghreb cluster than MS060 here, which reinforces the idea that his North African proportion of ancestry is greater than fifty percent.

The putative slave individual GOG59 appears well within this North African cluster (Figure 51) that has the addition of the individuals from Arauna et al. (2017) (although not all are displayed here, see Supplement II for complete Figures S13 and S14). In fact, when evaluated with the North Africa Affymetrix 6.0 dataset (Arauna et al., 2017) it becomes obvious that there is part of the North African population that carries European admixture and is distinct from the heterogeneous Berber groups that are characterized by isolation by drift and for having varying degrees of sub-Saharan admixture. In the North Africa Affymetrix 6.0 dataset, GOG59 appears in the sub-Saharan gradient of Berber populations together with three Algerian Berber individuals from Timimoun and two Moroccan Berbers from Errachida. Both Errachida (Morocco) and Timimoun (Algeria) are hinterland enclaves beyond the Atlas mountain range. This result points towards a specifically Berber origin of GOG59 rather than a generic North African Arabized source.

Figure 50: PCA with three poles of ancestry of and Iberian and Valencian prehistorical samples. Yamnaya are proxy for the Steppe Bronze Age related ancestry, whereas WHG and ANF represent the indigenous European hunter-gatherers and the Anatolian Neolithic farmers respectively. Only includes GOG samples, MS samples were under 10k SNPs, see Supplement II for Figure S12.

The other ancient Islamic samples (GOG24, GOG25, GOG26) and the Late Medieval ones (GOG56 and GOG57) still appear differentiated from the modern Spanish cluster and more similar to the cluster of modern Canary Islanders (Figure 51). The only surprise in regards to the general PCA is a male Islamic individual (GOG20) who is much closer to the Near Eastern pole. This PCA was useful to examine whether paternal lineages in Islamic samples (e.g. J2a in GOG24) can be indicative of higher Near Eastern ancestry. In the same way that E1b1b1b-Z827 and E1b1b1b-M81 seem to be an indicative proxy for higher North African ancestry.

In the case of J2a, this Y-chromosome haplogroup is rare in North Africa but very common in the Fertile Crescent area. However, given the results obtained here, I cannot conclude that J2a is indicative of higher Syrian or Arab (*sensu lato*) ancestry because GOG24 is not attracted to the Near East cluster more than others. On the other hand, individual GOG20 is a very likely candidate for having a Near Eastern grand-parent based on the distances between potential ancestries in the PCA, but sadly no paternal lineage information is available since she is a female individual.

Finally, the 17th century post-Medieval individual GOG60 remains clustering with all other modern Spanish individuals (Figure 51), hinting at the magnitude of the depopulation caused by the expulsion of the Moriscos, and the effect and genetic transformation derived from the wave of

repopulation by Aragonese, Navarros and Catalans after 1609 CE.

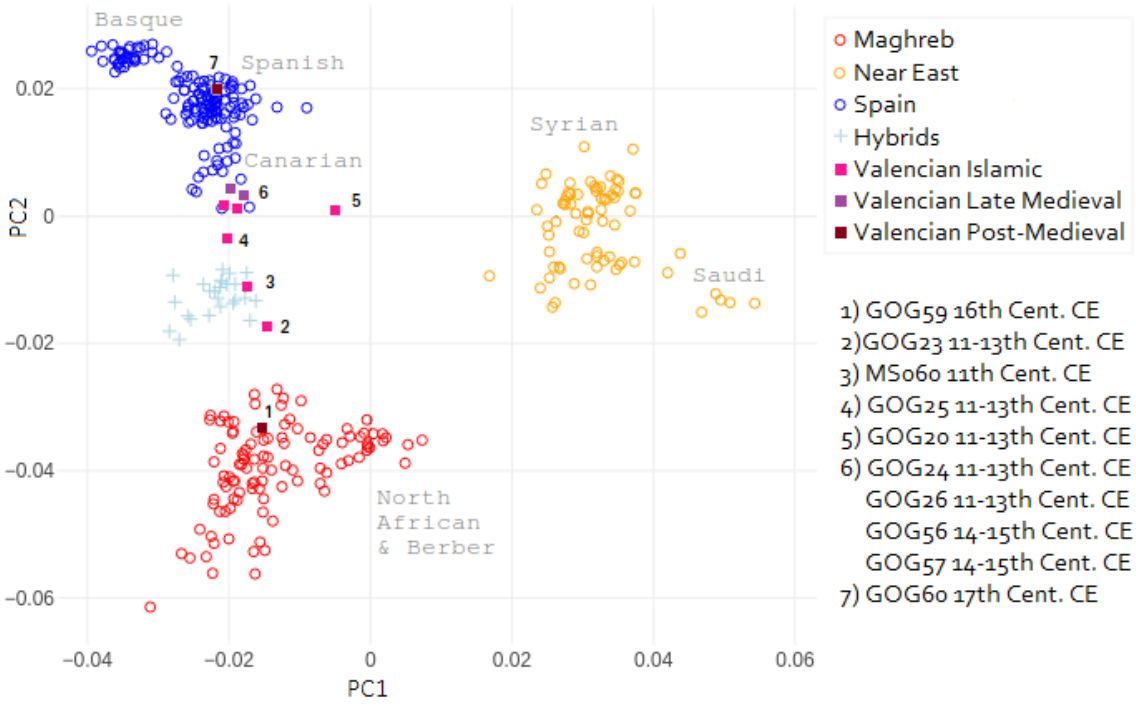


Figure 51: PCA with three poles of ancestry of historical ancient samples projected onto modern populations of the combined Human Origins and North African Affymetrix 6.0 datasets.

4.4.3 ADMIXTURE analysis

I ran different combinations of ADMIXTURE datasets, supervised and unsupervised, with only ancient samples above 10k SNPs and ancient samples together with European, West Asian and North African modern populations extracted from the Human Origins. I repeated these combinations using transitions and transversions together and filtering for transversions only. Results were fairly consistent across all modalities, although the ADMIXTURE runs are more stable and the cross-validation errors tend to be lower when more modern and diverse populations are added. Reducing the amount of ancient samples with low numbers of SNPs also helped to avoid artificial separations of modern and ancient into different clusters. In the ADMIXTURE plots in supplementary material (Figures S16 and S18) at the end of this chapter there are examples of different types of runs available.

In the supervised mode I selected as ancestral donor populations only hunter-gatherer groups (Figure 52).These groups acted as proxies to represent the ancestry layers present in modern European, in some cases introduced by later populations related to those hunter-gatherers and not directly by them. Typically this would require three sources (Western Hunter-Gatherers, Caucasus Hunter-Gatherers and Natufians) but since the information obtained from the PCA already points to Maghrebi genomic contribution to Iberian populations I increased that number to four (K=4) donor population by adding North African Hunter-Gatherers (NAHG). To represent European indigenous ancestry, I chose indigenous European hunter-gatherers (WHG) and for the Neolithic component I selected the Natufians whose ancestry predominates in the first Neolithic farmers that colonized Europe. I took the two individuals that make up the Caucasus Hunter-Gatherers group to represent the Steppe ancestry introduced into Europe with the Bronze Age by proto-Indo-European Yamnaya-related folk. Finally, I selected Iberomaurusians as the source of the Hunter-Gatherer ancestry indigenous to North Africa.

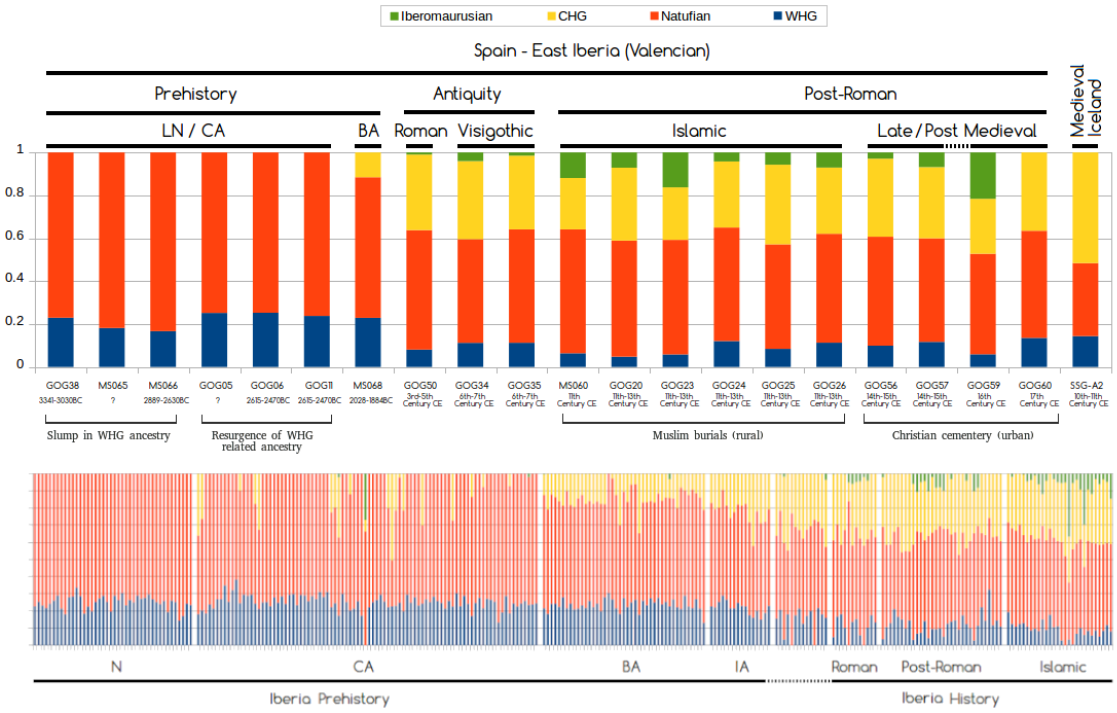


Figure 52: Supervised ADMIXTURE (K=4) with the HO 620k SNPs dataset. After LD and MAF (5%) filtering, over 200k variants remained. Only samples with at least 10k SNPs covered are displayed here. Top row shows a close-up of GOG samples and a re-analyzed high-coverage Medieval Icelandic individual from Ebenesersdóttir et al. (2018) for comparison. Bottom row displays all other public Iberian ancient genomes from the Neolithic to Middle Ages. The dotted lined corresponds to the section of Iron Age samples identified as Greek/Hellenistic and the transitional time when writing systems started to develop in Iberia.

ADMIXTURE profiles of the prehistorical samples

The supervised ADMIXTURE at K=4 for the prehistoric Valencian samples reveals patterns already hinted at by the PCA (top row of Figure 52). First of all, there are no surprises about the Late Neolithic and Copper Age samples who are a mixture composed of one-quarter WHG and three-quarters Natufian-Neolithic ancestry. There is an apparent decrease in WHG ancestry at the beginning of the Chalcolithic when ordering the LN/CA individuals chronologically according to their radiocarbon dates. However, this could well be the result of low sample size and chance rather than a local dynamic or phenomenon. However, it can also be minimally appreciated in the country-wide Spanish Chalcolithic genomes (see Supplement II Figure S20 and S21, after 5000 years BP).

The profile of the lone early Bronze Age sample MS068 in Figure 52 provides confirmation of what the prehistoric PCA pointed out above, a fraction of the ancestry of MS068 can be associated to genetic diversity introduced by individuals originally from the Pontic-Caspian steppe. The CHG-related ancestry is small (around 5%) but that is to be expected since even modern Spaniards carry low amounts of this ancestry compared with other Northern Europeans, although in the Iron Age and later this component increases its frequency as it can be seen in the bottom plot of Figure 52. This finding is interesting because it is from a multiple inhumation in cave (therefore not Indo-European tradition) and although a decades-old publication labelled it as early Bronze Age based on archaeological and fragmentary pottery, this has always been disputed, precisely because the dating sits at the very time of transition between periods. This result sheds light on the issue; perhaps the individual was not culturally associated to Bronze Age culture but his genomic profile is undoubtedly like those of the Iberian Bronze Age.

ADMIXTURE profiles of the Roman samples

The next sample GOG50 (a female dubbed Dafne) chronologically represents a big leap forward in time because it is dated to around the 3rd Century CE; this is the very Late Roman period, some 2500 years later. It can be seen that the yellow steppe-CHG component has increased, as mentioned above, and that the WHG-related ancestry has decreased considerably to a level at which it will remain until the Late Middle Ages. For the first time in this time transect there is also the subtle presence of North African Ancestry in a Valencian individual (1%). Overall her ADMIXTURE profile is similar to some other Roman and Post-Roman individuals from Iberia published before (Figure 52).

The ADMIXTURE proportions of GOG34 and GOG35 (father and daughter) are not too dissimilar to GOG50. These two samples are dated by radiocarbon to the 6th-7th Century CE, so only one or two centuries later than the Roman sample. The transition from the 6th to 7th Century CE coincides with the climax of the Visigothic-Byzantine war in southern and eastern Spain, precisely where these two samples where excavated. Father and daughter look very similar with the only distinction being the amounts of North African ancestry in each one of them. The daughter (GOG35) has about half (2-3%) of the Iberomaurusian-related ancestry that the father displays (5%).We do not know who the mother was, but this makes sense in a scenario of 50-50% parental contribution since it has already been discussed that the mother was of Spanish, Sardinian or Italian extraction and did not contribute any North African diversity to her offspring. The conclusions drawn from the PCA and the ADMIXTURE results coincide on this point.

ADMIXTURE profiles of the Islamic samples

The next few samples in line are six Islamic samples from burials in the Valencian rural world of the large farmland cultivated plains between the mountains and the sea (Vall d'Uixo). Sample MS065 is actually from this sample location but dated to the Chalcolithic, which helps us to appreciate the genetic shift over time. Sample MS060, dubbed the "Segorbe Giant" due to his considerable height compared to the other 40 individuals from the same Islamic necropolis, is from an important medieval fortified settlement (Segorbe) in the same mountain range but further up the course of the river Palancia. The most remarkable feature in these Islamic genomes is the consolidation of Iberomaurusian-related ancestry, ranging between 5% to 17%. Sample GOG23 is the one with the most (17%), and in the PCA is the one who appears closest to the North African populations, whilst still being intermediate (he is also a carrier of paternal lineage E1b). Whereas the Segorbe Giant (MS060) has a 10% Iberomaurusian ancestry proportion, and as commented on in the Principal Component Analysis section MS060 has an equidistant position between Spain and Western North Africa. This lets us conclude that an individual with 20% or more Iberomaurusian-related ancestry indicates that he or she is a native Maghrebi. I re-aligned and included a high-coverage Medieval genome from Iceland (SSG-A2) from the same centuries as the Islamic samples and the contrasts between the ADMIXTURE profiles of the same time period in Spain and Iceland is striking (Figure 52) regarding the amount of CHG-related ancestry. Another detail that becomes obvious among the Islamic samples is that a trade-off between WHG and Iberomaurusian ancestry exists, more of one generally means less of the other. This also applies for the CHG ancestry but not for the Neolithic-related ancestry because there is already similar Natufian ancestry in both Spanish and North African medieval populations and the dual contribution keeps the proportion stable.

The last four samples are uncharted territory because they are archaeologically dated between the 14th and 17th centuries, which makes them the only ancient genomes representative of this period among the ones published to date. The most complete temporal transect is the one from individuals in Olalde et al. (2019) but there are no post-Islamic samples in that dataset. What we have here are four individuals from the same Christian Medieval cemetery in Valencia city – now disappeared – that were buried in different centuries while the burial ground was in use before it

was dismantled in 1841 CE.

ADMIXTURE profiles of the post-Islamic samples

The ADMIXTURE profiles of the post-Christian conquest individuals from the 14th-15th century still carry significant Maghrebi ancestry. This is confirmation of what it was seen in the PCA but not less surprising (Figure 52 and 53). It is widely believed that James I expelled if not all, a huge majority of the original inhabitants of the city and distributed houses among his Aragonese, Catalan and French supporters. What is thought to have happened in Valencia city is in contrast with what happened in the rural regions where everything remained the same for another three or four centuries. However, what these two genomes reveal is that even after the conquest of the city in 1238 CE and chronicled replacement of the original inhabitants, the genetic make-up of the city did not change substantially compared to the Islamic period. People probably changed religion in order to remain living in the region, judging by the fact that these two individuals were buried following Christian custom.

It has been established already in the previous sections that the individual GOG59 had been identified archaeologically as a slave and genetically as of Berber origin. In his ADMIXTURE proportions, the Iberomaurusian-related fraction is slightly greater than 20%, which confirms the hypothesis discussed two paragraphs above (Figure 52). As explained in detail in the introduction, slavery was a common practice in Valencia during Medieval and post-Medieval times. However, there was great diversity in the provenance of the slaves (from sub-Saharan to Tatar). The results presented so far, along with the North Africa Affymetrix dataset PCA plot in the Supplement II (Figures S13, S14 and S15), now allow us to draw a firm link with Berber groups from the Maghrebi hinterland of southern Morocco and Algeria. DNA cannot inform us about the place of birth of an individual for certain, nor where a person lived, only where he or she died. Unlike the case of the Segorbe Giant (MS060) and some of his contemporaries, no isotopic data has yet been generated for this individual to shed light on the circumstances of his captivity.

Finally, individual GOG60 indicates that a genetic shift did occur in the city of Valencia, but much later than expected, given that the chronicles set the conquest of 1238 CE as a triggering episode. The ADMIXTURE profile of GOG60 is in line with his distinctiveness in the PCA; chronologically it is the first individual after twelve other samples since the Bronze Age not to have traces of Iberomaurusian ancestry in his genome (Figures 52 and 53). It is very interesting that this pattern of Maghrebi ancestry is broken with a sample from the 17th century, because it was at the start of that century, in the year 1609 CE, when the population of loose Moorish origin (the Moriscos) of the kingdom of Valencia was finally expelled and deported. The expulsion was enforced by the military, as was the suppression of the revolts that sparked following the decree. The reflection of this event has many examples in contemporaneous paintings. Perhaps GOG60 was someone who migrated from the neighbouring regions of Catalonia, Aragon or Navarra, attracted by the prospects of land redistribution to repopulate the emptied towns and orchards of the Valencian region. The origin of these new northern Iberian settlers, who likely did not carry signals of African ancestry in their genomes, is very well reflected in the surnames that are present in the modern Valencian populations. Many of the re-settlers surnames can be easily traced to their places of origin in Navarra, Aragon and Catalonia. Hundreds of thousands of Valencians nowadays carry surnamed from villages in those territories. For example, according to the Spanish National Institute of Statistics (INE) census over 35% of the town names from the Catalan province of Lleida have become surnames in Valencia, but still are rare elsewhere in Spain, even in the region they refer to. Navarro, Navarrete, Zaragozá, Catalá are only the most obvious ones. Perhaps, GOG60 was one of these re-settlers and new Valencian-to-be in the city but it would require a fine-detail analysis and a large Spanish dataset such as that in Bycroft et al. (2018) (not easily accessible) to do so.

Overall, the trends that the Valencian samples follow over time run in parallel to the ADMIXTURE results of the Iberian ancient samples in the lower plot of Figure 52. Especially relevant

is that with only twenty samples it is possible to observe that the North African ancestry starts to infiltrate the general Iberian population as well as the Valencian one well before the Islamic conquest, as suggested in Olalde et al. (2019). Since Olalde et al. (2019) is the only source of a handful Iron Age genomes from the Valencian region and a few extra Islamic ones, I decided to plot them in a ternary plot together with my own samples to see if there were any relevant differences to comment on. Each ternary plot is made with the supervised ADMIXTURE (K=4) proportions but excludes one of them and recalculates the other over one hundred. Figure 53 includes published Bronze and iron Age samples from the Valencian territory not included in the ADMIXTURE plot from Figure 52, and shows in a clear way that chronologically, the Bronze and Iron Age Valencian samples are the first to have significant CHG-related (as proxy for Steppe) ancestry. However it also shows that that proportion is yet to increase another 10-20% until Roman times and from then it stays stable in the subsequent periods. Regarding the newly added Islamic Valencian public samples, there is heterogeneity in the amounts of Maghrebi ancestry they carry. None reaches 20% but some are close to zero.

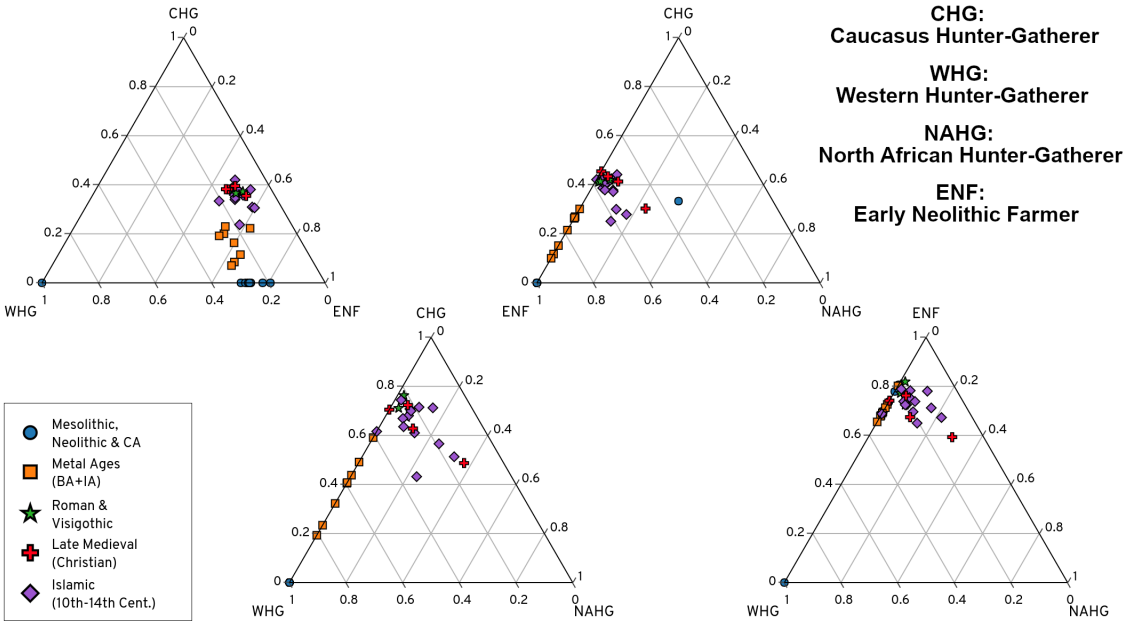


Figure 53: Terniary plots with supervised ADMIXTURE (Caucasus Hunter-Gatherer, Western Hunter-Gatherer, Early Neolithic Farmer and North African Hunter-Gatherer) proportions (K=4) for all Valencian samples from all periods.

4.4.4 Coefficients of Admixture Contribution by MuSaK

Figure 54 is an older unsupervised run of ADMIXTURE with a subset of ancient and modern populations from the Human Origins dataset, from a run of ADMIXTURE carried out in 2018 at a time when sample MS060 (Segorbe Giant) was the only Islamic sample sequenced by our lab. I used MuSaK and the ancestry proportions to calculate if any linear combination of these populations produce the profile of the last bar (individual MS060-Segorbe).

Figure 54 is the number of fractions K (K=9) with the lowest cross-validation error (K=9). The ADMIXTURE was carried out including all the populations in the Human Origins plus MS060 and the ancient populations in the left plot of Figure 54. The plethora of data from Olalde et al. (2019) was not yet in the public domain, so Iberian Chalcolithic samples were the best available proxy for the Iberian fraction. The figure illustrates the original motivation for the MuSaK software and making linear combinations of ADMIXTURE profiles. RSS stands for Residual Sum of Squares and it is the method to measure the error of the fit.

Figure 54 provides the top-10 ranking of combined population pairs with lowest RSS values, with the *Segorbe Giant* (MS060) as the target of the admixture. The least squares solution for some pair of combinations turned out negative for one of the coefficients. This was the case for 22 out of the 78 pairs attempted here. These results have no meaning in relation to the problem at hand and were discarded. For example, combining Yoruba and Ju'Hoan would never yield a valid solution. I found that for the case of MS060, using these ADMIXTURE proportions, the best combinations was some form of Iberian proxy plus a North African proxy in proportions close to 50% for each contributor, which makes sense in terms of the PCA, as mentioned earlier.

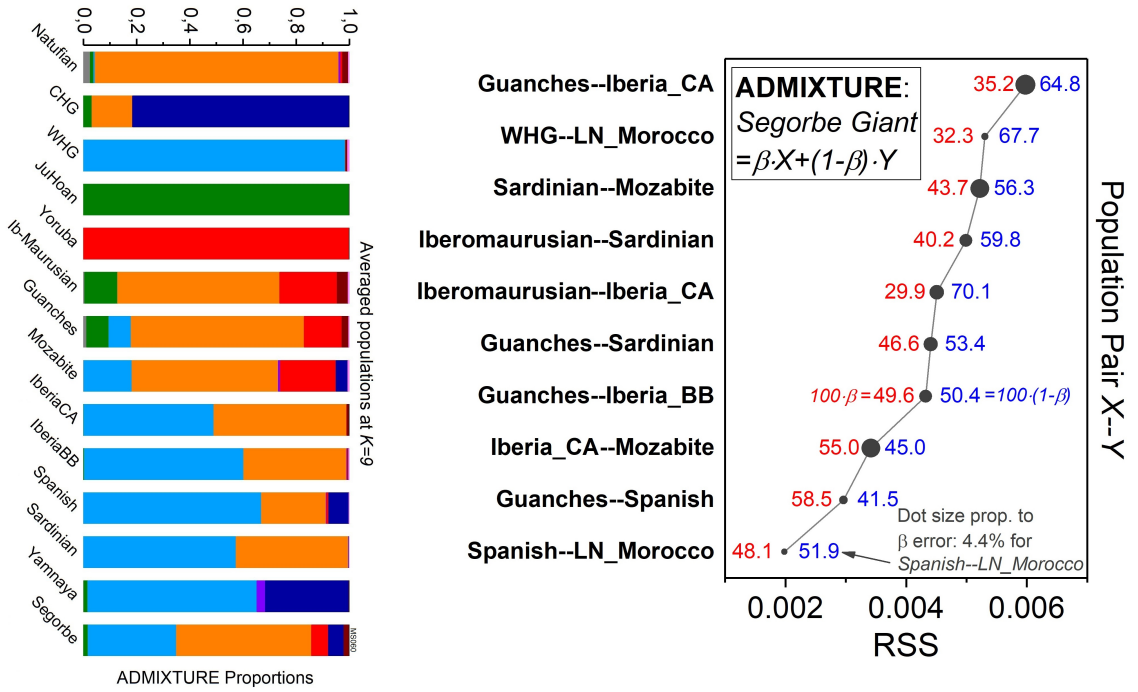


Figure 54: ADMIXTURE plot with K=9 ancestries (colours) and N=14 averaged populations (columns). The height of columns stands for the ancestry proportion (left). MusaK results for the top 10 results of pairwise combinations according the residuals sum of squares to measure the best fit (right).

However, since this is an old exercise and there is more data available, I repeated the analysis for some more of the GOG samples that appear to be product of admixture events based on the results shown so far. Because this was done with a different ADMIXTURE run, that only included a subset of populations from the Human Origins dataset, it was relevant to see whether there was consistency in the outcomes. This approach can be interpreted as an alternative to *qpAdm* for anyone who is not comfortable with the lack of a formal explanation of how *qpAdm* works. MuSaK is a method in development, however, and it is not limited to combining ADMIXTURE profiles of different number of *K*. I have also attempted the same using PCA coordinates, ten PCs in this case, because that is what *smartpca* output provides, but the more that can be provided, the better. When using 10 PCs the results were less clear than when using ADMIXTURE profiles but yielded the same conclusions. I decided not to include that part here to simplify the introduction of this methodology here. I aim to implement in the script the ability to do linear combinations using genotypes of hundreds of thousands SNPs directly, which is even more similar to how *f-statistics, D-statistics, qpAdm, etc...* make use of allele frequencies. However, sometimes one wants to interrogate a single ancient individual and I regard it as a methodological weakness to work with allele frequencies of a single individual, (they can only be 0 or 100% in pseudo-diploid genotypes which is rarely the case in a real populations).

Table 8 displays the best results according to the RSS values obtained performing linear com-

binations of the four K fractions of the unsupervised ADMIXTURE with the lowest CV error (K=4, supplementary Figure S19). For this instance I used only the ancient populations available as parental contributors since it intuitively makes more sense than using modern populations as contributors of the ancient individuals. However, I find it valid and not necessarily wrong to use modern populations as parental contributor as proxies.

$Target$	$\%P_1$	$Parental_1$	$\%P_2$	$Parental_2$	RSS	$Period$
GOG05	0.72	Anatolia N	0.28	WHG	<0.0001	Copper Age
GOG05	0.77	Levant N	0.23	WHG	0.0190	Copper Age
GOG05	0.80	Natufian	0.20	WHG	0.0784	Copper Age
GOG06	0.73	Anatolia N	0.27	WHG	<0.0001	Copper Age
GOG06	0.78	LevantN	0.22	WHG	0.0195	Copper Age
GOG06	0.81	Natufian	0.19	WHG	0.0807	Copper Age
GOG11	0.74	Anatolia N	0.26	WHG	<0.0001	Copper Age
GOG11	0.79	Levant N	0.21	WHG	0.0198	Copper Age
GOG11	0.81	Natufian	0.19	WHG	0.0817	Copper Age
GOG38	0.74	Anatolia N	0.26	WHG	<0.0001	Copper Age
GOG38	0.79	Levant N	0.21	WHG	0.0200	Copper Age
GOG38	0.82	Natufian	0.18	WHG	0.0827	Copper Age
GOG50	0.08	Armenia BA	0.92	Iberia Roman	0.0005	Late Roman
GOG50	0.26	Levant BA	0.74	Iberia Post-Roman	0.0009	Late Roman
GOG50	0.10	Levant BA	0.90	Iberia Roman	0.0011	Late Roman
GOG50	0.30	Levant BA	0.70	Iberia Hellenic	0.0013	Late Roman
GOG50	0.12	Armenia BA	0.88	Iberia Hellenic	0.0036	Late Roman
GOG50	0.04	Armenia BA	0.96	Iberia Post-Roman	0.0036	Late Roman
GOG50	0.02	Morocco LN	0.98	Iberia Post-Roman	0.0037	Late Roman
GOG50	0.05	Guanche	0.95	Iberia Hellenic	0.0052	Late Roman
GOG50	0.02	Morocco LN	0.98	Iberia Hellenic	0.0054	Late Roman
GOG20	0.26	Levant BA	0.74	Iberia Islamic	0.0003	Islamic
GOG20	0.39	Levant BA	0.61	Iberia Post-Roman	0.0024	Islamic
GOG20	0.19	Guanche	0.81	Iberia Roman	0.0027	Islamic
GOG20	0.29	Levant BA	0.71	Iberia Roman	0.0039	Islamic
GOG20	0.20	Guanche	0.80	Iberia Post-Roman	0.0047	Islamic
GOG23	0.55	Guanche	0.45	Iberia Roman	0.0013	Islamic
GOG23	0.59	Guanche	0.41	Iberia Hellenic	0.0019	Islamic
GOG23	0.30	Morocco LN	0.70	Iberia Islamic	0.0025	Islamic
GOG23	0.63	Guanche	0.37	Iberia IA	0.0028	Islamic
GOG23	0.57	Guanche	0.43	Iberia Post-Roman	0.0028	Islamic
GOG23	0.50	Guanche	0.50	Iberia Islamic	0.0032	Islamic
GOG23	0.35	Morocco LN	0.65	Iberia Post-Roman	0.0055	Islamic
GOG24	0.14	Guanche	0.86	Iberia Hellenic	<0.0001	Islamic
GOG24	0.08	Morocco LN	0.92	Iberia Post-Roman	0.0001	Islamic
GOG24	0.53	Iberia Hellenic	0.47	Iberia Islamic	0.0006	Islamic
GOG24	0.08	Guanche	0.92	Iberia Post-Roman	0.0008	Islamic
GOG24	0.08	Morocco LN	0.92	Iberia Hellenic	0.0010	Islamic
GOG24	0.08	Guanche	0.92	Iberia Roman	0.0010	Islamic
GOG24	0.77	Iberia Post-Roman	0.23	Iberia Islamic	0.0012	Islamic
GOG24	0.03	Morocco LN	0.97	Iberia Roman	0.0014	Islamic
GOG25	0.07	Guanche	0.93	Iberia Post-Roman	0.0043	Islamic
GOG25	0.10	Guanche	0.90	Iberia Roman	0.0082	Islamic
GOG25	0.17	Guanche	0.83	Iberia Hellenic	0.0091	Islamic
GOG26	0.24	Guanche	0.76	Iberia Hellenic	0.0002	Islamic
GOG26	0.18	Guanche	0.82	Iberia Post-Roman	0.0002	Islamic
GOG26	0.10	Morocco LN	0.90	Iberia Post-Roman	0.0009	Islamic
GOG26	0.05	Morocco LN	0.95	Iberia Islamic	0.0012	Islamic
GOG26	0.20	Iberia Hellenic	0.80	Iberia Islamic	0.0013	Islamic
GOG26	0.19	Guanche	0.81	Iberia Roman	0.0014	Islamic

Table 8: The most realistic 2-way combinations obtained with MuSaK for the Chalcolithic, Roman and Islamic samples that are clear results of suspected admixture events. ADMIXTURE profiles are shown in the supplementary information section (Figure S18).

MuSaK ancestry modelling: admixed prehistorical samples

The results in Table 8 for four of the Chalcolithic samples (GOG05, GOG06, GOG11, GOG38) make sense and are very much in line with results in the literature using qpAdm to model the same scenario. Iberian individuals from the Chalcolithic can be explained as a mixture of three-quarters Anatolian Neolithic and one-quarter Western Hunter-Gatherer. These tests were more of a sanity check rather than a challenging question. Since I was happy with the results, I tried to model some of my more complex individuals using the MuSaK method (Table 8). Note that this approach, like *qpAdm* or any other, is only as good as how close the two proxy contributors are to the real contributors of the admixture product. There are combinations that are mathematically possible but that could have never happened in real life. Note that the Moroccan Late Neolithic here is represented by the only samples available, from only a single site (Khelif el Bourud), and this group shows substantial evidence of incoming European gene flow in its genomic and maternal lineages (Fregel et al., 2018). This is why I argue that Guanches are a better proxy (Rodríguez-Varela et al., 2017).

MuSaK ancestry modelling: admixed Roman samples

For the modelling of the Late Roman sample GOG50 I found that Roman (90-92%) and Post-Roman (74-96%) Iberia show up as the majority contributor. The averaged ADMIXTURE of both profiles are very similar, but the earlier Iberian Roman source appears to be an even better fit overall (see Figure 52 and supplementary material Figure S18 for reference). The Iberian Hellenic ADMIXTURE period profile is also similar to the Roman and Post-Roman periods but lacks the North African component and that is why it appears as a worse fit as majoritarian parental contributor in the ranking, so I discard it in favour of the two mentioned above (Table 8). For the minority contributor I found two options: a Bronze Age Armenian-like like source and a Bronze Age Levantine-like source (Table 8). With the Armenian scenario the admixture proportions are 4-8% Armenia BA and 92-96% native Roman Iberian. In the Levantine scenario, the admixture coefficients range between 10-26% for the Broze Age Levant source and 74-90% for the Roman Iberian-like source (Table 8). To simplify the scenario, perhaps it is more reasonable to talk about a broad West Asian contribution of around 15% to Dafne/GOG50 and the rest of the 85% of the ancestry deriving from the local Iberian population of the Roman period. This Asian-like contribution to her ancestry may also help to understand why she was a carrier of an eminently East Asian mitochondrial lineage (D4e1) (Table 7), very rare today in Europe but common in Central, East and South Asia. However, following the publication of the data of the temporal transect of Rome by Antonio et al. (2019), Dafne is not quite so alone any more. Like in Olalde et al. (2019), some of the Roman ancient genomes from the Imperial period also appear genomically similar to Dafne and there is even one who carries a mitochondrial haplogroup of type D4j11 (another female, sample R78 from Necropoli Salaria dated between 0-200 CE). This MuSaK result reinforces one of the fundamental conclusions in Antonio et al. (2019), which states that substantial gene flow from the Levant to Rome happened since Roman conquest of Asia Minor and the Near East by the end of the Republic and the beginning of the Principate. I can further add that this gene flow went into Rome and beyond, to other regions of the Empire.

MuSaK ancestry modelling: admixed Islamic samples

Moving on to the modelling of ancestries of the Islamic samples, as shown in Figure 51 samples GOG24, GOG25 and GOG26 are the ones closer to the European cluster than to the North African cline whereas for GOG23 (carrier of E1b paternal lineage) the situation was the opposite, individual GOG20 followed a perpendicular trend compared to these previous ones. If we agree that the Guanches are the best proxy available for pre-Islamic North Africa and that post-Roman Iberia is the best for the Spanish side, then we obtain the results that individuals GOG24, GOG25

and GOG26 carry an average 11% of native North African (Iberomaurusian) ancestry (Table 8). This is very much in agreement with what has been discussed in sections above. For the case of GOG23, we observe that the proportion of Guanche contribution, as suspected, indeed comes up as much higher, at 57% if we choose post-Roman Iberia as the other contributor (with 43%) (Table 8). The distinctness of GOG20 (Figure 51)is also reflected in this analysis because a substantial Near Eastern contribution to her genome, in the form of Levant BA, is revealed. But I faced a problem at this point: what Iberian population is the best proxy? Because depending on whether we use a post-Roman or Iberian Islamic source we get almost 40% or just 25% respectively. The RSS values indicate that the fit is much better for three-quarters Iberian Islamic and one-quarter Levantine contributions. This is also easier to reconcile with the scenario of GOG20 having one grandparent originally from the Near East and the others being admixed local Islamic Iberian-Maghrebis. Although speculation, this would provide a simpler explanation than other fractions, such as the 60-40% case using post-Roman Iberia as parental (Table 8).

4.4.5 Runs of Homozygosity

For the four Chromosomes 22 imputed for samples GOG11 (Chalcolithic), GOG50 (Roman), GOG23 (Islamic) and GOG26 (Islamic) the length of each homozygous stretch (RoH) in the chromosome was measured. The RoH approach serves to inform about past demography. The information was plotted as a cumulative function of the sum of the length of each RoH in ascending order but without creating discrete RoH length categories as it is usually done. This method of visualization allows to see what curves/individuals have shorter RoH composing their chromosome (curve saturates faster) and which ones indicate longer RoH fragments in the chromosome (curves that saturate slower). The faster the curves saturate to the maximum length possible, the faster short RoH stretches are being added to the cumulative total. If more short RoH fragments are found, it implies there will be fewer long fragments of RoH in the genome. This is a useful indication of inbreeding, bottlenecks, etc. Genomes, individuals or populations with lower diversity will display RoH curves that take longer to reach the plateau of the curve because they have longer RoH which are only added after the short ones.

Figure 55A shows the RoH length cumulative curves (using transitions and transversions) for the imputed samples and a series of modern populations from the 1000 Genomes Project panel for comparison. Yoruba were selected as a reference of upper boundary as a population with expected short RoH fragments. Peruvians were chosen as a lower boundary because it is known that Native Americans derive from a severe bottleneck and therefore have the longer RoH fragments (Cassidy et al., 2016; de Manuel et al., 2020). Spaniards, Italians, Britons and Finns are also included for a more proximal comparison with the four ancient Iberians. On the right (Figure 55B) is the same plot but only displaying the four ancient imputed chromosomes and distinguishing the two curves made using all mutations (Ts+Tv) and transversions only (Tv).

Among the ancient imputed genomes, the sample from a collective burial in a Chalcolithic cave (GOG11) is the one that has more longer homozygous runs in chromosome 22 (it can be extrapolated to the whole genome). Compared to the modern European populations it also has more longer fragments but it is comparable to the Finnish curve and not too distant from the British, Spanish and Italian. This pattern of slightly higher RoH in Neolithic-Chalcolithic, but nonetheless comparable to modern Europeans has already been shown in Cassidy et al. (2016).

The two Islamic samples and the Roman individual display lower RoH patterns, intermediate between Yoruba and the modern European references included here Figure 55A. This is not unexpected since at least for the two Islamic samples I already knew they were admixed with North African sources, especially GOG23 who is heavily admixed.

The Late Roman female individual GOG50 also shows a pattern of shirt RoH fragments, intermediate between European and Yoruba. This still holds when considering the transversions only curve, and what is more, the RoH pattern of GOG50 is very similar to high-Maghrebi ancestry GOG23. This is interesting because GOG50 appears to have a generic Mediterranean ancestry

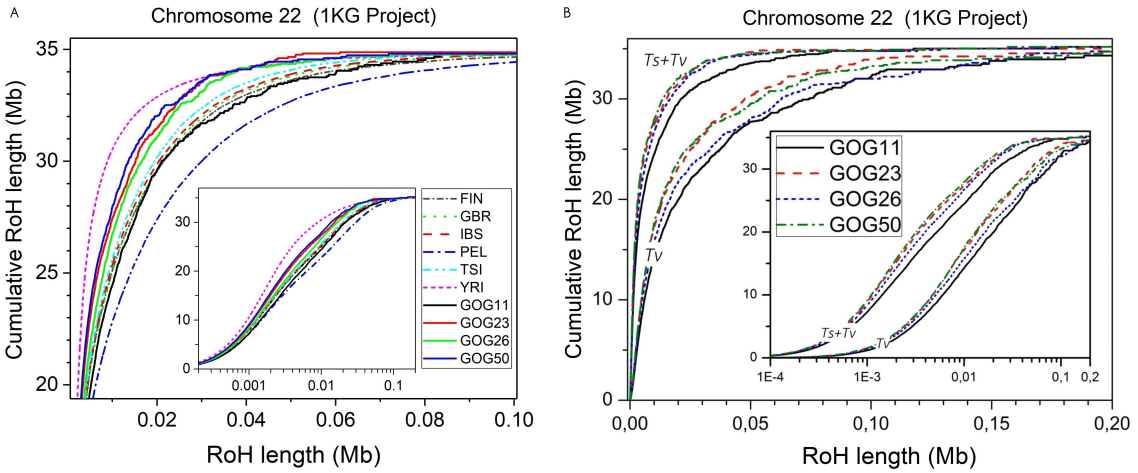


Figure 55: Runs of Homozygosity analysis performed after imputing Chromosome 22 in the four ancient GOG samples with overall genomic coverage >0.5X. It includes 1000 Genomes Project data from modern populations (FIN: Finnish, GBR: British, IBS: Spanish, PEL: Peruvian, TSI: Italian, YRI: Yoruba) for comparison (A), and the same plot but only including transversions (B).

sitting on a no man's land position on the PC2 axis, with Sicilians and Spanish as the closest groups on each side. This position in the PCA could suggest admixture between Spanish and Sicilian sources. However, the RoH pattern similar to GOG23 suggests that GOG50 is a mixture of two groups or populations more divergent than Iberians and Italians or Sicilians. The marginal fraction of Iberomaurusian ancestry in GOG50 ADMIXTURE profile suggests a scenario where one of the sources already carried some North African ancestry, the Roman Iberian population of the time is a perfect candidate in that regard as we have explored above. An ancient population from the western Mediterranean would be sufficiently differentiated genetically to explain the RoH pattern observed in GOG50, and thus is a good fit to fill the role of the second parental source. Taken all together, this result is further evidence of mixture between Western and Eastern Mediterranean sources to form GOG50, although not in equal proportions as already seen above. Haplotype sharing analysis will be required in the future to further break down and confirm what PCA, ADMIXTURE and RoH are telling us about the ancestry of GOG50.

4.4.6 Formal Tests with *f*-Statistics & *D*-Statistics

I performed an outgroup-$f3$ test to measure the amount of shared drift for the group of samples that conform the Valencian Late Neolithic and Chalcolithic population (Figure 56). I included several Neolithic populations from across Europe and the Fertile Crescent, and some other later populations.

Not surprisingly, the populations that shared the most drift are the Iberian Early and Mid-Late Neolithic together with the Iberian Chalcolithic population. However, several other ancient Neolithic populations of the Mediterranean and the Atlantic regions (Serbia, Italy, Scotland and Ireland) also share high levels of drift with the Valencian LN/CA group. The similarity between the Spanish and Atlantic Irish Neolithic has already been noted in previous works (Cassidy et al., 2016). The Italian affinity is also expected due to proximity and since it is largely assumed that the sea route of Neolithic expansion came to Iberia through the Italian peninsula. Hungarian and German Neolithic populations appear later but with similar levels of shared drift, confirming the genetic homogeneity of the land and sea routes of Neolithic expansion. The human origin of these two routes are the Anatolian Neolithic farmers (ANF). In Figure 56 they appear behind of all the Neolithic groups from Mediterranean, Atlantic and Central Europe. This makes sense since these later Neolithic groups are derived from the ANF that are more distant chronologically.

What is interesting is that the Italian Mesolithic population seems to have more affinity with the Valencian LN/CA than the Iberian Mesolithic. This is a trend I also observed by looking at the PCA (see Figures S12 and S23), the Iberian Neolithic cluster appears to have a better alignment in directing with Italian hunter-gatherers than with Iberian hunter-gatherers, which opens up the possibility that Iberian WHG actually contributed very little to later populations (Figure 56).

I have already presented evidence that supports the Late Roman sample GOG50 (also named Dafne) being admixed from two different Mediterranean sources. However, I nevertheless performed outgroup-$f3$ tests, available in the supplementary material (Figure S24), with modern and ancient populations (an admixture-$f3$ test is impossible to perform with AdmixTools since it is a single sample). I also performed a number of *D*-statistic tests with modern and ancient populations, and several $f4$-ratio combinations to calculate admixture coefficients and to compare with the MuSaK results.

Both the outgroup-$f3$ and the *D* tests with modern populations show that Dafne displays the most affinity towards other Mediterrenean groups like Sardinians, Spaniards, Basques and Italians (Figure 57). An important detail is that she appears equally related to North Italians and Spanish, note that *D* value is greater than zero with North Italians but error overlaps with zero.

The outgroup-$f3$ tests (see SM) provide basically the same information as Figure 57. GOG50 shares some of the highest drifts with two ancient Sardinian individuals from the Roman Imperial period (Antonio et al., 2019) and another one found in the Crusader Pit in an archaeological excavation in the Near East (Haber et al., 2019). However, a puzzling results from the outgroup-$f3$ test is that ancient Neolithic populations (from Anatolia, Serbia, Iberia, among others) appear as consistently sharing high levels of drift despite being separated in time by thousands of years (see SM). I suspect this is an artefact created by convergence to similar allele frequencies through different routes. It also shares high drift with modern groups with high Neolithic-related ancestry (Sardinians, Basques), and I would be prone to accept a full Sardinian ancestry if it was not for the RoH profile which suggests admixture from two differentiated sources; Sardinians are too isolated and inbred (more so than average mainland Europeans) to reconcile with the RoH results.

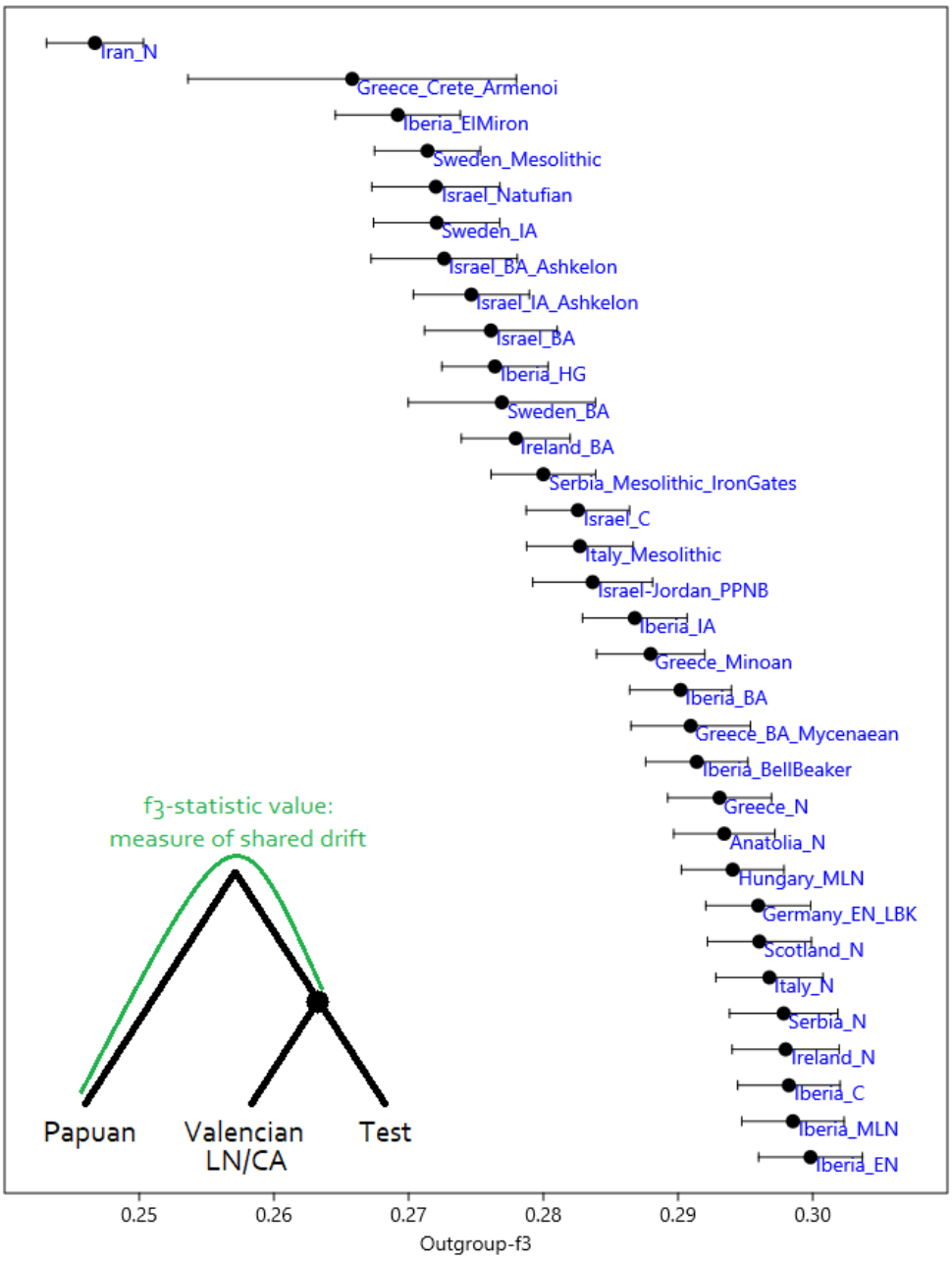


Figure 56: Outgroup-$f3$ statistics for prehistoric Iberian samples

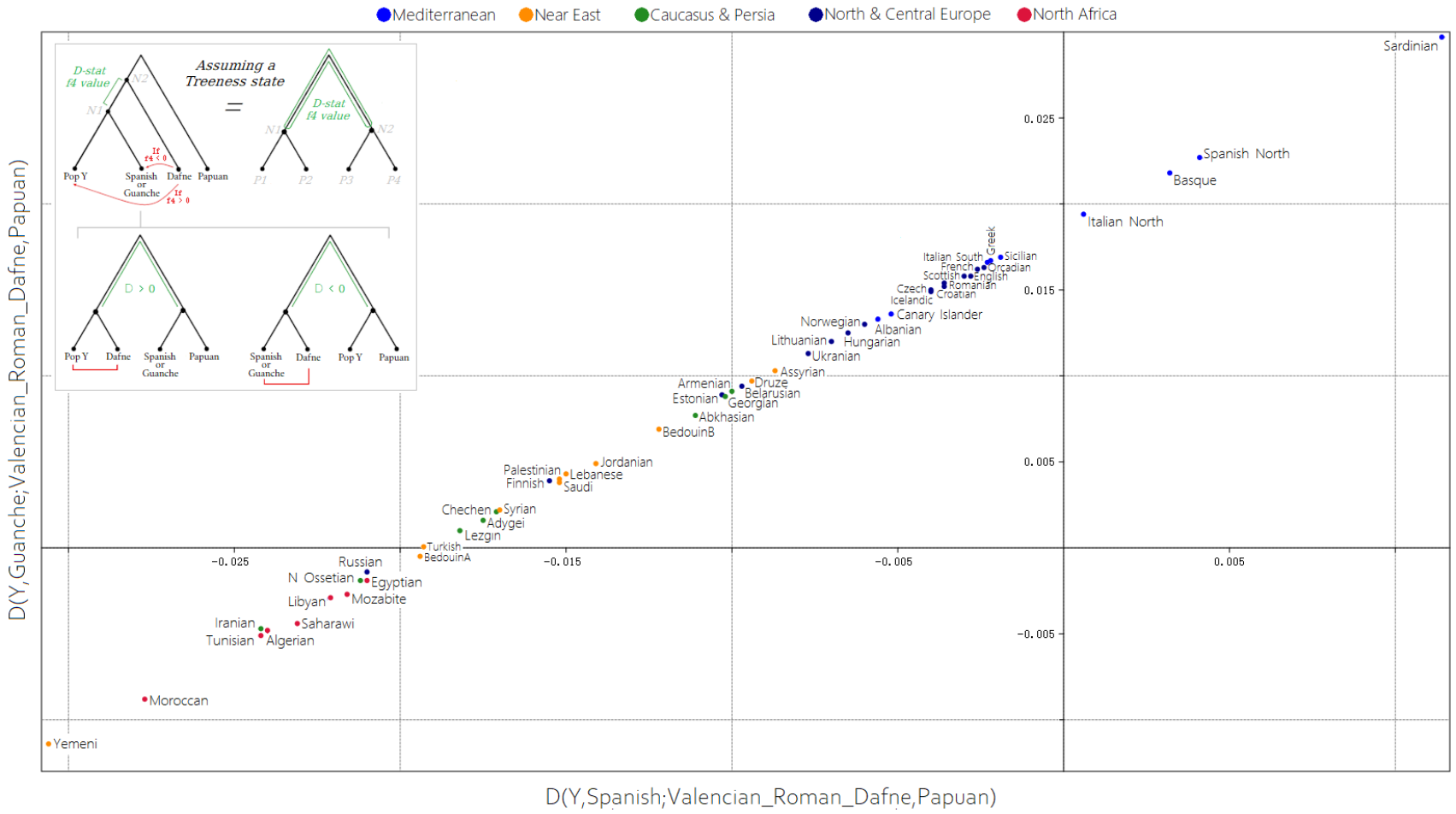


Figure 57: Crossed D values to condense information of two tests for Late Roman sample GOG50 (Dafne). This should not be confused with an $f4$-ratio test because the slope does not indicate an alpha coefficient value contributing to Dafne. The numbers used correspond with D values not with $f4$ values.

Because GOG50 must be an admixed individual, $f3$ tests are not the right approach to investigate her ancestry. For the purpose of further looking into the possible contributing sources I performed $f4$ test using ten different combinations of outgroups (*f4(Outgroup1, Outgroup2; Dafne, Test)/f4(Outgroup1, Outgroup2; Alpha, Test)*). I did so to reduce the noise and spurious effects introduced by the use of particular outgroups.

I found that the outgroups had to be genetically distant between each other as well as distant from the test populations in order for the test to work but in a capricious way that is yet to be explained by the developers of the AdmixTools software. For some reason, even when I used two African outgroups that are supposed to be differentiated, for example Mbuti and Ju'Hoan, the software was never able to produce realistic results even though it should have. The nature and use of what can be considered an *outgroup* or more recently rebranded as *reference population* in Harney et al. (2020), as well as what is a good set of them in terms of number remains to be properly explained.

Based on previous results I selected various modern populations (Basques, Sardinians and Spanish), and some ancient populations (Roman Iberia and Post-Roman Iberia) as proxies for the main ancestry contributors (Alpha value). As minor contributors (Test) I selected many modern populations from the Caucasus, Near East and North Africa (although based on the PCA I only regard Egyptians as a plausible source). I repeated the $f4$ ratio analysis but grouping the modern *Test* population into meta-population groupings (see Supplement II, Table S4).

As shown in Figure 58, the results varied depending on the assumptions made to decide which are the best and most likely sources. The genetic diversity in the Mediterranean during Roman times experienced mixing and homogenization to a degree that the boundaries that we can distinguish in a PCA with modern populations become blurry, and this introduces uncertainty and an arbitrary element when deciding what is the best proxy for ancestry sources. However based on previous results (PCA, ADMIXTURE and MuSaK) I had some hints that either Armenian-BA-like ancestry or Levantine-BA-like ancestry could be potential sources. Although North Africa also appears as mathematically plausible in Figure 63, I do not think a Maghrebi population was directly involved in the making of Dafne. In fact, the error bars fall outside realistic proportions. In light of the $f4$-ratio results (Figure 58) I can confirm previous the MuSaK results from the above section and say that the genome of Dafne is made up of 70-90% of a West Mediterranean source (Ibero-Sard-Italic cluster) and a 10-30% West Asian ancestry (ranging from the North Caucasus and Asia Minor to the Levant). Again, this would help to explain her Asiatic maternal lineage.

For the Medieval period, I selected two unmixed samples excavated in the same Christian cemetery and separated by no more than a century. These two samples are the slave (GOG59) and sample GOG60. I plotted the outgroup-$f3$ values of a subset of modern and ancient populations in the same figure to have a direct visual comparison in Figure 59. The difference between the two samples is basically that GOG59 (of Berber extraction) always shares more drift with every modern and ancient North African group when compared with GOG60, which genetically resembles modern Spaniards. It is a sharp contrast for two samples that shared the same burial ground in the city of Valencia only 400 years ago. This is evidence that not only ancient societies, but also post-medieval ones, were more diverse than popular belief would suggest. The reason for this, perhaps, is that modern European nation-states have had an interest in burying this part of their history. The recent history of institutionalized slavery happened not only in the New World; sometimes it is buried in our own backyards.

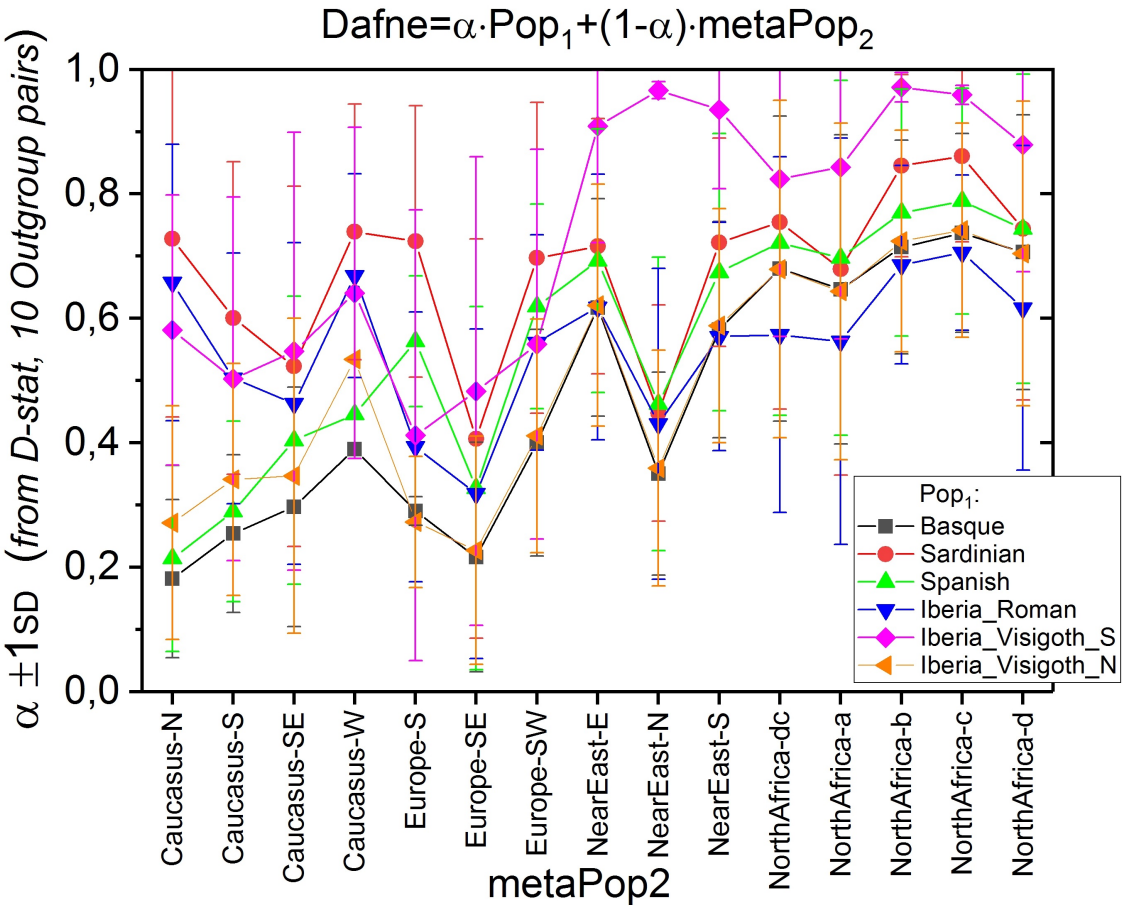


Figure 58: Series of $f4$-ratios calculated for Dafne assuming various primary donors: Basque, Sardinian, Spanish, Iberia Roman and Iberia Visigothic South and North (alpha). Multiple populations used as secondary donor (1-alpha) were grouped into meta-populations according to their region of origin (Caucasus, Near East and North Africa).

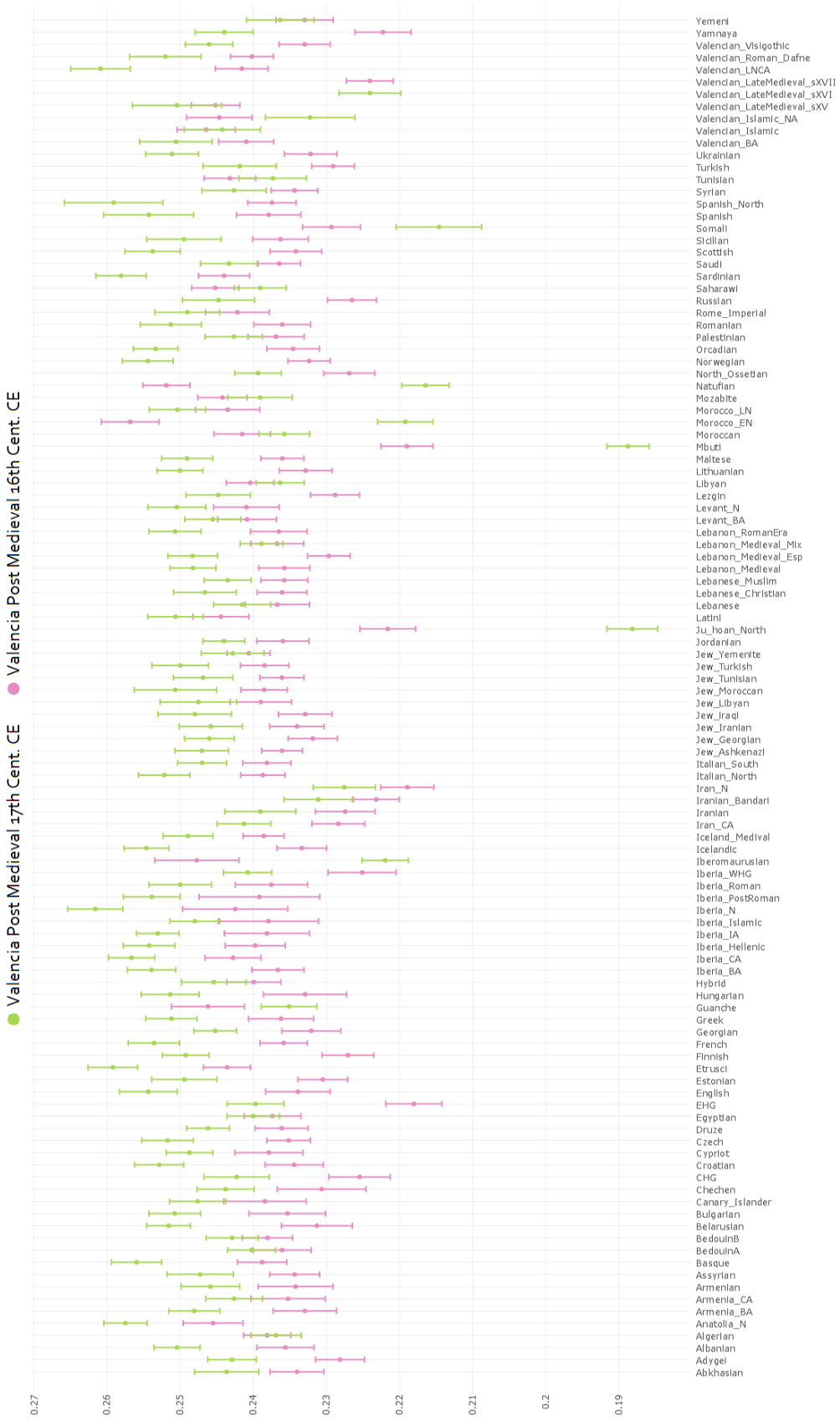


Figure 59: Outgroup-$f3$ results for individuals from the 16th century CE (GOG59) and another from the 17th century CE (GOG60) from the same Christian cemetery.

4.4.7 Dating the North African Admixture Event in Iberia

In comparison to Europe, the genetic past of North Africa is understudied in terms of numbers of samples. There are only a handful of early and late Neolithic genomes from Morocco available and some extremely valuable Moroccan hunter-gatherers from the Iberomaurusian culture which have proved so central for these thesis. However, there are some good datasets with modern populations from Henn et al. (2012) and Arauna et al. (2017) which might theoretically be good proxies for modelling scenarios with DATES.

Combination	*Target*	*Generations*	*Years*
Basque+Guanche	Islamic	143±52	4144±1511
Basque+Moroccan	Islamic	30±5.5	879±160
Basque+Mozabite	Islamic	25±5.7	718±166
Spanish+Guanche	Islamic	130±55	3761±1584
Spanish+Moroccan	Islamic	27±6.7	785±194
Spanish+Mozabite	Islamic	23±7.8	658±225
Basque+Guanche	Late Medieval	Fail	Fail
Basque+Moroccan	Late Medieval	Fail	Fail
Basque+Mozabite	Late Medieval	13±11	382±320
Spanish+Guanche	Late Medieval	Fail	Fail
Spanish+Moroccan	Late Medieval	Fail	Fail
Spanish+Mozabite	Late Medieval	Fail	Fail

Table 9: DATES results for time of admixture event that formed the Islamic (n=6) and Late Medieval (n=2) Valencian population using GOG samples as target in different combinations.

Since I observed clear signals of Maghrebi admixture in the Islamic Valencian individuals, something that is also visible in the Islamic Iberians from Olalde et al. (2019), I decided to use DATES to estimate the time of admixture in my Islamic genomes (Narasimhan et al., 2019). However, I found that the outputs provided by DATES are limited by the input. Ideally several individuals should make up the parental and target populations because results are not reliable when using single samples. I also noticed that mixing ancient and modern (acting as proxy) populations tend to yield ages that do not make sense or fail. An example for this instance is the combination of Basque and Spanish (modern) + Guanche (ancient) in Table 9, the ages obtained are way to old to fit any migration or archaeological scenario.

The best performing scenarios turned out to be when using two modern populations acting as proxy for North Africa and Iberia (Basque, Spanish, Mozabite and Moroccan), with ancient Islamic (n=6) and Late Medieval (n=2) samples as the targets of the admixture (Figure 60). When using Basque+Moroccan the time of the admixture that formed the Valencian Islamic population is estimated to have happened some 879 years before the age of the samples (11th-13th Centuries of the Common Era); with Basque+Mozabite it is 718 years before 11th Century CE; with Spanish+Moroccan it is 785 years prior to the 11th century, and with Spanish+Mozabite the admixture is estimated 658 years earlier. The overall range indicated that the admixture (likely a product of continuous genetic influx from North Africa) happened sometime between 430 years and 1040 years before the time of the samples around the 11th and 13th Centuries CE. However, the mean value is around 750 years which indicates the admixture happened at some point between the 3rd and 5th century CE. This makes sense looking at the Roman and Post-Roman cluster in the PCA, and in light of the Iberomaurusian-related ancestry in the ADMIXTURE profiles of pre-Islamic samples GOG50, GOG34 and GOG35.

Reality must have been closer to a scenario where the increased mobility during the times of the Roman Empire favoured the introduction of North African ancestry into Hispania in the last

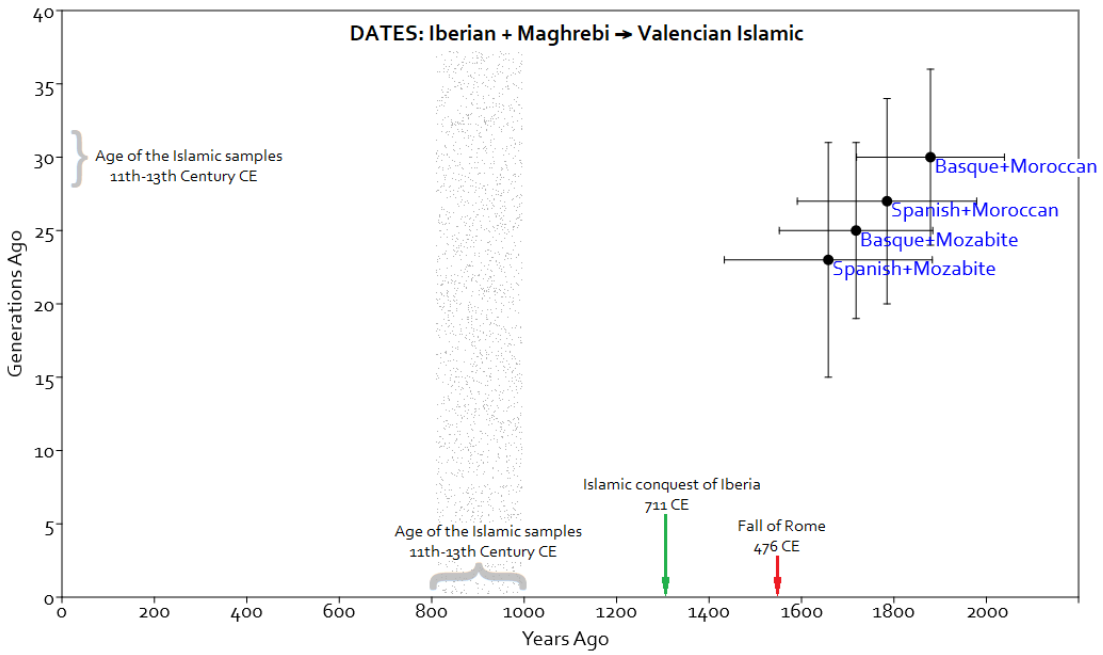


Figure 60: Timeline of the dating of the admixture of Islamic Iberian GOG samples obtained with DATES.

centuries of the Late Empire (3rd-5th centuries), especially in Mediterranean coastal provinces. This is probably a case of a constant stream of genetic admixture over centuries, at least to the Mediterranean lands of Spain. The phenomenon seems to predate the Islamic conquest in 711 CE by a couple of centuries, so when the Islamic newcomers arrived in the Valencian region they probably encountered a population that was already closer than suspected genetically to native North Africans of the time.

When I replaced the Islamic samples as target with the two Late Medieval individuals the output results were a complete failure or with unacceptable error margins (Table 9). I have not been able to find or provide a satisfactory answer to this outcome which should be further explored in the future. Perhaps, using only two individuals is not enough because causes distortions in the calculations performed by DATES.

4.5 Conclusions

The upheaval that made Europe

The sequencing of ancient genomes in the work of Haak et al. (2015); Allentoft et al. (2015) settled, in part at least, one big archaeological debate: the source of the Indo-European languages in Bronze Age Europe and Asia. In the last five years a series of works Haak et al. (2015); Furtwängler et al. (2020); Linderholm et al. (2020); Schroeder et al. (2019) have confirmed that the Neolithic to Metal Ages transition is characterized by a strongly male-mediated shift towards what we call steppe ancestry. Archaeologist Marija Gimbutas posited the Kurgan hypothesis, which argued that steppe migrants brought Indo-European languages to Europe. She further argued that the steppe pastoralists from the Yamnaya culture, hierarchical and aggressive, replaced the more peaceful matriarchal Balkans Neolithic tradition she called Old Europe (Mathieson et al., 2018).

In the last stages of the Neolithic, sometime around 6000 to 5000 years ago, the genetic make-up of the inhabitants of Europe and Iberia was a composite of European hunter-gatherer and Anatolian/Levantine early farmer ancestry. However, the Yamnaya, a herding culture from the Pontic-Caspian steppe that connects Europe and Central Asia who arose in the centuries before 5000 years ago, derived their ancestry from a different mix of eastern European hunter-gatherers with a distinct Neolithic ancestry from the Caucasus and Iranian plateau region. Technological innovations (wheels, carts, warfare) and horse domestication developed by the Yamnaya allowed bands of young male individuals to move further into Europe, presumably for cattle raiding, trade, and also establishing client networks and patriarchal structure (Anthony, 2010). Around 4500 years ago, the Yamnaya presence in Europe had already crystallized by contributing the major part of the ancestry of the Late Neolithic Corded Ware people (Kristiansen et al., 2017; Linderholm et al., 2020), changing the continent and its culture forever with the introduction of Indo-European customs (Haak et al., 2015; Immel et al., 2020). Yamnaya ancestry is nowadays widespread in present-day Europeans, with an approximate NE-SW gradient visible in the Prehistoric PCA (Figure 50).

There is one issue however, where ancient DNA evidence does not agree with Gimbutas because matrilocality does not appear to be the rule in Neolithic Central Europe at least. Reconstructed kinship networks with genetic and stable isotope data of family structures from the Central European Neolithic and Bronze Age appear to remain unchanged after the arrival of Yamnaya ancestry. In very recent years we have started to unveil the first evidence of how patrilocality seems to have been the predominant social structure in some prehistoric European societies of central Europe (males stay where they are born, females come from other groups) (Mittnik et al., 2018; Schroeder et al., 2019; Linderholm et al., 2020; Furtwängler et al., 2020).

The genomes sequenced in Olalde et al. (2018, 2019), have offered clarification about how this phenomenon translated from Central Europe to Iberia. The Yamnaya-like or steppe ancestry arrived in Iberia shortly before the 2000 BCE horizon, quite possibly through the same trade networks that helped the Bell Beakers spread through much of continental Europe. The genetic impact of the migrants that introduced this novel ancestry in Iberia was less, however it is not entirely clear the reason behind this milder impact in the south.

The results I have obtained here confirm that the Valencian Chalcolithic population was little different in terms of population from the preceding Neolithic one, and that this Late Neolithic and Copper Age population was smaller but only slightly more inbred than modern Europeans. Combining genetic data and radiocarbon dates of some of my samples, I also draw the conclusion that by 4000 years ago the steppe ancestry had arrived in the Mediterranean region of Spain. Nonetheless, although this coincides with published findings for other parts of Iberia, note that this conclusion is based on the sole Bronze Age sample successfully sequenced in this project since the other two failed the screening.

I have also observed that, based on *D*-tests and visual linear combinations of the two-dimensional PCA, Western Hunter-Gatherer genomes from North Spain (La Braña 1 and Chan do Lindeiro) are

not a good fit to have formed the Iberian Neolithic populations. More likely, the Neolithic farmers that arrived in Iberia already carried WHG ancestry from admixture events that happened with Mesolithic populations such as Italian WHG. Alternatively, WHG could have had more population structure than what we currently observe and perhaps no appropriate Iberian candidates have been sequenced to date. However, the door is open to the possibility that Iberian WHG contributed very little and were poorly integrated into the resulting Middle and Late Neolithic population. To try to explain this phenomenon, an intuitive explanation could be that the Neolithic migrants arriving to Iberia had much more experience in colonizing new land and exporting their improved agricultural culture than their predecessors that left Anatolia and colonized the Balkans and Italy centuries to millennia before. In this regards the Neolithic colonizers of Iberia could have been a sort of professional pioneers of the time that had to rely much less on interactions with local hunter-gatherer populations. Furthermore, in my results I do not appreciate significant increase of WHG ancestry between Early and Late Neolithic individuals. The explanation for this observed difference has to be found in different genetic clusters of Upper Palaeolithic and Mesolithic individuals. The published Iberian Mesolithic WHG listed above, as well as El Miron, are closer to the Goyet-Aurignacian/Magdalenian cluster (Villalba-Mouco et al., 2019) than other Mesolithic WHG from Europe and Italy, who are more similar to the Villabruna cluster (Villalba-Mouco et al., 2019; Catalano et al., 2020; Rivollat et al., 2020; van de Loosdrecht et al., 2020). The Villabruna type seems to have been the ancestry dominant in Italy and hinterland of Europe and contributed more to later Neolithic farmers than the mixed Magdalenian-Villabruna Iberian hunter-gatherers. However, this result might be biased since most Iberian WHG are from the north of Spain and not from the southern and eastern Mediterranean fringe.

In summary, there is general agreement that the arrival of the Yamnaya from the Eurasian steppe created the European genetic pool as we understand it nowadays. However, as I have mentioned in the results and discussion section, in Iberia later shifts in the proportions of the three types of ancestries were yet to happen. For example, the steppe ancestry increases greatly during the Iron Age, and it is yet to be confirmed how this happened. Very likely, the arrival of the Urnfield culture by the Late Bronze Age and Early Iron Age played a role, or maybe other later Celtic-related movements of people from Europe had an influence as well. Whatever the case, and leaving the Iron Age period aside, the Bronze Age migrants from the steppe were not the last massive genetic transformation of Mediterranean Iberia because the future Roman and the Islamic conquerors.

The Roman Empire and the establishment of the pan-Mediterranean genome

Until very recently, despite being a period very well studied from the point of view of historiography, the Roman era of domination over the Mediterranean had not been approached by ancient DNA studies. Martiniano et al. (2016) sequenced the first Roman individuals from an excavation in York (England) proving the high mobility of the period by detecting a Near Eastern individual, but nonetheless all the other individuals were similar to Bronze and Iron Age Britons. On the other side across the Mediterranean sea, the Roman samples from Lebanon in Haber et al. (2019) also show little change and genetic continuity with previous Levantine Iron Age populations (Haber et al., 2020; Skourtanioti et al., 2020; Agranat-Tamir et al., 2020). It has been the Roman genomes from the Imperial period in Olalde et al. (2019) and Antonio et al. (2019) that have highlighted the immense genetic heterogeneity in the Central and Western Mediterranean possibly reflecting the high mobility of the time. Although it might seem contradictory, the observed genetic heterogeneity within particular lands (York, Spain, Rome) is the result of a process of homogenization by admixture across the West, East and South Mediterranean that led to the establishment of an early stage of what could be called a pan-Mediterranean (meta)genome.

It is in these Late Imperial times that sample GOG50 (Dafne) is framed chronologically (3rd-4th century CE). This is based on the archaeological dating, which is very reliable in this case, but no radiocarbon dates are available. Dafne was excavated in the lower layers of the Christian

Medieval cemetery of San Lorenzo (Saint Lawrence) in Valencia (Olmos and Marcos, 2000). This cemetery is also where four of the post-Islamic samples sequenced here were buried centuries later and excavated in the last decades. The first thing learnt about Dafne was that her mitochondrial lineage was D4e1. This is not a common European haplogroup but instead Asiatic. Our lab's worldwide mitochondrial database indicates that the only modern carriers of this haplogroup in Europe were an Austrian and a Bohemian individual (plus an unpublished Cypriot sample generated by a colleague in the group). This mtDNA finding initially lead me to believe, taking into account the dating too, that the sample could be Gothic-related since it seems plausible that the Germanic invasions could have introduced it. However, the position of Dafne in the PCA suggests that she was not a Germanic individual but clearly of some kind of Mediterranean background. The suspicion that she could be a mixed genome was immediately raised, but it was yet unclear what kind of mixture she could be since she also has a small but non-neglectable African-related ancestry fraction. Perhaps a mix of Iberian and Sicilian or Greek was my first intuition (similar to the individuals labelled as Greek/Hellenic in Olalde et al. (2019)).

The history and urban archaeology of the city of Valencia is fairly well documented, and a manuscript from 1979 (Pereira Menaut, 1979) which contained a compilation of all the Roman inscriptions ever found around the perimeter of Roman Valentia offered some interesting clues. I found out that a lot of Imperial, especially 3rd century, inscriptions in Valentia display Greek names. A few examples of these Greek named are Hyginus, Nysus, Tyche-Eutyches-Eutychia (variations of these were very common among slaves and freedmen), Atimetus, Nymphe, Glyce, Antitheus, Onesicratia, Coimothoe, Meliae, Protis, Zoe, Alypion and for some also their occupation is known, like the case of Apolaustus the silvermonger. There is even an inscription with Greek names of three family members. The inscription goes *"Philete annorum XVII. Hic sita est. Apollonius et Helene cognatae de suo fecerunt."* Which roughly translates as *Philete of 17 years of age. Here he rests. Apollonius and Helene, his relatives, made this*. Of course a Greek name is not definite proof of direct Greek origin but these inscriptions point towards an important body of inhabitants/citizens of Greek or at least eastern Mediterranean origin. Actually, a West Asian source such as the Caucasus, Asia Minor and/or the Northern Levant fits better the minor contributor role in models obtained in the results (I assume the primary one is West Mediterranean). Judging by the inscriptions available this broad Asiatic or eastern Mediterranean ancestry could have already been established in the Valentia of the 3rd century CE. Sadly, there is still a lack of data from the Roman period in Greece to have a direct confirmation. All we know at the moment is that Bronze Age Mycenaean Greeks were genetically different from modern Greeks, and that a modern Greek-like source according to their position in a PCA for example, is not as good proxy for a secondary source in Dafne's genome.

Since I do not have a large sample size such as we see in many recent ancient DNA publications (Olalde et al., 2019; Antonio et al., 2019), I feel compelled to try to explain how an individual with such a genetic background ended up at the far end of the West Mediterranean in an obscure and perhaps decaying Roman colony. After some discussions with lab colleagues I have developed a speculative scenario that I call the Black Sea hypothesis. The basic idea is that the Asiatic D4e1 lineage was picked up by one of the maternal ancestors of Dafne around the Black Sea coastal regions of Caucasus or Anatolia where the ancient Greek colonies were established at some point in the past, but at least sufficient time for any trace of the Asian ancestry to have disappeared. Then, during the Roman Empire period, another ancestor of Dafne migrated along the Mediterranean, as a slave or otherwise, and arrived in Iberia. Nonetheless, with the current data, I find myself unable to decide, without adding further speculation, whether a Sardinian-like, a Roman Iberian or Post-Roman Iberian source is the most adequate for the primary source. I believe haplotype sharing analysis would help to shed light on the issue. Time constraints and computational limitations have prevented me from achieving this goal during the completion of my PhD, but I intend to do so in the future.

Both post-Roman samples dating to the time of the war between Visigoths and Byzantines in

the Iberian Peninsula of the 6th and 7th centuries are another example of the pan-Mediterranean homogenization. However, in this case, they are a better example of another type of ancestry, the North African one, that permeated Roman Hispania under the umbrella of a cosmopolitan Empire. A reason behind the spread of this ancestry through Roman territories, besides increased mobility, could be the destruction, of Carthage a century before the end of the Republic in 146 BCE, and the enslavement of thousands of its inhabitants who were transported back to Italy, and also the dispersal of the survivors. It is the adult male individual GOG34 (the father) who looks genetically very similar, if not identical, to the later Islamic population in the same Valencian region. This is interesting at many levels, genetically and linguistically; to start with because, if the general post-Roman population in the Valencian region was anything like GOG34, it would mean that the North African newcomers that arrived with the Islamic conquest of Spain a couple of centuries later were not too different genetically from the natives and perhaps even less so in the language they spoke. This might also help us to understand the speed of the conquest. Quite possibly, the most striking example of how deep the North African connection with Iberia was by the end of the Roman Empire and beyond, was an unlikely discovery in the Catalan Pyrenees. The discovery was the burial of a macaque, dated by radiocarbon to the 6th century, found while excavating the old Roman town of Iulia Livica or castrum Lybiae (Llivia, Spain). The macaque of Llivia is extremely well preserved, allowing researchers to notice that it was buried with care and wrapped with military paraphernalia, as well as having had a good diet during his life. The context indicates that it was a military mascot of a Visigothic army (Guardia et al., 2007). Archaeological evidence of macaques as private military mascots is rare but not unheard-of in the earlier Roman period (Poitiers, Pompeii, Wroxester, Catterick, Dunstable, Rainau-Buch) but this one provides evidence of contacts, even during the Migration Period, between Iberia and the Maghreb: a surprisingly high degree of mobility after the imperial era, which even allowed movement of apes from the Atlas mountains to the Pyrenees.

A genetic bridge linked Iberia and North Africa for over a millennium

The sequencing of Valencian Medieval genomes dated to between the 11th and 17th centuries CE points towards a pattern that is best described as a genetic bridge between the coasts of North-western Africa and the Iberian Levant. Every single one of the Islamic, Late Medieval and post-Medieval samples fall along a gradient of Maghrebi admixture. In this gradient, regardless of whether they are Muslim or Christian; rural or urban, we can find representatives for almost any combination. There are individuals, the majority, with an average of 10% Berber-related ancestry in Muslim rural cemeteries as well as in urban Christian post-Islamic cities. Another individual appears to be a 50-50% mix, yet both his uniparental markers are of clear Amazigh or Berber origin, which hints at the possibility that he was product of an indigenous Valencian population heavily admixed with Maghrebi ancestry rather than being a first generation Ibero-Marghrebi mixture himself. Another is 75% North African. On top of that, there are two more individuals separated by no more than one hundred years (16th and 17th centuries), buried in the cemetery but the younger one for the first time shows no evidence of North African admixture in his genome, while the other one is a Berber slave buried in Christian tradition.

Remember that the Medieval Iberian world of both Muslim Al-Andalus and the Christian kingdoms was defined by a social structure into which people fitted depending on the cultural background and circumstances. It is worth noting that some of these concepts were only developed *a posteriori*. Moors, in Valencia more commonly known as Saracens, were the terms used to defined the Muslim *others*, encompassing Maghrebis, Syrians and Arabs, although technically the first one should only refer to North Africans and the second one only to Arabs. In later times, Berbers and Ottoman Turks were also confused under the term Barbary pirates. Nevertheless, as we have discussed already, the resulting population of Al-Andalus by the 11th century in the Mediterranean region of Valencia was a mixture of native and imported ancestries. Remember that at the zenith of Islamic territorial domination other terms were applied to describe indigenous

people who converted to Islam (*Muladi*), and Christians living in Muslim territory (*Mozarab*). Alternatively, in later centuries when the tides had changed and most of the territory was under Christian control, new terms with derogatory connotations appeared to describe Muslims living in Christian rule (*Mudejar*) and Muslims that converted to Christianity (*Morisco*).

The timeline of the rural Islamic samples situates us between the 11th and 13th centuries CE. These centuries came just before the conquest of James I, when the territory was under Islamic rule, so since the genetic cluster formed by the samples from Vall d'Uixo clearly comes from burials in the Islamic tradition we can identify this group under the Muladi category. Genetically they are 90% Iberian but their religion was Islam. They tick all the boxes because interestingly, the original meaning of the term *Muladi* in Arabic (*muwallad*) was a person whose ancestry was a mix of Arab (typically a Muslim father) and non-Arab (typically a non-Muslim mother) parents and educated in Islamic society. This would reconcile the etymology with the North African introgression and hints that religious transition was driven also by social mixing, favoured by the fact that both groups still shared a form of Latin as a common tongue. However, that meaning of the word *Muladi* does not survive any more in the modern Spanish language.

Moving forward along the timeline, there are two urban samples from Valencia city (14th-15th century CE) found in Christian burials from the end of the Medieval period. These two samples resemble genetically those rural Islamic peasants discussed above, still carrying significant Maghrebi ancestry, but they are Christians. It appears safe to assume genetic continuity, which can only mean that they are part of a population, formerly known as the *Muladis*, that converted to Christianity, forcibly or not, in order to be allowed to stay. Therefore, the two of them were most likely representatives of the *Morisco* population of the time. The dates imply that they could have been *Mudejares*, but since they are buried in a Christian cemetery it is safer to assume they are early and voluntarily converted Moriscos (if the 14th-15th century dates are accurate). Forced conversions did not start in the kingdom of Valencia until the year 1525 CE.

Although there is some speculation in these assumptions, archaeological dates and historical evidence seem to point in the same direction. To my knowledge this is the first exercise attempting to link observable genetic clusters with social, cultural and ethnic groups in Medieval Iberia. The final conclusion is that we now know that although the Expulsion of the Moriscos in 1609 CE from Valencian lands was probably conceived of at the time as a suppression of a parallel culture, it ended up being an ethnic cleansing, since a large proportion of centuries-old original indigenous population that happened to carry with them significant North African ancestry was effectively deported. That is where the tragedy of the Moriscos started – not Christian enough for the Spaniards, not Muslims enough for the North Africans – cursed to forever wander the Mediterranean for a new place to live, or to disappear altogether.

One final point that the survival of significant amounts of North African ancestry until the 17th century CE highlights, is the wide-spread presence of such ancestry in South Americans as reported in Chacón-Duque et al. (2018). Officially by law, Christian converts were not allowed to migrate to American colonies, and only a few were able to make a clandestine journey to the New World. The South/East Mediterranean ancestry signature of colonial migrants to South America is too high (1-5º%) to be satisfactorily explained by special cases, or what historical records indicate. The estimates of South/East Mediterranean ancestry in Latin Americans suggests colonial migration to Latin America involved people with higher levels of South/East Mediterranean ancestry than the average modern Spaniard (around 5-8% (Botigué et al., 2013; Bycroft et al., 2018)). Furthermore, the time estimates since the Maghrebi in South America admixture are not significantly different from Iberian mixture event, which indicates that most of this ancestry is consistent with being introduced simultaneously by the initial colonial immigrants. The two Late Medieval individuals from Valencia shed light on a simpler answer to this issue, a population with increased Maghrebi ancestry (around 10%) did exist at the time in Southern Iberia. Given that cities in the South such as Sevilla and Cadiz were the main ports for the American voyages, it is reasonable to think that this was the actual Spanish source that introduced the Maghrebi ancestry because that

is where the majority of sailors and colonist came from. In summary, the results presented here about the Late Medieval genomes fit well with a previous hypothesis based on modern genomes described in Chacón-Duque et al. (2018).

5 Chapter III: Exploring the Phenotypic Space of Skin Pigmentation

5.1 Introduction

The popularity of phenotype prediction models has grown in recent times and extended to cover complex traits, such as eye colour, hair colour and even skin pigmentation. These particular traits are determined by a significant number of interacting genes and a large number of SNPs. This is especially true for skin colour (Shriver et al., 2003), although there is less consensus about the variants concerned than in the case of eye and hair colour. For example, one of the best known genes in the world of biology is the melanocortin 1 receptor (*MC1R*) locus which regulates pigmentation in several mammalian species. There are some alleles of *MC1R* carrying mutations that cause partial loss of function, the carriers of these alleles are associated with pale skin and red hair. It has been shown that not only modern humans (of European ancestry) carry such alleles for reduced functionality but they also evolved independently in the Neanderthal line (Lalueza-Fox et al., 2007). *MC1R* is nowadays accepted as a good proxy to explain or predict red hair. However, its additive effect on skin pigmentation can be countered by the effect of other variants in other genes associated with pigmentation, so it is not safe to assume an individual has light skin only because he or she carries a particular allele of *MC1R*.

In recent years, many more SNPs related to skin, hair, eye colour and freckles have been identified in genome-wide association studies. One of the studies that started this trend was the work by Stokowski et al. (2007) that used genotypes and phenotypes determined by reflectance spectrometry. This and the studies that followed have allowed researchers to identify genes linked to pigmentation in human and other mammals. Among the most relevant genes in this category we find *HERC2*, *OCA2*, *SLC45A2* and *MC1R*. As previously noted, *MC1R* was the one used to infer red hair in the Neanderthal individuals from El Sidron (Spain) (Lalueza-Fox et al., 2007; Lalueza-Fox, 2013). All these four traits likely share a molecular basis related to melanin production in one way or another (Barsh, 2003; McEvoy et al., 2006; Wilson et al., 2011; Deng and Xu, 2018). GWAS have been undeniably beneficial for the field of biomedicine in identifying risk mutations linked to genetic diseases, and also to characterize human genetic diversity, although they are not exempt of limitations (Parra et al., 2004).

Figure 61: Artistic reconstructions of European hunter-gatherer individuals: La Braña 1 (Spain), Cheddar Man (England) and Lola (Denmark). Source: press releases from Olalde et al. (2014); Brace et al. (2019); Jensen et al. (2019).

Despite the genetic complexity underlying skin pigmentation this trend has expanded even further into the field of ancient genomes (Mathieson et al., 2015; Günther et al., 2018; Deng and Xu, 2018). One of the earliest attempts at reconstructing the skin colour of an ancient individuals came in 2014 (Olalde et al., 2014) with the genome of a 7000-year-old Mesolithic Hunter-Gatherer (La

Braña 1) from Spain (Figure 61). The conclusions drawn from this pioneering work established a new vision that has guided later research on the matter. The study concluded that what we refer to nowadays as western European hunter-gatherers (WHG) carried variation compatible with a skin pigmentation darker than present-day Europeans (although unclear to what extent), and possibility of light eye colour. This indicated that a genetic sweep leading to lighter skins must have occurred later in Europe (McEvoy et al., 2006). The publication of this work was accompanied with a the release of an artistic facial reconstruction of the individual, which turned out to be very popular and has been widely reproduced in the media. However, one caveat can be pointed out. The conclusions were based solely on two SNPs in the genes SLC45A2 (rs16891982 in the hg19 build) and SLC24A5 (rs1426654 in the hg19 human reference genome) (Cheng and Canfield, 2006). However, from the position of the field today it would seem precipitate to make such inferences based on two variants alone. For example, even though eye colour is supposed to be a simpler trait to predict, a re-examination of La Braña 1 did not predict blue eyes for the individual (Brace et al., 2019) as previously advertised by the artistic reconstruction (Figure 61).

It was probably the interest in having more reliable classifications that led to the development of new methods (Walsh et al., 2013; Maroñas et al., 2014, 2015; Kayser, 2015). To make such predictions for complex characters, these methods rely on regression-like strategies after having trained the model with a large dataset of samples for which phenotypic and genotypic data are known. Note though that for various reasons the training datasets from the cited works above are not in the public domain, hindering the reproducibility of the research. Generally these kind of datasets are never made public, making access to pigmentation-related reliable phenotypic-genotypic data extremely difficult.

Some of these pieces of work have gained popularity despite the afore mentioned lack of reproducibility and established prediction models for eye and hair colours, such as the IrisPlex and HIrisPlex (Walsh et al., 2013, 2017). Typically, the classification system used in these methods for skin type prediction is based on in the Fitzpatrick scale (Fitzpatrick, 1975) which has six categories and therefore imposes a discrete classification. The Fitzpatrick scale replaced the older Von Luschan chromatic scale by means of recording the response of the skin to ultraviolet light (Figure 62A). The skin colour categories are usually built by phenotyping participants of the model training dataset with reflectance spectrometers and assigning them to one of the six categories on the Fitzpatrick scale. It is not difficult to see how this procedure results in loss of information when summarizing the gradient of human skin colour diversity into a few discrete categories. Again, this is especially problematic for the case of skin colour, given the great heterogeneity observed even within groups of similar genetic background. This scale is of course useful nonetheless and its use is a valid approach, but the Fitzpatrick scale was clearly not designed for evaluating the molecular basis of skin pigmentation. In other words, it assumes that, for example, Melanesians and sub-Saharan Africans are dark-skinned in the same genetic way (Gibbons, 2017), or that Europeans and East Asians are pale-skinned in the same genetic manner too (Figure 62B). We now know that this is not the case.

One reason for the differences and similarities in this regard between regions and populations is the correlation between skin colour and latitude, which in turn is the ultimate cause behind intensity of UV radiation (Figure 62A). The darkest skin phenotypes in human populations across the world are found in indigenous groups that inhabit the land masses overlapping with the high-intensity UV radiation band (Figure 62). This is to be expected since skin pigmentation is a highly adaptive trait. Perhaps one of the most adaptive *Homo sapiens* phenotypes, because it is under such powerful environmental selection pressure due to its constant interaction with the environment UV radiation. The dual example of modern Europeans and East Asians independently evolving a loss of pigmentation as an adaptation to a new environment is well known. Then a question arises, would it be possible to approach the issue of skin pigmentation but accounting for the evolutionary convergence of identical phenotypes from different regions of the globe? (Figure 62).

In ecology, the Hutchinsonian definition of niche is described as a space of multiple dimen-

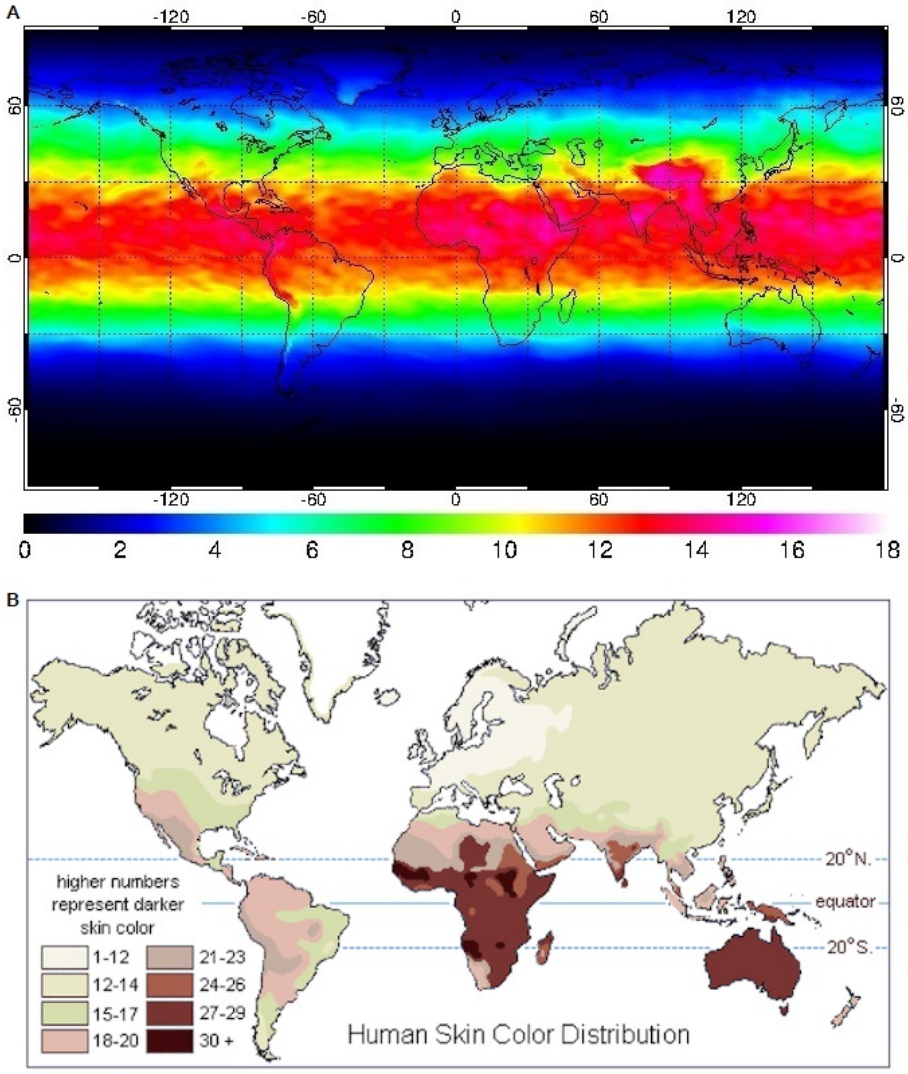


Figure 62: A) Map representation of the band of high UV incidence and how it overlaps with land masses. Source: European Space Agency, 19th of August 2019. B) Classic map of skin colour variation based on the data collected by the Italian geographer Renato Biasutti in the early 20th century. From Parra et al. (2004).

sions, or to quote literally: an *n-dimensional hypervolume* (Hutchinson, 1957; Holt, 2009). The dimensions correspond to the environmental resources, conditions and any other factor, biotic or abiotic, that can potentially impact on survival for a species. Hutchinson's concept of niche can be usefully borrowed and adapted to the field of genetics because skin colour is a highly continuous character (Figure 63) determined by several loci which define a multidimensional space. This conceptualization of the pigmentation niche would be ideal to escape the categorization imposed by the Fitzpatrick scale. When applied to a complex phenotypic trait like human skin colour, which is dictated by the additive effects of great numbers of SNPs, then it is not difficult to picture how the variants become the *n*-dimensions that define the hypervolume of skin pigmentation phenotypic space. We can therefore speak of phenotypes in a Hutchinsonian space. Thanks to dimensional reduction techniques like principal component analysis (PCA) or t-distributed stochastic neighbour embedding (tSNE), it is possible to condense this information into 2D or 3D figures (van der Maaten and Hinton, 2008; Li et al., 2017).

In summary, this theoretical Hutchinsonian approach offers an alternative to a complex trait prediction using regression models (Zaorska et al., 2019). There are some instances where similar

approaches have been attempted (Maroñas et al., 2014, 2015). When working with one of these predictive regression models it is important to take into account certain issues that might affect the outcome. One is the number categories created when building a model based on a discrete classification system. If a methodology uses only few categories, dark *vs* light skin would be the extreme, distinguishing between the two phenotypes might seem intuitive but also simplistic because intermediate pigmentations will be difficult to assign to either category and there will be a great degree of uncertainty. Creating more categories suits continuous traits better, however having as many categories as individuals is also problematic because it results in an over-fitted model. So if an arbitrary line has to be drawn somewhere, why not avoid discretization altogether?

Finally to further summarize all the ideas presented, another issue to watch out for, is how baseline values are established. This can be an arbitrary process and completely relies on the composition of the training dataset. How the scale is defined based on the training data as well as the measuring tool can introduce bias and have large effects on the predictive outcome. Representative diversity is important but most studies fail to account for this issue, and like most GWAS, they are biased towards groups of European ancestry. The HIrisPlexS is no different: 74% of the individuals from the dataset are Polish and Irish alone (Walsh et al., 2017) and a further 6% are Greeks. Africans as a whole represent only 2.5% of the training data, and Oceania (represented by Papuans) covers less than 1% of the individuals. This is a tremendous imbalance for an ambitious piece of work that claims *"Global skin colour prediction from DNA"* in the title. Worldwide diversity harboured by groups of non-European ancestry remains understudied and this is not only for the case of skin pigmentation (Barsh, 2003; Tang and Barsh, 2017). This is potentially hindering predictions at the wider spectrum (Parra et al., 2004; Martin et al., 2017). However, there have been recent efforts to address this shortcoming (Figure 63).

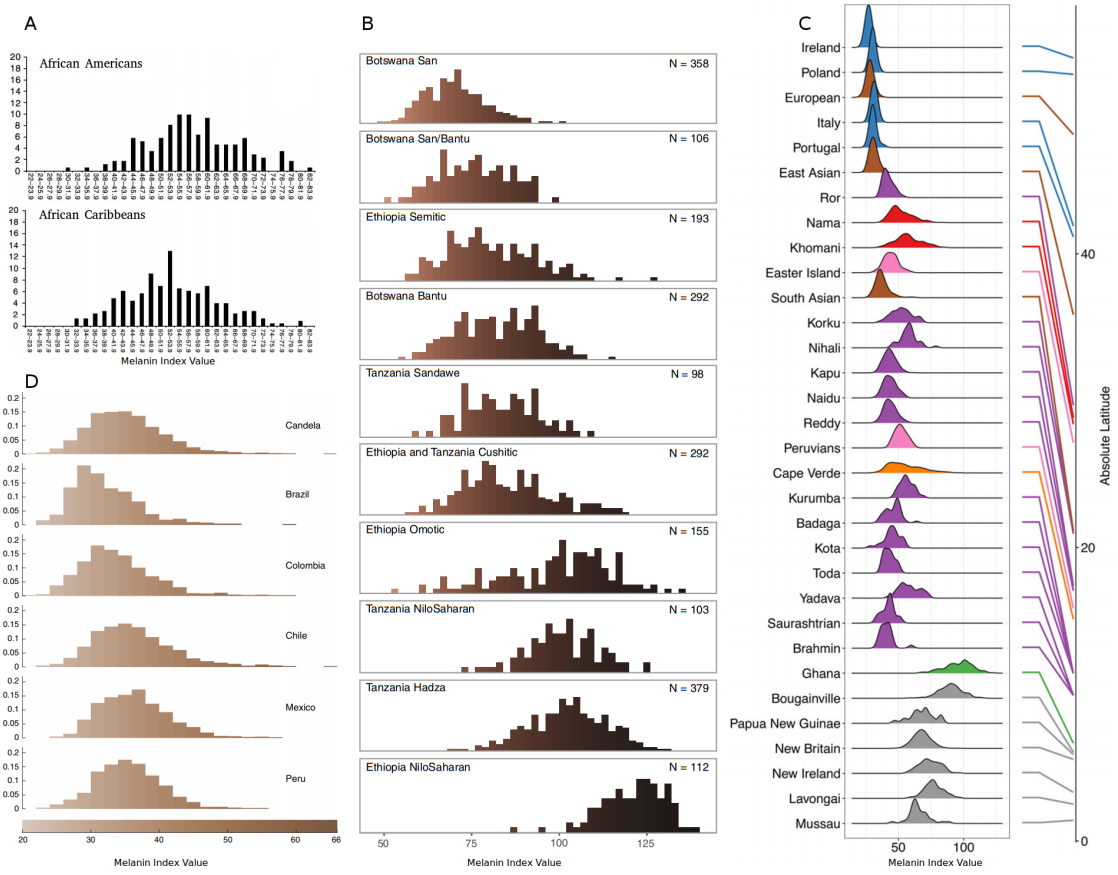


Figure 63: Skin pigmentation is a continuous trait as exemplified by melanin index values in different modern populations from Africa and America. Note that in all cases melanin values follow an approximate normal distribution in each population. Modified or taken from Parra et al. (2004) (A), Crawford et al. (2017) (B), Martin et al. (2017) (C) and Adhikari et al. (2019) (D).

5.2 Methods

5.2.1 Genotyping of the samples

The sample set used in this work comprised the 279 publicly available VCF files from the Simons Genome Diversity Project (SGDP) (Mallick et al., 2016), last updated Wednesday 24th January 2018 (Version 2). These samples are aligned to the hg19/GRCh37 human reference genome using the BWA *mem* algorithm, as stated on the SGDP webpage. The metadata file for the 279 samples was also downloaded from the same SGDP site, no changes were made to the labels or any other information contained in the metadata. No phenotypes were assigned *a priori* to the samples but it was expected that African, Oceanian some South Asian samples would represent darker phenotypes and European as well as other Asian samples would represent lighter phenotypes.

The samples were genotyped based on the subset of 36 SNPs selected in Walsh et al. (2017) as significantly correlated with skin pigmentation. This selection was made in order to have a direct comparison with the phenotypic results predicted in Brace et al. (2019). The selected SNPs in the variant call format files of the SGDP were parsed into individual genotype files using an in-house developed Python script. The individual genotype files were translated into a trinary matrix (with the two homozygous allele states and the heterozygous state) following the reference and alternative alleles present in the human reference genome (build hg19/GRCh37.p13).

The data for the ancient genomes was downloaded from the EBI European Nucleotide Archive in the form of FASTQ files or BAM files when available. The FASTQ files from each sample were aligned to the human reference (hg19/GRCh37.p13) using the BWA algorithm. The resulting BAM files were sorted using Samtools and duplicates were marked using Picard. From the BAM files, gVCF files were generated with GATK HaplotypeCaller to genotype for the 36 pigmentation SNPs.

Two extra samples of known origin and phenotype (Sicilian; brown eyes/dark hair and Irish; blue eyes/blonde hair) for which I had 23andMe data available were tested also in the HIrisPlex-S DNA Phenotyping Webtool for eye, hair and skin colour. The data from 23andMe for both samples were downloaded and the information about the SNPs required by the HIrisPlex compiled and introduced in the online webtool of the HIrisPlex-S (https://hirisplex.erasmusmc.nl/).

5.2.2 Gathering pseudo-phenotypic data and skin palette build

The SGDP samples do not have any phenotypic data associated to them so I created pseudo-phenotypes for each region by proxy. For this, I gathered around one hundred photo portraits of individuals from different parts of the world and mesured the RGB values from their faces, more specifically three measurements were taken from both cheeks and the area just above the space between eyebrows whenever the light appeared natural and there were no reflections. I then recorded the averaged RGB values from their faces using the freely available ColorInspector3D software (https://imagej.nih.gov/ij/plugins/color-inspector.html) originally developed as a plugin of ImageJ by Kai Uwe Barthel. In this way I was able to reconstruct an artificial palette of skin tones that can be used as a proxy for real phenotypic data. Since the individuals in the SGDP cannot be linked to any of the specific photographs I grouped the individuals from the photo dataset following the regions in the SGDP metadata file, with the addition of some categories, and I averaged the RGB values for each region. Then the individuals from those regions in the SGDP dataset were assigned the averaged values obtained. I acknowledge the great deal of measurement error associated to the approach since the photos have diverse origins and were taken with different cameras. However, the goal is not accuracy or a real outcome but to develop a methodology.

The photo portraits measured originate primary from The World in Faces project by Alexander Khimushin (khimushin.com) because it comprises a myriad of diverse portraits of non-European individuals across the world and they were a good proxy to represent the groups present in the SGDP datset. Other portraits used can be found at the webpages of the following photographers: Wendy Simmons (wendysimmons.com), David Lazar (davidlazarphoto.com) and Steve McCurry

(stevemccurry.com). I tried to contact Alexander Khimushin unsuccessfully for permissions, so to avoid copyright problems I have tried to not reproduce any photos used with the exception of one of a Mursi girl dispalyed. All photos were used purely for academic purposes, all credit and rights of the portrait reproduced here belong to Alexander Khimushin, creator of The World in Faces project.

Ideally there would be no categories because all individuals in a real dataset should have phenotypic information associated to them, however since this was not possible I had to create 10 categories based on the division by region from the SGDP metadata. I ended up measuring ten portraits for each of the ten regions and assigned the averaged RBG values of the different faces to create the palette of the regions.

5.2.3 Dimension reduction and clustering techniques

To explore the phenotypic hypervolume of skin pigmentation I used three dimensional reduction methods. One was PCA in R, to capture linear relationships between samples. The seconds was tSNE in R, to explore non-linear information between the samples in the 2D and 3D space. I used R packages *prcomp* and *Rtsne* to make the reduction of dimensions calculations and obtain the new coordinates. The parameters for the tSNE were 10.000 iterations, perplexity=30, verbose=FALSE and three dimensions. The third was Autoencoder in Python. I plotted the results in two and three dimensions. To plot the distribution of the samples in the phenotypic hypervolume I used *ggplot* and *plotly* packages in R. To group the samples into clusters in the phenotypic space, I ran K-means in R (for values 2 to 6). Since this is an iterative method it was run several times and a probability of belonging to a cluster was assigned to the predicted samples. To test confidence in the classification of the samples in the different clusters I ran Silhoutte analysis in R.

5.2.4 Geography and Procrustes transformation

On the grounds that skin pigmentation correlates with geography, I explored the correlation of the phenotypes drawn by the set of 36 SNPs and its correlation with the geographic origin of the individuals of the SGDP using Procrustes. I did so by performing a procrustean transformation of the resulting 3D coordinates with 36 SNPs and the geographical coordinates of the samples as target matrix. I transformed geocoordinates into spherical coordinates to check whether it helps overcome the deformations derived from projecting worldwide coordinates in a two dimensional Mercator space. I used the *procrustes* function from the *vegan* package in R to perform the transformation in three dimensions. I later tested the significance of the transformation using the *protest* function from the same *vegan* package with 999 permutations each.

5.2.5 Two dimensional interpolation and RGB colour map

The mathematical problem I wanted to address consisted in interpolating from irregularly-spaced data, a scatter plot with tSNE coordinates in this case, to build up a continuous surface. I have n triplets (x_i, y_i, z_i). The pair (x_i, y_i) stands for the locational coordinates of the data point D_i (tSNE, PCA, geography,...) and z_i is the corresponding feature (RGB colour channel intensity). An interpolation function $z = f(x, y)$ assigns a value to any location $P(x, y)$.

Let $d[P, D_i]$ (shortened to d_i) be the Euclidean distance between P and D_i. In Cartesian coordinates $d_i^2 = (x - x_i)^2 + (y - y_i)^2$. I used the Shepard method to build up the interpolated value at P, which is given by

$$f(P) = \frac{\sum_{i=1}^{n} d_i^{-2} z_i}{\sum_{i=1}^{n} d_i^{-2}} \quad \text{if } d_i \neq 0, \tag{9}$$

and $f(P) = z_i$ if $d_i = 0$. This is the basic procedure where all data points intervene with a weight inversely proportional to the squared Euclidean distance. Besides, the shadowing of the influence

of a data point by a nearer one in the same direction is considered to modulate the weighting function. Eventually, local slopes are taken into account to provide correct partial derivatives.

The interpolation function $f(P)$ is based on the quadratic Shepard method for bivariate interpolation of scattered data (Algorithm 660) (Renka, 1988). The surface defined by $f(P)$ is continuously differentiable and assumes the required value at all data points. The final interpolation can extrapolate up to 10% outside the range of initial values. Interpolated values for Red, Green and Blue were capped according to the lower end (at 0) and higher end (at 255) of the RGB scale because there may be extreme value in areas with no informative points. Three interpolation maps were made, one for each of the Red, Green and Blue values associated to the genetic coordinates of the 279 SGDP samples. The three interpolation maps were merged into a portable pixmap format file (ppm) which reconstructs approximate skin colours. This image was finally edited to remove excess of primary colours.

5.3 Results

5.3.1 Testing the reproducibility of the HIrisPlexS

I downloaded the Simons Genome Diversity Project (SGDP) genomic data in the form of VCF files and the metadata with the information about the geographic origin of each sample, with the knowledge that despite being a very complex character, skin pigmentation roughly correlates with latitude and broad geographic origin. The 279 SGDP individuals were genotyped at the same 36 SNPs used to predict the skin pigmentation of the Cheddar Man in Brace et al. (2019) Then I reproduced the formula in Walsh et al. (2017) and predicted for the 279 individuals their probability of being very pale, pale, intermediate, dark and black. I hoped to see a correlation between locations and categories. However, the values of the alpha coefficients necessary to adjust the multinomial logistic regression (MLR) model are not provided in the publication and the authors in Walsh et al. (2017) and Brace et al. (2019) (S. Walsh, M. Kayser and Y. Diekmann) refused to provide us that information after consultation. With no further information about the alpha coefficients, I proceeded assuming alpha values equal to zero. Details about the model MLR parameters can be found in Walsh et al. (2017) supplementary material, the absence of some key parameters makes it impossible to confidently reproduced the predictive model.

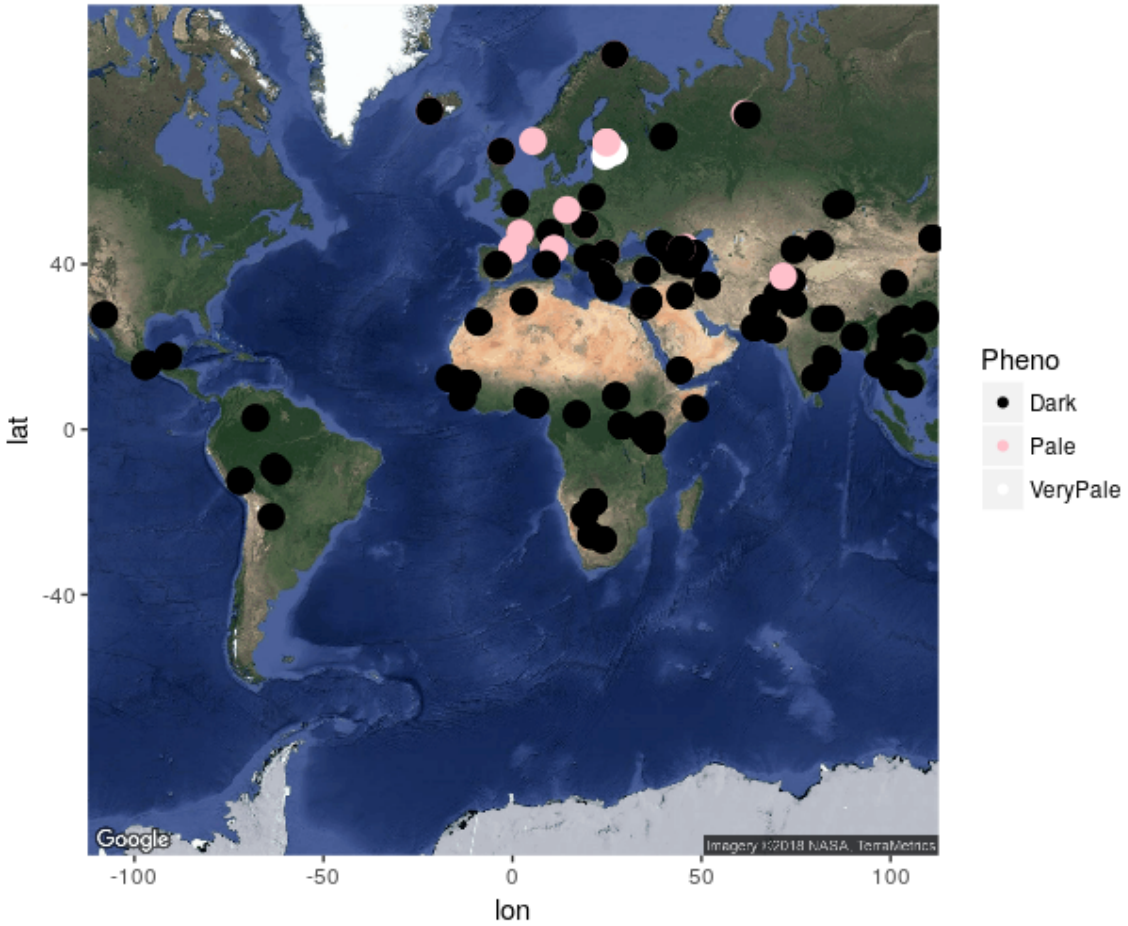


Figure 64: Predicted skin pigmentation phenotype for the publicly available samples from the Simons Genome Diversity Project (SGDP). Prediction was made following the multinomial logistic regression (MLR) from Walsh et al. (2017).

There was some relative success with this assumption. As it can be seen in Figure 64, very pale and pale skin phenotypes are confined to Europe only. However, my reproduction of the HIrisPlexS model using the same 36 SNPs only yielded individuals classified in three (*Very Pale,*

Pale and Dark) out of the five categories (*Very Pale, Pale, Intermediate, Dark and Very Dark*). The two categories that did not feature in the predicted outcome of any SGDP sample were Intermediate and *Very Dark*. No African individuals were assigned to the *Very Dark* category instead they all fell in the *Dark* category. In fact, all non-European samples fell in the *Dark* category including the SGDP individuals from South America, East Asia, South Asia, Central Asia and West Asia. Also, as seen in Figure 64, a few Europeans also were classified within the *Dark* skin category.

To further test whether this was a limitation derived from lacking the real values of the alpha coefficients of the HIrisPlexS MLR or the actual behaviour of the model, I also tried the online HIrisPlexS web-tool. I was referred to this online tool by the authors of Walsh et al. (2017) after my failed request to gain access to the alpha coefficient values. I made use of 23andMe data from two members of our own research group (one Sicilian and one Irish) for whom the phenotypes were known. I genotyped the Sicilian and Irish individuals for the 36 SNPs necessary in the HIrisPlexS model to make a prediction, and I introduced the genotypes in the online HIrisPlexS tool (Figure 65). For the Irish individual the predicted outcome was blue eyes, blond hair, and pale skin. This prediction matches the actual phenotype of the Irish individual. However, the predicted phenotype for the the Sicilian individual also was blues eyes, blond hair, and intermediate skin. This outcome does not match the phenotype of the Sicilian individual from Figure 65 (except the intermediate skin type). Like our mimicked HIrisPlexS MLR formula, the online tool also offers dubious results. As a sanity check, prior to the genotyping of the 36 pigmentation SNPs to asses the reliability of the 23andMe data I tested a trait simpler than pigmentation. In order to do so, I genotyped three SNPs in chromosome 9 necessary for blood group determination (rs8176747, rs8176746 and rs8176719). The prediction of the blood type for the Irish and Sicilian individuals based on the 23andMe data matched their known blood groups.

I further explored the ability to predict blood groups from DNA in a subset of SGDP samples and in the high coverage ancient genomes used for the pigmentation prediction I will discuss below (see Supplement III, Figure S41).

Predicted phenotype

	p-value	AUC Loss
blue eye	0.848	0
intermediate eye	0.088	0
brown eye	0.065	0
blond hair	0.702	0.01
brown hair	0.258	0.007
red hair	0.01	0.034
black hair	0.03	0.002
light hair	0.955	-0.001
dark hair	0.045	-0.001
very pale skin	0.014	0.022
pale skin	0.392	0.017
intermediate skin	0.586	0.024
dark skin	0.004	0.003
dark to black skin	0.004	0.004

Figure 65: Phenotypic prediction by the HIrisPlexS model using 23andMe data from a Sicilian individual. Blue eyes and blonde hair are predicted not matching the known phenotype. There was missing data for eleven SNPs.

5.3.2 Dimensional reduction and correlation with geography via Procrustes

I first tried to plot the 279 individuals in a PCA with the same 36 SNPs (from Walsh et al. (2017)). Some structure and loose clusters can be appreciated but not enough for a confident assignment of a skin colour type since there is great overlap (Figure 66).

I then tried t-Distributed Neighbour Embedding (tSNE) in R as an alternative to PCA to explore non-linear relationships. It becomes clear that the novel tSNE technique introduced here provides a much better resolution than the classical PCA approach, at least when working with an extremely small dataset of markers (note again that this is only 36 SNPs) related to skin pigmentation. I plotted the tSNE outputs made with the 36 SNPs in 2D and 3D. Plots of tSNE coordinates in 2D show clear clusters although each run is marginally different from each other. However, in 3D runs were stable and distinguished clear clusters for Europeans, Africans, Oceanians, East Asians, South Asians and Native Americans. Although this is very hard to display printed in 2D, there are interactive 3D files available on request (Figure 66).

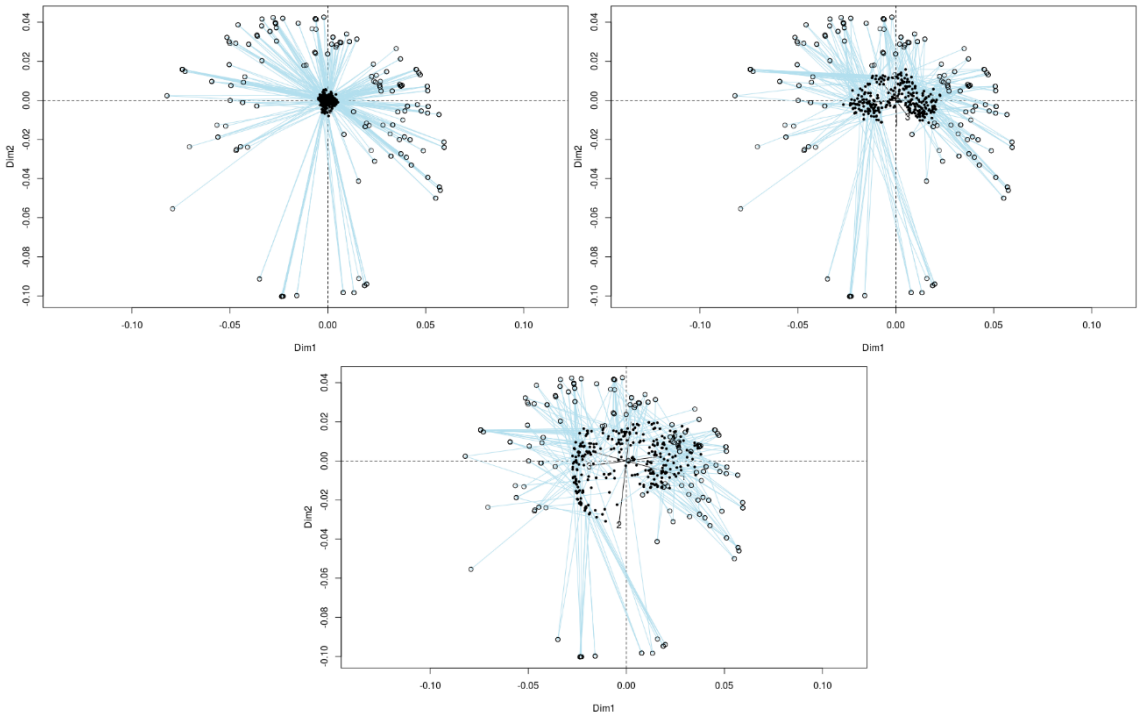


Figure 66: Procrustes transformation of three dimensional reduction techniques used (PCA, tSNE, Autoencoder). The first 3 genetic coordinates of each dimensional reduction (PCA top left, tSNE top right, Autoencoder bottom) of each individual from the SGDP were transformed into their corresponding spherical geographic coordinates.

A third dimensional reduction technique was also tried, it is known as Autoencoder and was defined as a non-linear Principal Component Analysis that uses auto-associative neural networks. The clusterization results seemed at least as good as the ones seen in the tSNE plots but following a spherical distribution. However, I did not proceed further with Autoencoder because handling the data with tSNE was faster and more convenient (Figure 66). All these three methods proved that the dataset of SNPs assembled with the 36 SNPs and used in the Cheddar Man prediction is able to distinguish - at least in part - different types of skin pigmentation (Figure 66).

It was immediately appreciated that the clusters obtained with tSNE with only 36 SNPs in genes with functions related to pigmentation seemed to show overall correlation with the geographical distribution of the SGDP samples. To test this observation formally, I took the latitude and longitude geographical coordinates for the samples from the SGDP metadata. I converted the 2D geographical coordinates into spherical coordinates to have 2D and 3D coordinates for the

Procrustes transformations and to overcome issues of projecting a globe in a 2D map (Figure 66 and Figure 67).

I found that there is a positive correlation, tested with Protest in R, between the genetic coordinates of the individuals and their geographic location when projected with Procrustes using a 2D and 3D coordinates for the cases of PCA, tSNE and Autoencoder (Figure 66 and Figure 67). To test whether these SNPs were giving this outcome only by chance, I generated a dataset simulating random genotypes. Each SGDP individual had its personal genotype recreated with random SNP genotypes. When plotting this random genotypes dataset I observed that both PCA and tSNE plots lost cluster structure they had with their original genotypes of the 36 SNPs, ergo, correlations have been lost. The proportion of homozygous and heterozygous sites in the genotypes was preserved to avoid the risk of loss of structure in tSNE because too many homozygous or heterozygous sites to reference positions have changed. This confirmed that there is genuine information about pigmentation contained in the original set of 36 SNPs defined in Walsh et al. (2017).

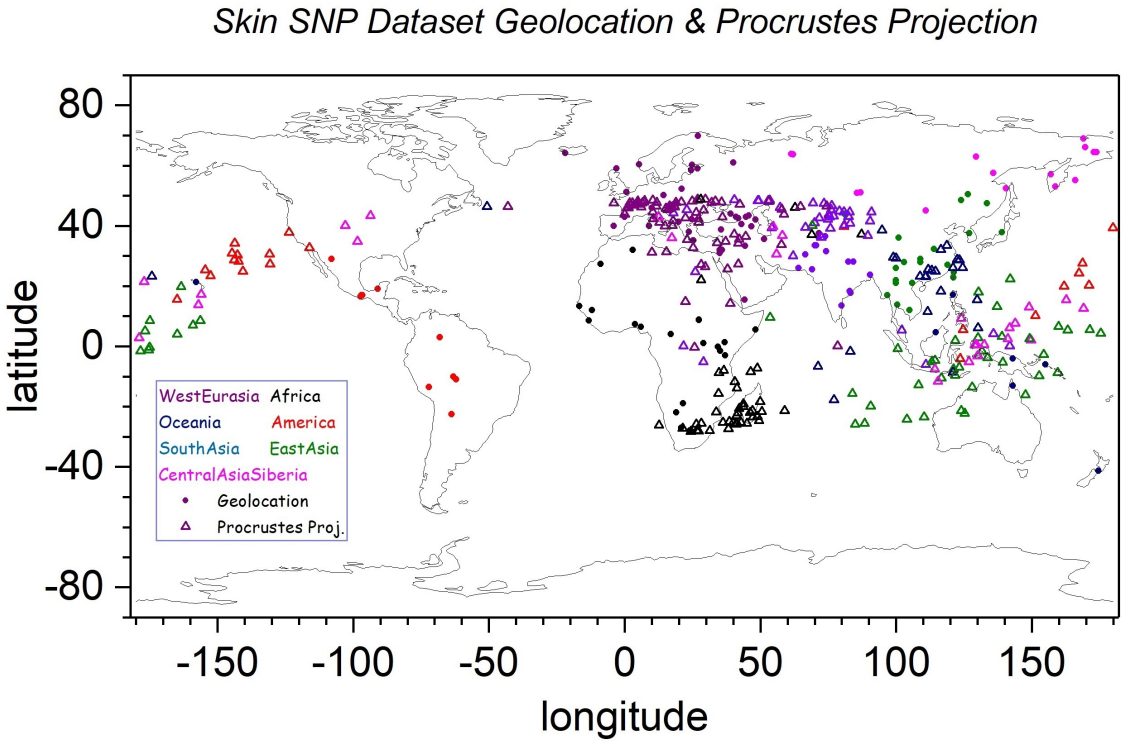
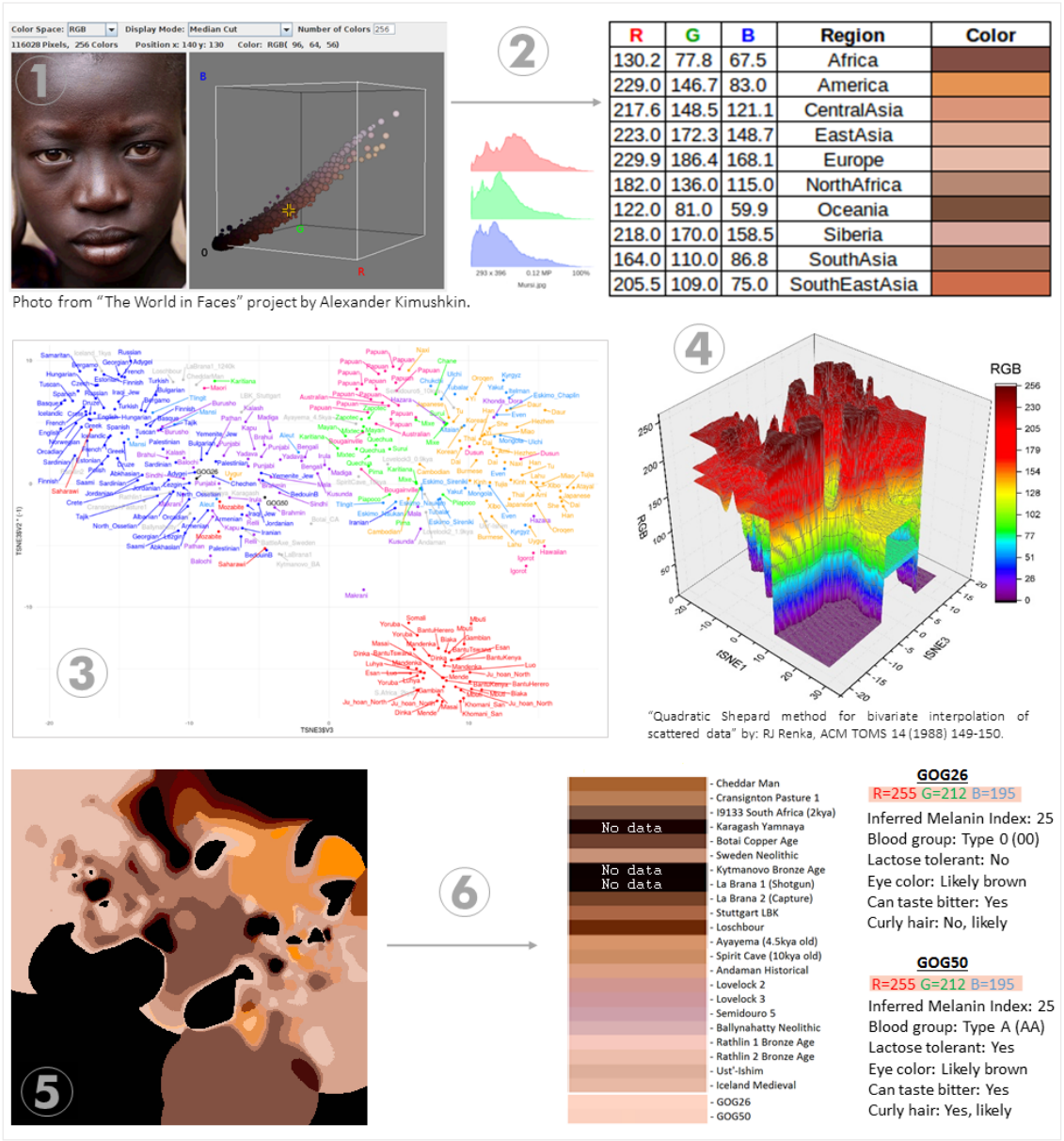


Figure 67: Procrustes transformation of Autoencoder coordinates in 2D.

5.3.3 Building a new methodology for skin type prediction

My results so far have shown that the set of 36 SNPs linked to pigmentation in humans, identified in Walsh et al. (2013, 2017) carry sufficient information to distinguish clusters of broad continental groups based on their skin pigmentation. However, I have also shown that the predictions of the HIrisPlexS model lack reproducibility and accuracy for reasons I will discuss later on.

To overcome the HIrisPlexS limitations encountered, an alternative method for pigmentation prediction in ancient genomes was developed. For comparison the new approach still makes use of all 36 SNPs from the HIrisPlexS used for Cheddar Man´s prediction. The development of the idea can be divided in a series of steps (Figure 68).



R	G	B	Region	Color
130.2	77.8	67.5	Africa	
229.0	146.7	83.0	America	
217.6	148.5	121.1	CentralAsia	
223.0	172.3	148.7	EastAsia	
229.9	186.4	168.1	Europe	
182.0	136.0	115.0	NorthAfrica	
122.0	81.0	59.9	Oceania	
218.0	170.0	158.5	Siberia	
164.0	110.0	86.8	SouthAsia	
205.5	109.0	75.0	SouthEastAsia	

Figure 68: Workflow methodology developed in the present work to predict pigmentation based on 36 SNPs from the HIrisPlexS system.

Firstly, I generated pseudo-phenotypes for the skin colour palette by grouping the hundred photo portraits into world regions according to their origin, and I then measured the RGB values from the photo portraits with Color Inspector 3D software (Step 1 in Figure 68). The resulting colour palette for each world region, by averaging and merging RGB values of the individual

portraits in each group, can be appreciated in Step 2 of Figure 68. The dimensional reduction made with tSNE (perplexity 30) in three dimensions was plotted in 2D and selected for the later interpolation base. The coordinates of SGDP and high-coverage ancient samples in this plot were used as the interpolating points. This tSNE was made using only the set of 36 SNPs used in HIrisPlexS for predictions about pigmentation (Step 3 in Figure 68). I then associated the palette values of RGB colour to each individual point, based on geographic origin, in the tSNE plot in three different layers, one for Red, one for Green and one for Blue. The values of Red, Green and Blue were then interpolated independently in each layer based on the pseudo-phenotypic values attributed to each point (Step 4 in Figure 68). Finally, the three interpolated layers were merged to recreate real colours similar to the ones in the pseudo-phenotype palette. The outcome of this resulted in what I have labelled as an Interpolated Colour Map (Step 5 in Figure 68). From the Interpolated Colour Map I was able to recover RGB values at the positions in the tSNE space where the ancient samples with no previous RGB data where located (Step 6 in Figure 68).

I managed to obtain results (Figure 68) for 19 out of the 22 public ancient high coverage genomes included here in addition to two samples with more than 1X coverage from the previous chapter (GOG50 from the Roman period and GOG26 from the Islamic period in Spain). I previously removed four samples that had more than 10% of missing data. The workflow for the making of an Interpolation Colour Map is complete but this approach is still in development as an open alternative to the HIrisPlexS black box.

Based on the results presented in Crawford et al. (2017) and Adhikari et al. (2019) I reversed engineered a system to estimate melanin indices (MI) from RGB values, using a polynomial regression (top row in Figure 69A). On average the melanin index, of three European hunter-gatherers successfully used here, was 85. This value is four times higher than in modern Europeans on average. It was determined that La Braña 1 Red-Green-Blue values were 117-66-39 (84 MI), for Loschbour RGB values were 108-40-5 (105 MI), and for Cheddar Man were 168-98-46 (67 MI) (Figure 69B). The average melanin index for four Neolithic European individuals was 38. Ballynahatty had RGB values of 218-176-178 (23 MI), LBK Stuttgart had 171-102-69 (45 MI), Carsington Pasture 1 had 185-126-86 (49 MI), and RISE98 had 200-146-120 (35 MI) (Figure 69B). The two Irish Bronze Age samples (Rathlin1 and Rathlin2) had R-G-B values of 247-200-190 and 237-190-172. These RGB values are at the higher end of the interpolation curve for MI and both can only be classified as having a value of 22 for MI (Figure 69B). This value of MI is already similar to the one seen in modern Europeans (Figure 69B). For two European Middle Age samples (Spanish Islamic and Medieval Icelandic) of around 1000 years old each, I also determined an average MI of 24. The Icelandic sample SSG-A2 RGB values were 232-185-165 (25 MI) and for the Islamic samples from Spain the values were 255-212-195. These RGB values are also extreme for the interpolation curve so I can only determine a values of 22 for the MI (Figure 69).

Finally, I plotted these MI I obtained for different periods, with minimum two samples in each, and adjusted a polynomic regression line. A steady decrease in MI values can be observed from Mesolithic times to the Bronze Age. Mesolithic Europeans have the highest MI values, Neolithic individuals have intermediate values, and post-Bronze Age samples have values of MI similar to the modern European average as seen in Figure 69.

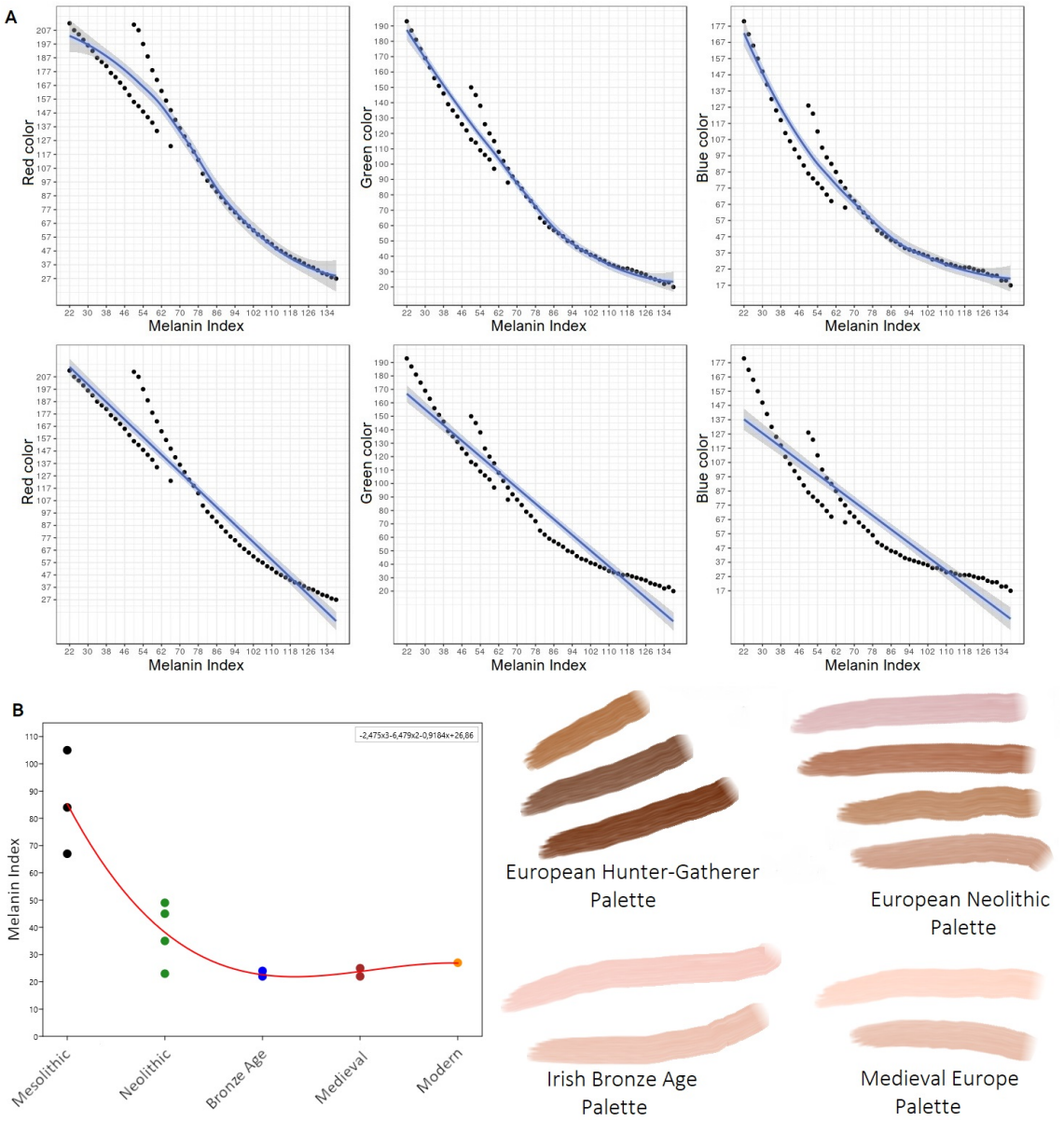


Figure 69: A) Reversed inference of melanin indices from reconstructed RGB values. The values are RGB values with their corresponding melanin indices recovered from Crawford et al. (2017) and Adhikari et al. (2019). Adjusted linear regression (top row) and adjusted multinomial logistic regression (bottom row). The calibration bias of the measuring tool from both studies can be appreciated in the different trends of the two different group of dots from each study. B) Evolution of Melanin Index through time in Europe and the reconstructed RGB colours for the periods of some of the ancient European genomes used here (Loschbour, La Braña 1, Cheddar Man, Stuttgart-LBK, Ballynahatty, Carsington Pasture 1, RISE98, Rathlin1, Rathlin2 and SSG-A2).

5.4 Discussion

The motivation for the work that unintentionally led to the completion of this chapter, were the skin phenotypes described for a series of European hunter-gatherers, including La Braña 1 (Spain), and Cheddar Man, (Olalde et al., 2014; Brace et al., 2019). The genome of Lola, a Danish hunter-gatherer girl recovered from a prehistoric chewing gum had not been published yet and therefore never included in the dataset, but it certainly will be interesting to include her in the future re-analyses. The initial goal was simply to reproduce the HIrisPlexS results obtained for the Cheddar Man in Brace et al. (2019). This was in anticipation of the interest for my own samples, since one of my aims was also being able to give an estimation about skin pigmentation in the ancient genomes that I was due to sequence myself.

However, I soon encountered serious limitations (Figures 64 and 65), namely in the form of implementing into my Python script the mathematical model that serves as the basis of the HIrisPlexS. The alpha coefficients necessary to adjust the multinomial linear regression (MLR) are not disclosed in the original paper, and nor have they been made available by Walsh et al. (2017). I contacted the corresponding authors, S. Walsh and M. Kayser, directly but they refused to provide us with the values of the coefficients or the raw data to re-calculate the MLR independently. As a result, the implementation of the mathematical model in the in-house code was handicapped from the start.

I tried nevertheless to validate and explore the information contained in the current best set of SNPs for human pigmentation (Maroñas et al., 2014, 2015; Walsh et al., 2017) (Figures 64 and 65). It was then when I also realized the imbalance in the training data used to build the model. Three quarters of the dataset are composed of Polish and Irish individuals alone. This underlines a tremendous bias affecting the predictive ability of the model. This may well explain the accuracy for the results of the Irish individual obtained using 23andMe data, and the inaccuracy for the prediction of the Sicilian (Figure 65). A training dataset composed of many Irish samples might be powerful for predictions involving a person of Irish ancestry but weak for a person of Mediterranean ancestry. My conclusion is that although the HIrishPlexS has gained a lot of popularity, the reliability of its predictions is questionable.

Another worrying issue is the inability to reproduce the mathematical model underlying the pigmentation estimates of the HIrisPlexS. The lack of knowledge about the alpha values to adjust the MLR in my code meant that none of the African samples from the SGDP dataset was assigned to the very dark category, which should be taken as a warning regarding the lack of transparency. I suspect the assumption as zero of the values of alpha plays a role; there might still be issues, but this was impossible to test due to the refusal to cooperate by the authors.

Given the limitations and over-prediction of dark phenotypes (Figure 64), I decided to approach the prediction in a different way (workflow in Figure 68). The new approach is category-free and non-discrete because skin pigmentation is possibly the most continuous trait of any human phenotype. In this work I have introduced the technique of tSNE, chosen because the clusters generated were clearly differentiated and served a purpose for the issue of skin colour prediction. I found that tSNE outperformed the PCA and Autoencoder approaches in terms of resolution by outputting four discernable clusters. The tSNE approach also handles outliers better while retaining non-linear relationships between samples.

The fact that I obtain such clear clusters, differentiating with only 36 SNPs between Europeans, East Asians, Oceanians, Africans and Americans was interpreted as an indication that although probably biased towards Europeans, these 36 SNPs have sufficient information to discriminate differences in allelic composition in genes related to pigmentation (Steps 3 and 5 in Figure 68). This is why a positive correlation with geography was detected, although this can also be related to ancestry to some degree and not only differences in pigmentation.

In the gene–geography correlation, Oceanians are separated from Africans in a different cluster (Figure 68, Step 3), much like Europeans and East Asians, because they have evolved dark pigmentation on an independent genetic trajectory. However, this fact is systematically over-

looked when grouping Melanesians and sub-Saharan Africans under the same dark skin category in predictive regression models, which is the case of the HIrisPlexS.

It is very exciting to work with RGB colour but I wanted to be able to work with a quantifiable unit that was more manageable to study the evolution of pigmentation through time. This need led me to the idea of recovering the RGB values associated to melanin values (MI) from the works of Crawford et al. (2017) and Adhikari et al. (2019) (Figure 69A). The RGB-MI interpolation curves I obtained allowed me to reconstruct melanin values from the estimated RGB values predicted for the European ancient genomes included in this study (Figure 69B).

This work never intended the results to be taken as hard evidence since I did not have genuine phenotypic data, in fact this work should be taken just as an example of how the new methodology presented works. However, the results obtained approach very closely what has been reported before about pigmentation of European hunter-gatherers (Olalde et al., 2014; Brace et al., 2019; Jensen et al., 2019).

Based on the melanin index, the results presented conclude that Mesolithic individuals (a total of three analysed here) from Europe had melanin levels twice as high as Neolithic farmers and four times higher than Bronze Age people and modern Europeans. This conclusion is very much in line with the current agreement that light skin pigmentation is a very recent innovation that appeared somewhere in Europe in the last few millennia. It is also in agreement with the evolutionary perspective and model proposed by McEvoy et al. (2006). This change in pigmentation was probably linked to other changes in diet and favoured by natural selection of genetic variants associated to metabolism, as well as a relaxation of UV levels. However, if darker pigmentation in Europe survived so long after the out-of-Africa event (60 kya) (Manica et al., 2007) as confirmed by the European hunter-gatherer phenotype (10–7 kya), this must mean that the relaxation of natural selection for UV radiation could have been not so crucial after all or that darker pigmentation still provided certain advantages (Figure 69B).

Without more testing, and better and more data, it is impossible to be more precise about the nature of hunter-gatherer pigmentation in Europe. However, based on their position in the I would suggest that they were dark-skinned in a different way to that in which modern Melanesians, South Asians and sub-Saharan Africans have evolved their pigmentation. This also highlights the complexity of pigmentation as a continuous trait, andthe fact that different genetic pathways can converge in similar phenotypes. In any case, European hunter-gatherers appear to be much darker than any later prehistoric or modern European population in light of both my results and previous results (Figures 68 and 69B).

In conclusion, although advances have been made in the topic, we should not forget that it is difficult to judge the accuracy of the prediction for ancient DNA data. Since we do not have hard evidence of what people from the distant past looked like, there is a high risk of inaccuracy right now in such predictions. We should also not forget that the SNPs currently being used derived mostly from studies on Europeans or East Asians and probably have negative bias on the predictive outcome of other groups. For modern populations ancestry can be an indicative proxy of what their skin looks like (Parra et al., 2004; McEvoy et al., 2006). But we don't have that advantage with ancient samples because we will rarely have any phenotypic data other than that deriving from the analysis of the skeletal remains. Inaccurate predictions can be misleading and have repercussions given the media impact that these kind of publications can have (Brace et al., 2019).

I hope that this new method to predict different skin type phenotypes in a hypothesis-free manner keeps developing. A series of weaknesses in how the current approach was built have been identified, and I have also devised some strategies to address these limitations. To move forward further it would first be necessary to gather real phenotypic data, and looking ahead, skin colour prediction models should expand the SNPs datasets and abandon discrete category classification. This will help avoid bias introduced by the composition of the training dataset, but diverse datasets should still be actively promoted. I have also noticed that some ancient samples

with no RGB information participating in the dimensional reduction tend to appear in regions of no reconstructed data colour within the Interpolation Map. This is a drawback and projecting them instead of making them participate in the tSNE perhaps would be better. However, the best route might be skip the dimensional reduction step altogether and generate a multidimensional colour map. This would be more in line with the idea of the phenotypic hypervolume presented in the introduction. The trade-off would be that that there will be no visual maps to show the continuity of a trait like human pigmentation in a 2D Interpolation Colour Map.

6 Supplementary Material

6.1 Supplement Chapter I

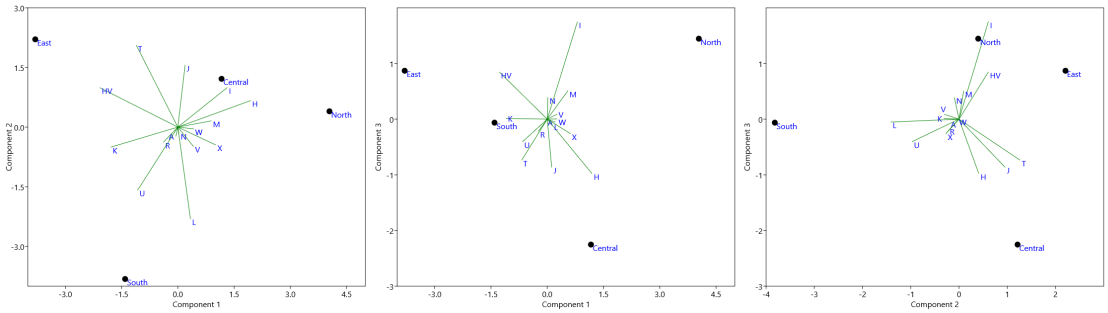


Figure S1: PCAof the cardinal sub-divisions of Spain based on the mitochondrial composition. Calculated with the macro-haplogroups indicated in the biplot green lines.

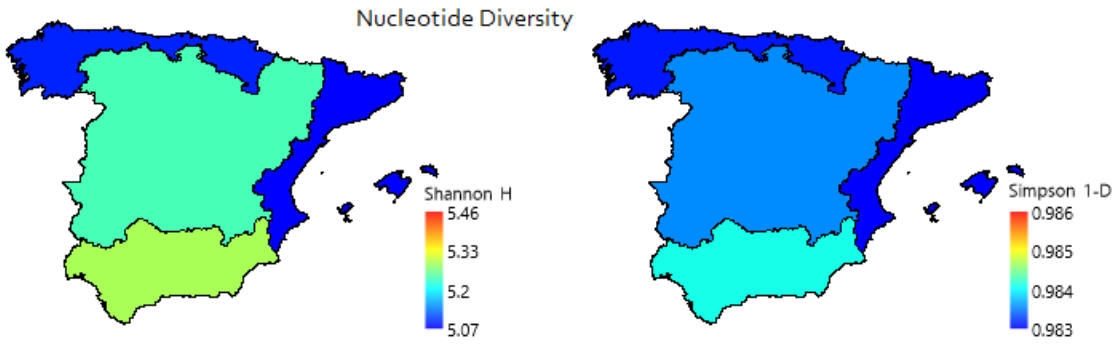


Figure S2: Nucleotide diversity in Spain, as measured with Shannon Index (H) and Simpson Index (1-D), by cardinal area grouping.

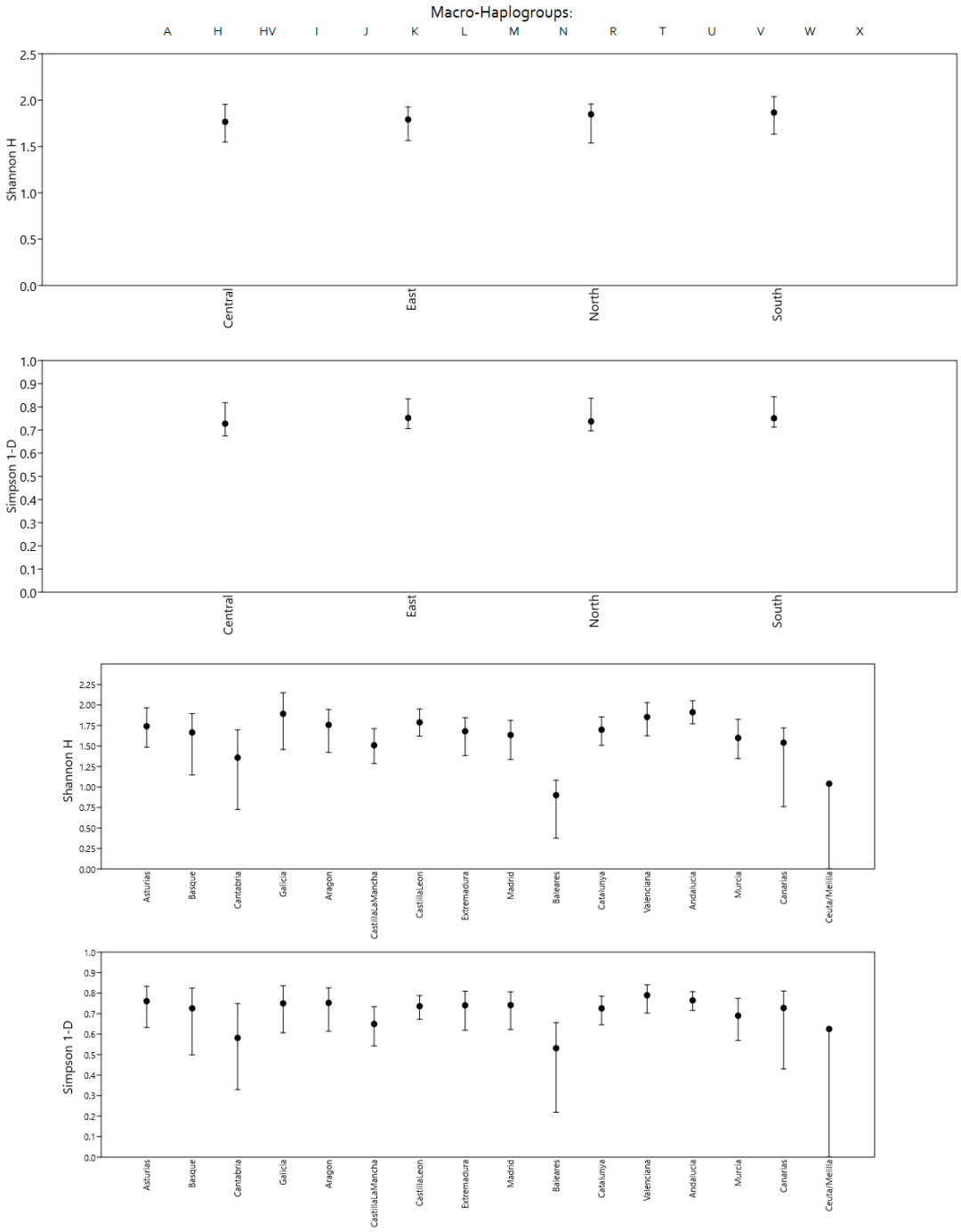


Figure S3: Haplogroup diversity indices for the Spanish cardinal divisions (North, Central, South East) and autonomous regions, and the, calculated with macro-haplogroups.

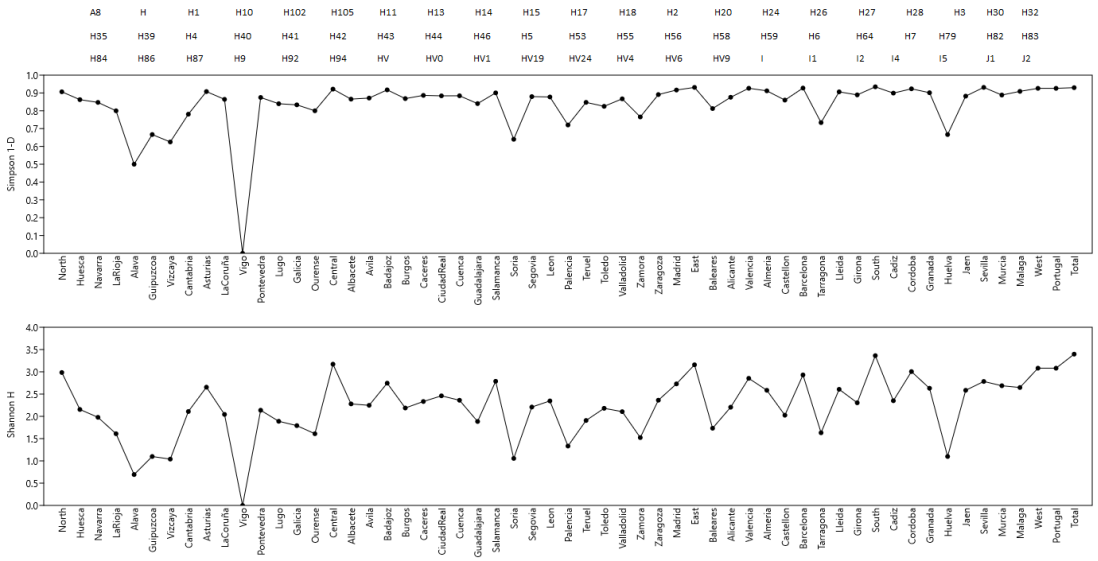


Figure S4: Haplogroup diversity indices for the Spanish provinces, cardinal division (North, Central, South East) and the country as a whole, calculated with sub-haplogroups.

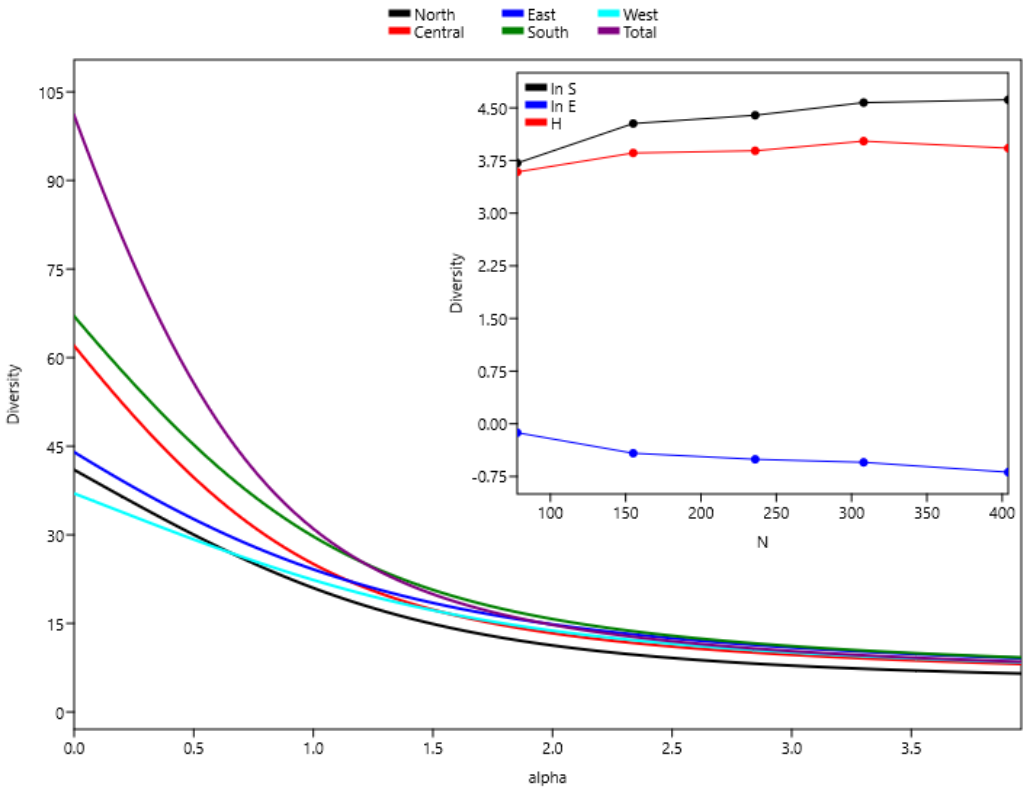


Figure S5: Diversity profile (Renyi Diversity Index Curve) for each geographical division of Spain (North, Central, South and East) and its corresponding SHE analysis, calculated using sub-haplogroups.

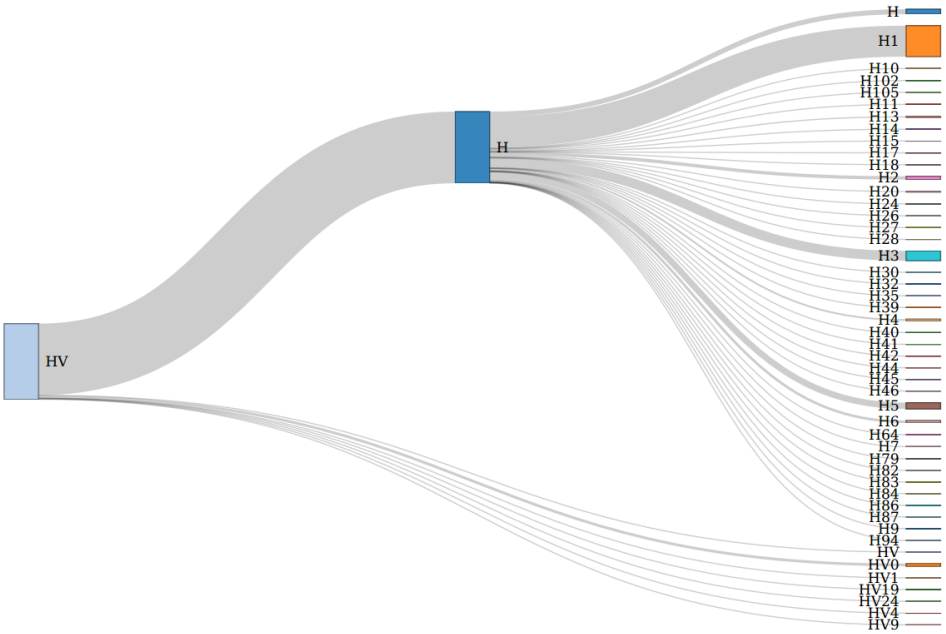


Figure S6: Phylogenetic relations of the haplogroups from the HV clade found in the Spanish dataset with the branches showing frequency. Plotted in R with sankeyNetwork function from networkD3 package.

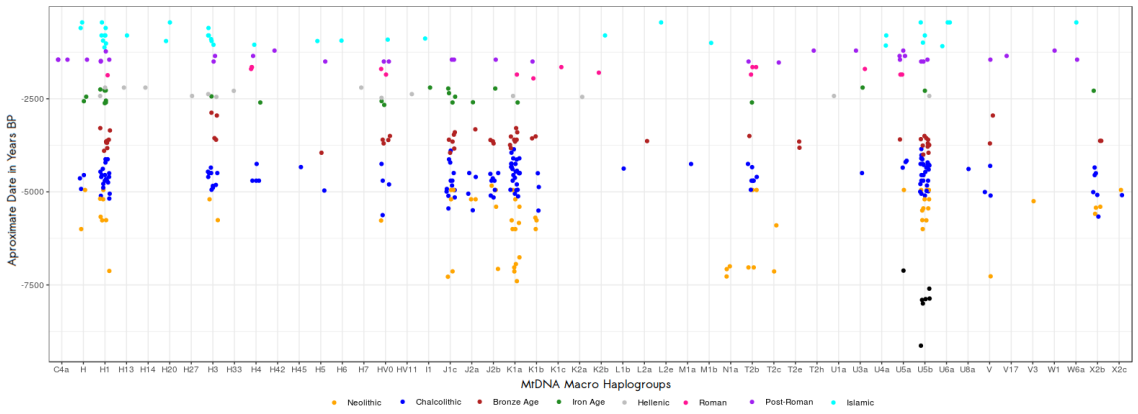


Figure S7: Timeline of published ancient mitochondrial genomes in Iberia sorted by more detailed sub-haplogroups.

Sample	*Haplogorup*	*Haplotype*
GOG05	H3	263 310 750 1438 4769 6776 8860 15326 16519
GOG06	K1b1a	73 152 263 310 513 750 1189 1438 1811 2706 3480 4769 5913 7028 8860 9055 9698 9962 10289 10373 10398 10550 11299 11467 11719 11923 12308 14167 14766 14798 15257 15326 15946 6224 16319 16463
GOG11	U5b1	73 150 263 310 750 1438 2706 3197 4769 5656 7028 7768 8860 9477 11467 11719 12308 12372 13617 13754 14182 14766 15326 16270 16519
GOG20	J1c1b	73 185 228 263 295 310 462 482 489 750 1438 2706 3010 3394 4216 4769 7028 7184 8860 10398 11251 11719 12612 13056 13708 14766 14798 15326 15452 16069 16126
GOG23	HV	195 263 302 310 750 1438 2706 4769 6935 7028 7692 8860 11101 12549 15326 16235 16298
GOG24	U4a1d	73 152 195 198 263 302 310 499 750 1438 1811 2706 4646 4769 5999 6047 7028 8209 8818 8860 11332 11467 11719 12308 12372 12937 14620 14766 15326 15693 16134 16356 16519
GOG25	L3d1	73 152 242 263 302 513 750 921 1438 2706 4769 5147 6197 6680 7028 7424 8618 8701 8860 9540 10398 10873 11719 11914 12705 13105 13135 13886 14110 14284 14766 15237 15301 15326 15826 16124 16223
GOG26	H1	263 310 750 1438 3010 4769 8753 8860 14353 15326 16261 16519
GOG34	HV34	263 310 750 1438 2706 4769 7028 8860 9801 10205 10920 15326 15514 16311
GOG35	H2a1e1a	263 310 575 750 751 951 8860 9052 15326 16124 16354
GOG38	K1a1	73 114 263 310 497 750 1189 1438 1811 2706 3480 4769 7028 8860 9055 9698 10398 10550 11299 11467 11719 11914 12308 12372 13708 14167 14766 14798 15326 16093 16224 16311 16519
GOG50	D4e1	73 263 489 750 1438 2706 3010 3316 4769 4883 5178 6366 7028 8414 8701 8860 9536 9540 10398 10400 10873 11215 11719 12705 14668 14766 14783 15043 15301 15326 16188 16223 16362
GOG56	H+7720	263 750 1438 4769 5558 7720 8860 15326 16519
GOG57	R0a4	58 64 150 263 750 1438 2351 2442 2706 3847 4769 7028 8860 9531 13135 13188 14124 14766 15326 16126 16362
GOG59	H5+152	152 263 456 750 1438 4769 8860 15326 16304 16519
GOG60	K1a+195	73 195 263 497 750 1438 1811 2706 3480 4769 7028 8860 9055 9698 10398 10550 11299 11467 11719 11926 12308 12372 14167 14766 14798 15326 16224 16311 16519

Table S1: Spanish mitochondrial genome haplotypes for the ancient samples sequenced in this thesis work.

Table S2: Relation of all the Spanish mitochondrial genomes sequenced in this project along with its haplogroup classification and region of origin.

ID	*Haplogroup*	*Region*	*Area*
ESP0001	H3	Andalucia	South
ESP0002	U5b1f1a	Navarra	North
ESP0003	H32	Madrid	Central
ESP0004	H20a1a	Catalunya	East
ESP0005	H3	Castilla_LaMancha	Central
ESP0006	V14	Castilla_LaMancha	Central
ESP0007	H1h1	Castilla_LaMancha	Central
ESP0008	H6a1a	Andalucia	South
ESP0009	H3	Castilla_LaMancha	Central
ESP0010	H5a3a1	Castilla_LaMancha	Central
ESP0011	H13a2b1	Andalucia	South
ESP0012	H10h	Andalucia	South
ESP0013	T2b25	Castilla_LaMancha	Central
ESP0014	H86	Extremadura	Central
ESP0015	H1	Castilla_LaMancha	Central
ESP0016	H1ap1	Catalunya	East
ESP0017	HV1a2	Andalucia	South
ESP0018	T2b3+151	Extremadura	Central
ESP0019	H1e2	Catalunya	East
ESP0020	U3a1	Castilla_y_Leon	Central
ESP0021	H	Castilla_LaMancha	Central
ESP0022	T2c1+146	Castilla_y_Leon	Central
ESP0023	H1b3	Catalunya	East
ESP0024	H3ab	Castilla_y_Leon	Central
ESP0025	K2b1a	Andalucia	South
ESP0026	H1	Extremadura	Central
ESP0027	H1aq	Andalucia	South
ESP0028	H1+16355	Castilla_LaMancha	Central
ESP0029	T2b	Catalunya	East
ESP0030	X2b+226	Madrid	Central
ESP0031	H1j1a1	Castilla_y_Leon	Central
ESP0032	U5b1+16189+@16192	Castilla_LaMancha	Central
ESP0033	H3+152	Aragon	North
ESP0034	H2a2a1	Catalunya	East
ESP0035	H1e2	Castilla_y_Leon	Central
ESP0036	U8b1b1	C_Valenciana	East
ESP0037	V	Aragon	Central
ESP0038	H1u2	Catalunya	East
ESP0039	T2b17a	Andalucia	South
ESP0040	H1ao1	Cantabria	North
ESP0041	H2a2a	Madrid	Central
ESP0042	H24	Castilla_LaMancha	Central
ESP0043	T2b3b	Andalucia	South
ESP0044	V	Extremadura	Central
ESP0045	H1e1a1	Asturias	North
ESP0046	H87	Madrid	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0047	H1e8	Castilla_y_Leon	Central
ESP0048	H1a3	Castilla_y_Leon	Central
ESP0049	H1+152	Andalucia	South
ESP0050	U8a2	Galicia	North
ESP0051	H	Castilla_y_Leon	Central
ESP0052	H3	Castilla_y_Leon	Central
ESP0053	J1+16193	Murcia	South
ESP0054	K1a+195	Castilla_y_Leon	Central
ESP0055	H3+152	Catalunya	East
ESP0056	J1c3	Castilla_y_Leon	Central
ESP0057	H94	Extremadura	Central
ESP0058	H11a	Catalunya	East
ESP0059	H4a1a	Castilla_y_Leon	Central
ESP0060	H1j1	Navarra	North
ESP0061	J1c2e2	Madrid	Central
ESP0062	H6a1b4	Catalunya	East
ESP0063	U5b1c	Castilla_y_Leon	Central
ESP0064	H1	Catalunya	East
ESP0065	J1d1b1	Catalunya	East
ESP0066	H5a1	Catalunya	East
ESP0067	K1a	Andalucia	South
ESP0068	U5b2b5	Castilla_y_Leon	Central
ESP0069	H24	Catalunya	East
ESP0070	X3a	Galicia	North
ESP0071	H2a2a1	Aragon	Central
ESP0072	K1a4a	Castilla_LaMancha	Central
ESP0073	T2c1d1a	Castilla_y_Leon	Central
ESP0074	X2m'n	Castilla_y_Leon	Central
ESP0075	X2b+226	Castilla_y_Leon	Central
ESP0076	K1a+195	Madrid	Central
ESP0077	U1b3	Madrid	Central
ESP0078	U1a1b	Castilla_LaMancha	Central
ESP0079	H1r	Aragon	Central
ESP0080	U2e2a1c	Andalucia	South
ESP0081	H5a3	Castilla_y_Leon	Central
ESP0082	J1c2	Madrid	Central
ESP0083	U5b1i	Castilla_LaMancha	Central
ESP0084	J1b	Galicia	North
ESP0085	U	Andalucia	South
ESP0086	H2a2b	Castilla_y_Leon	Central
ESP0087	H6a1b2	Andalucia	South
ESP0088	H2a5a1	Galicia	North
ESP0089	H6a1b2	Catalunya	East
ESP0090	H	Castilla_y_Leon	Central
ESP0091	U6a1a1	Andalucia	South
ESP0092	I3a	Canarias	Other
ESP0093	T2c1d+152	Extremadura	Central
ESP0094	H3ab	Extremadura	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0095	X2d2	Andalucia	South
ESP0096	J1c2h	Castilla_y_Leon	Central
ESP0097	U5b3	Baleares	East
ESP0098	U8b1b	C_Valenciana	East
ESP0099	H2a5a1	Galicia	North
ESP0100	H1j1a	Castilla_LaMancha	Central
ESP0101	K2a	Andalucia	South
ESP0102	H55	Andalucia	South
ESP0103	H	Galicia	North
ESP0104	T2c1a2	Canarias	Other
ESP0105	H1+152	Navarra	North
ESP0106	H20a1a	Andalucia	South
ESP0107	U2e1a1	Castilla_y_Leon	Central
ESP0108	V	Navarra	North
ESP0109	M1a3a	La_Rioja	Central
ESP0110	I2	Navarra	North
ESP0111	H1	Galicia	North
ESP0112	H3an	Galicia	North
ESP0113	T2b1	Castilla_LaMancha	Central
ESP0114	H35	C_Valenciana	East
ESP0115	V+150	Andalucia	South
ESP0116	H1j2a	Cantabria	North
ESP0117	U4c1	Canarias	Other
ESP0118	HV4a1a2	Castilla_y_Leon	Central
ESP0119	J1c3e2	Catalunya	East
ESP0120	J1c2e2	Andalucia	South
ESP0121	V	Andalucia	South
ESP0122	U3b2	Madrid	Central
ESP0123	J1c7	Castilla_LaMancha	Central
ESP0124	H102	Andalucia	South
ESP0125	H17a	La_Rioja	Central
ESP0126	U5b1c	Galicia	North
ESP0127	U4b1b1	C_Valenciana	East
ESP0128	J2b1a	C_Valenciana	East
ESP0129	M1a3b1	Galicia	North
ESP0130	H5g	Galicia	North
ESP0131	K1a4a1e	Castilla_LaMancha	Central
ESP0132	H	Extremadura	Central
ESP0133	H1ag1	C_Valenciana	East
ESP0134	J1c1g	Catalunya	East
ESP0135	V	Andalucia	South
ESP0136	H2a2a	Madrid	Central
ESP0137	J2a1a1a2	Castilla_y_Leon	Central
ESP0138	H	Galicia	North
ESP0139	J2b1a	Galicia	North
ESP0140	U2e1a1	Castilla_y_Leon	Central
ESP0141	H1j1a	Andalucia	South
ESP0142	T2h2	Cantabria	North

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0143	H1+16189	Castilla_y_Leon	Central
ESP0144	U5b1b1g	Aragon	Central
ESP0145	H1bv1	Murcia	South
ESP0146	U2e1a1	Castilla_y_Leon	Central
ESP0147	H1c5a	Navarra	North
ESP0148	R	Andalucia	South
ESP0149	U5b1g	Castilla_y_Leon	Central
ESP0150	X2b11	Extremadura	Central
ESP0151	H1j1	Extremadura	Central
ESP0152	U5a1a1	Aragon	Central
ESP0153	H4a	Castilla_y_Leon	Central
ESP0154	U5a1c2a1	Castilla_y_Leon	Central
ESP0155	X3a	Extremadura	Central
ESP0156	T2b33	Castilla_LaMancha	Central
ESP0157	H1bf1	Andalucia	South
ESP0158	H1	La_Rioja	Central
ESP0159	X2m'n	Andalucia	South
ESP0160	J1c2e2	Castilla_y_Leon	Central
ESP0161	L2a1'2'3'4	Castilla_LaMancha	Central
ESP0162	V7	Asturias	North
ESP0163	H1ap1	Andalucia	South
ESP0164	H1c3	Castilla_y_Leon	Central
ESP0165	H5a1b	Andalucia	South
ESP0166	J2b1a1	Castilla_LaMancha	Central
ESP0167	L3e4a	Canarias	Other
ESP0168	H7	Castilla_LaMancha	Central
ESP0169	K1a+195	Catalunya	East
ESP0170	X2i+@225	Castilla_y_Leon	Central
ESP0171	H53	Canarias	Other
ESP0172	H1j1a	Andalucia	South
ESP0173	U6a1a1	Andalucia	South
ESP0174	H1av1	Castilla_LaMancha	Central
ESP0175	H11b1	Castilla_LaMancha	Central
ESP0176	X2b+226	Aragon	Central
ESP0177	H5a	Andalucia	South
ESP0178	V	Basque_Country	North
ESP0179	H1bw	Castilla_y_Leon	Central
ESP0180	K2b1a	Aragon	North
ESP0181	J1c1	Andalucia	South
ESP0182	H1j1	Andalucia	South
ESP0183	H1ao	Castilla_y_Leon	Central
ESP0184	H6a1a	C_Valenciana	East
ESP0185	H2a2a	Andalucia	South
ESP0186	H4a1	Castilla_y_Leon	Central
ESP0187	U5b2a1a2	Andalucia	South
ESP0188	H1t1a	Castilla_LaMancha	Central
ESP0189	H9a	Catalunya	East
ESP0190	H3	Castilla_LaMancha	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0191	L3d1b3a	Andalucia	South
ESP0192	H1bd	Castilla_LaMancha	Central
ESP0193	X2b+226	Andalucia	South
ESP0194	K1a+195	Catalunya	East
ESP0195	H1e1a6	Murcia	South
ESP0196	V	C_Valenciana	East
ESP0197	T1a	Catalunya	East
ESP0198	H3+152	Aragon	Central
ESP0199	U3a1	Catalunya	East
ESP0200	H4a1a	Aragon	North
ESP0201	U3a1	Murcia	South
ESP0202	H5a	Murcia	South
ESP0203	K1a4a1e	Castilla_y_Leon	Central
ESP0204	H+152	Baleares	East
ESP0205	H1at	Asturias	North
ESP0206	H20c	Aragon	North
ESP0207	K1b1a1	Extremadura	Central
ESP0208	H1+16189	Extremadura	Central
ESP0209	J2b1a+16311	Castilla_LaMancha	Central
ESP0210	H3at1	Extremadura	Central
ESP0211	H1j1a	Castilla_LaMancha	Central
ESP0213	H1aa	Castilla_y_Leon	Central
ESP0214	H1q	Castilla_y_Leon	Central
ESP0215	H1	Murcia	South
ESP0216	H1e1a1	Andalucia	South
ESP0217	K1b1a1c	Murcia	South
ESP0218	H4a1	Andalucia	South
ESP0219	K1a4a1	Asturias	North
ESP0220	T1a2	Catalunya	East
ESP0221	HV0+195	Catalunya	East
ESP0222	H4a1a4a	Castilla_y_Leon	Central
ESP0223	H5a1d	Castilla_LaMancha	Central
ESP0224	H1+152	Castilla_y_Leon	Central
ESP0225	L2a1c1	Murcia	South
ESP0226	H1ae3	Murcia	South
ESP0227	H1c9a	Castilla_y_Leon	Central
ESP0229	J1c8b	Castilla_y_Leon	Central
ESP0230	H5	Catalunya	East
ESP0231	H1j1	Castilla_LaMancha	Central
ESP0233	H1av1	Extremadura	Central
ESP0234	X2b6a	Cantabria	North
ESP0235	J1c2	Catalunya	East
ESP0236	U5a2c1	Andalucia	South
ESP0237	H17	Castilla_y_Leon	Central
ESP0238	U6a7c1	Castilla_y_Leon	Central
ESP0239	J1c2	Murcia	South
ESP0240	H1c	Andalucia	South
ESP0241	H	Castilla_y_Leon	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0242	H6a1b2	Andalucia	South
ESP0243	U5a1i	Asturias	North
ESP0244	T1a1	Andalucia	South
ESP0245	H3+152	Catalunya	East
ESP0246	H3ap	Andalucia	South
ESP0247	H40a	Murcia	South
ESP0248	T2c1d+152	Castilla_y_Leon	Central
ESP0249	H10+(16093)	Castilla_y_Leon	Central
ESP0250	H3	Murcia	South
ESP0251	J1c3	Murcia	South
ESP0252	T2a1a	Catalunya	East
ESP0253	H	Castilla_y_Leon	Central
ESP0254	U6b1a	Murcia	South
ESP0255	V	Aragon	Central
ESP0256	H+152	Asturias	North
ESP0257	H1j	Cantabria	North
ESP0258	T2b3b	Asturias	North
ESP0259	K2b1a	Catalunya	East
ESP0260	X1c	Andalucia	South
ESP0261	H1e1a7	Murcia	South
ESP0262	J1c2	Castilla_y_Leon	Central
ESP0263	J1c3a	Catalunya	East
ESP0264	K1a+195	Andalucia	South
ESP0265	H1av1	Aragon	Central
ESP0266	J2b1a	Andalucia	South
ESP0267	H5c	Andalucia	South
ESP0268	J2b1a1	Castilla_y_Leon	Central
ESP0269	H2a1	Catalunya	East
ESP0270	T1a1p	Aragon	Central
ESP0271	U2e1c1	Castilla_y_Leon	Central
ESP0272	T2b3+151	Castilla_y_Leon	Central
ESP0273	H2a1	Andalucia	South
ESP0274	K1a	Castilla_y_Leon	Central
ESP0275	K2b1a	Andalucia	South
ESP0276	T2b21	Andalucia	South
ESP0277	U5b2b	Castilla_y_Leon	Central
ESP0278	I5a2	Castilla_LaMancha	Central
ESP0279	K1a4a1	Andalucia	South
ESP0280	H3ah	Catalunya	East
ESP0281	U6a3a2a	Catalunya	East
ESP0282	H+152	Castilla_y_Leon	Central
ESP0283	U5b1i	Catalunya	East
ESP0284	I4a	Andalucia	South
ESP0285	H1c	Murcia	South
ESP0286	K2a3	Andalucia	South
ESP0287	H3	C_Valenciana	East
ESP0288	H6a1b2	Catalunya	East
ESP0289	U5b2b3	Castilla_y_Leon	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0290	H3ar	Extremadura	Central
ESP0291	H3w	Castilla_LaMancha	Central
ESP0292	H1b	Murcia	South
ESP0293	H5a1	Catalunya	East
ESP0294	H32	Castilla_LaMancha	Central
ESP0295	H1+16311	Cantabria	North
ESP0296	H1+16189	Cantabria	North
ESP0297	L3e5a	Andalucia	South
ESP0298	U5b1c	Extremadura	Central
ESP0299	J1c7	C_Valenciana	East
ESP0300	V22	Aragon	Central
ESP0301	U4a	Catalunya	East
ESP0302	U5b3	Andalucia	South
ESP0303	K1a4a1	Madrid	Central
ESP0304	X2b+226	Andalucia	South
ESP0305	H1	Castilla_y_Leon	Central
ESP0306	J1c3f	Galicia	North
ESP0307	X3a	Extremadura	Central
ESP0308	I1a1	Catalunya	East
ESP0309	H3	Murcia	South
ESP0310	H1ao1	Galicia	North
ESP0311	H2a2a1	Madrid	Central
ESP0312	K1a1b1g	Aragon	Central
ESP0313	L2a1c3a	Murcia	South
ESP0314	H5a	Murcia	South
ESP0315	H2a1	Andalucia	South
ESP0316	T1a1	Castilla_y_Leon	Central
ESP0317	T2	Catalunya	East
ESP0318	H6a1b2	Andalucia	South
ESP0319	K1b1a	Catalunya	East
ESP0320	H18	Madrid	Central
ESP0321	J1c5c1	Castilla_LaMancha	Central
ESP0322	H56	Andalucia	South
ESP0323	J1c3e2	C_Valenciana	East
ESP0324	H1at	Aragon	North
ESP0325	J1d6	Extremadura	Central
ESP0326	H28a	Cantabria	North
ESP0327	U5b2a1a2	Cantabria	North
ESP0328	HV0f	Catalunya	East
ESP0329	H26a1	Asturias	North
ESP0330	HV0h	C_Valenciana	East
ESP0331	L3f1b	Asturias	North
ESP0332	H2a5a	Andalucia	South
ESP0333	U4a1a	Galicia	North
ESP0334	U5b1	Catalunya	East
ESP0335	H1b	Castilla_y_Leon	Central
ESP0336	X2b7	C_Valenciana	East
ESP0337	J1c2c1	Extremadura	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0338	H6a1b	Andalucia	South
ESP0339	K1a4a1	Castilla_LaMancha	Central
ESP0340	H3g1	Murcia	South
ESP0341	H1bf1	Murcia	South
ESP0342	U5b1+16189+@16192	Asturias	North
ESP0343	N1b1a	Catalunya	East
ESP0344	H1e1a6	Murcia	South
ESP0345	V	C_Valenciana	East
ESP0347	H1j1a	Andalucia	South
ESP0348	H3m	Navarra	North
ESP0349	T2b33	Andalucia	South
ESP0350	H11a2a	Andalucia	South
ESP0351	U6a3b	Andalucia	South
ESP0352	K1c1	Castilla_y_Leon	Central
ESP0353	H1c1	Andalucia	South
ESP0354	H1b	Catalunya	East
ESP0355	H11a2	Madrid	Central
ESP0356	H3k	Madrid	Central
ESP0357	H1ai	Castilla_y_Leon	Central
ESP0358	M1b2	Castilla_LaMancha	Central
ESP0359	U8b1b1	Galicia	North
ESP0360	U5b1f1a	Navarra	North
ESP0361	H1t	Andalucia	South
ESP0362	H5d	Extremadura	Central
ESP0363	K2b1a	Catalunya	East
ESP0364	H9a	Murcia	South
ESP0365	U5b3	Catalunya	East
ESP0366	H1a1	Cantabria	North
ESP0367	J1b2	Andalucia	South
ESP0368	H3av	Murcia	South
ESP0369	H	Andalucia	South
ESP0370	I4a	C_Valenciana	East
ESP0371	H2a2b1	Extremadura	Central
ESP0372	J2b1g	Castilla_y_Leon	Central
ESP0373	J2b1a	Catalunya	East
ESP0374	H41a	Murcia	South
ESP0375	H3g1	Murcia	South
ESP0376	H4a1	Castilla_LaMancha	Central
ESP0377	H1	Asturias	North
ESP0378	H1br	Extremadura	Central
ESP0379	HV0+195	Castilla_y_Leon	Central
ESP0380	H3w	C_Valenciana	East
ESP0381	U5b1+16189+@16192	Basque_Country	North
ESP0382	H1	Catalunya	East
ESP0384	M1a2a	C_Valenciana	East
ESP0385	T2c1d1	Castilla_y_Leon	Central
ESP0386	U3a1c1	Castilla_y_Leon	Central
ESP0387	I2	Andalucia	South

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0388	H5g	Castilla_y_Leon	Central
ESP0389	H105a	Castilla_LaMancha	Central
ESP0390	H59a	Castilla_LaMancha	Central
ESP0391	K1a+195	Galicia	North
ESP0392	V2b	Castilla_y_Leon	Central
ESP0393	K1a+195	Andalucia	South
ESP0394	U5b1e	Andalucia	South
ESP0395	H1e1a	C_Valenciana	East
ESP0396	J1c9	Castilla_y_Leon	Central
ESP0397	U3a1	Murcia	South
ESP0398	H1aa	Castilla_y_Leon	Central
ESP0399	H1g1	Castilla_LaMancha	Central
ESP0400	H1c3	Andalucia	South
ESP0401	H6a1a	Castilla_LaMancha	Central
ESP0402	W5	Andalucia	South
ESP0403	H82	Andalucia	South
ESP0404	I5a1b	Castilla_y_Leon	Central
ESP0405	K1a+195	Castilla_y_Leon	Central
ESP0406	K1a1a	Andalucia	South
ESP0407	H3h6	Andalucia	South
ESP0408	H3an	Galicia	North
ESP0409	U5b3	Castilla_LaMancha	Central
ESP0410	H	Castilla_y_Leon	Central
ESP0411	T2	Catalunya	East
ESP0412	V10b	C_Valenciana	East
ESP0413	L4b2b	Andalucia	South
ESP0414	K1a3a	Catalunya	East
ESP0415	H44b	Castilla_y_Leon	Central
ESP0416	J2b1	Castilla_y_Leon	Central
ESP0417	H3	Catalunya	East
ESP0418	T2e	C_Valenciana	East
ESP0419	H1e1a1	Andalucia	South
ESP0420	K1a+195	Andalucia	South
ESP0421	H46	Castilla_y_Leon	Central
ESP0422	K1b2b	Aragon	Central
ESP0423	H3c2a	Catalunya	East
ESP0424	H1bf	C_Valenciana	East
ESP0425	H1x	Andalucia	South
ESP0426	T2b3+151	Africa	Other
ESP0427	U5b2b5	Castilla_y_Leon	Central
ESP0428	W	Castilla_y_Leon	Central
ESP0429	H7b	Andalucia	South
ESP0430	H1c	Galicia	North
ESP0431	H3	Andalucia	South
ESP0432	H+195	Castilla_y_Leon	Central
ESP0433	U5a1a1	Andalucia	South
ESP0434	U5a1a2a	Castilla_y_Leon	Central
ESP0435	T1a13	Castilla_LaMancha	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0436	H1e1a	Andalucia	South
ESP0437	T2f	Castilla_y_Leon	Central
ESP0438	U5b2b3	Andalucia	South
ESP0439	H1ar	Castilla_y_Leon	Central
ESP0440	H1cf	Asturias	North
ESP0441	H7d3	Castilla_y_Leon	Central
ESP0442	T2b+152	Andalucia	South
ESP0443	J2a1a1e	Andalucia	South
ESP0444	K2b1a	Castilla_y_Leon	Central
ESP0445	V7a	Extremadura	Central
ESP0446	H1j1	Castilla_LaMancha	Central
ESP0447	T2b33	Andalucia	South
ESP0448	H7	Andalucia	South
ESP0449	H3y	Madrid	Central
ESP0450	H3ao1	Madrid	Central
ESP0451	H3c2a	Castilla_y_Leon	Central
ESP0452	K1a+150	Aragon	North
ESP0453	H24a1	Castilla_LaMancha	Central
ESP0454	H1j1	Extremadura	Central
ESP0455	H3c	Castilla_y_Leon	Central
ESP0456	V22	C_Valenciana	East
ESP0457	H3c	Castilla_y_Leon	Central
ESP0458	H+152	Castilla_y_Leon	Central
ESP0459	H1j1a	Andalucia	South
ESP0460	V	Madrid	Central
ESP0461	I1a1	Cantabria	North
ESP0462	U5a1a1	Andalucia	South
ESP0463	H1c3	Castilla_LaMancha	Central
ESP0464	H79	Asturias	North
ESP0465	H1r	Castilla_LaMancha	Central
ESP0466	X2b11	Catalunya	East
ESP0467	H	Castilla_y_Leon	Central
ESP0468	J2a1a1e	Andalucia	South
ESP0469	J1c5a	Castilla_LaMancha	Central
ESP0470	H	Andalucia	South
ESP0471	H3	Castilla_LaMancha	Central
ESP0472	H+152	Canarias	Other
ESP0474	K1b2b	Extremadura	Central
ESP0475	W1	Castilla_y_Leon	Central
ESP0476	V25	Andalucia	South
ESP0477	H1b1e	Galicia	North
ESP0478	U5a2d1	Murcia	South
ESP0479	R1a1a1	Andalucia	South
ESP0480	HV24	Andalucia	South
ESP0481	U3a1	Castilla_y_Leon	Central
ESP0482	H5m	Aragon	Central
ESP0483	T1a	Extremadura	Central
ESP0484	K1a13a	Castilla_LaMancha	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0485	V+150	Castilla_y_Leon	Central
ESP0486	U5a1a1	Castilla_y_Leon	Central
ESP0487	J1c1b	Cantabria	North
ESP0488	T2	Andalucia	South
ESP0489	V3a	Andalucia	South
ESP0490	H3k	Andalucia	South
ESP0491	T2b33	Andalucia	South
ESP0492	H1+16189	Catalunya	East
ESP0493	H7h	Extremadura	Central
ESP0494	K2b1a1a	Castilla_y_Leon	Central
ESP0495	HV0f	Catalunya	East
ESP0496	H2a1	Extremadura	Central
ESP0497	T2	Andalucia	South
ESP0498	H1+16189	Andalucia	South
ESP0499	V3a	Extremadura	Central
ESP0500	H1c	Castilla_LaMancha	Central
ESP0501	T2b	Castilla_LaMancha	Central
ESP0502	U5a2c	C_Valenciana	East
ESP0503	H1e1a7	Andalucia	South
ESP0504	H1j1a	Madrid	Central
ESP0505	J1c2e2	Basque_Country	North
ESP0506	U5b2a1b	Madrid	Central
ESP0507	H3k1	Andalucia	South
ESP0508	H3s	Catalunya	East
ESP0509	H1c	Catalunya	East
ESP0510	N1b1a7	Galicia	North
ESP0511	H1r	Castilla_LaMancha	Central
ESP0512	J1b1a1+146	Castilla_y_Leon	Central
ESP0513	H1ak1	Catalunya	East
ESP0514	T2a1a	Castilla_LaMancha	Central
ESP0515	H	Catalunya	East
ESP0516	H5	Castilla_LaMancha	Central
ESP0517	H5a	Andalucia	South
ESP0518	M1b1a	Castilla_y_Leon	Central
ESP0519	I1a1	Catalunya	East
ESP0520	T2n	Navarra	North
ESP0521	T2b4	Catalunya	East
ESP0522	H5a1	Castilla_y_Leon	Central
ESP0523	K1a2a	C_Valenciana	East
ESP0524	U5b1+16189+@16192	Madrid	Central
ESP0525	H40a	Andalucia	South
ESP0526	H1e5a	Catalunya	East
ESP0527	H	Galicia	North
ESP0528	T1a1	Galicia	North
ESP0529	K1c1	Andalucia	South
ESP0530	H3	Murcia	South
ESP0531	T2b3+151	Castilla_y_Leon	Central
ESP0532	N1b1a2	Andalucia	South

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0533	J1c1g	Catalunya	East
ESP0534	U5b1c	Andalucia	South
ESP0535	H1e1c	Castilla_y_Leon	Central
ESP0536	U6a7a1b	Andalucia	South
ESP0537	I2a2	Galicia	North
ESP0538	H7	Catalunya	East
ESP0539	T2n	La_Rioja	Central
ESP0540	U5a1a1	Madrid	Central
ESP0541	T1b	Castilla_LaMancha	Central
ESP0542	T1a1p	C_Valenciana	East
ESP0543	H13a1a1	Catalunya	East
ESP0544	HV0h	Andalucia	South
ESP0545	H1	Catalunya	East
ESP0546	T2b	Castilla_y_Leon	Central
ESP0547	H30a	Murcia	South
ESP0548	H3c	Madrid	Central
ESP0549	U5b1e	Castilla_LaMancha	Central
ESP0550	H1ba	Aragon	Central
ESP0551	I2a3	Cantabria	North
ESP0552	H1aa1	Castilla_LaMancha	Central
ESP0553	T1a1	Asturias	North
ESP0554	H42a	Andalucia	South
ESP0555	U4c1	Murcia	South
ESP0556	H4a1a1a3	Catalunya	East
ESP0557	H11a2	Catalunya	East
ESP0558	H1	Castilla_y_Leon	Central
ESP0559	H	Castilla_y_Leon	Central
ESP0560	HV0a	Catalunya	East
ESP0561	H1	Catalunya	East
ESP0562	H10e	Andalucia	South
ESP0563	H3ah	Andalucia	South
ESP0564	H2a1	Andalucia	South
ESP0565	M1a1i	Andalucia	South
ESP0566	K1a1a	Castilla_y_Leon	Central
ESP0567	H1bm	Catalunya	East
ESP0568	J1b2	Andalucia	South
ESP0569	T2b	Castilla_y_Leon	Central
ESP0570	H2a2b	Castilla_LaMancha	Central
ESP0571	H1	Castilla_LaMancha	Central
ESP0572	H5g	Madrid	Central
ESP0573	H1e1a	Extremadura	Central
ESP0574	H1?	Catalunya	East
ESP0575	J1c2m	Galicia	North
ESP0576	H10+(16093)	Andalucia	South
ESP0577	X2+225+@153	Cantabria	North
ESP0578	T2a1a	Murcia	South
ESP0579	H1a	Cantabria	North
ESP0580	K1a13a	Cantabria	North

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0581	H6a1b2	Catalunya	East
ESP0582	H1aa1	Castilla_LaMancha	Central
ESP0583	H1ag1	Catalunya	East
ESP0584	U6a6b1	Catalunya	East
ESP0585	T2c1d1a	Andalucia	South
ESP0586	H6a1b2	Catalunya	East
ESP0587	U5b2c	Andalucia	South
ESP0588	J1c2e2	Asturias	North
ESP0589	K1a4a1	Castilla_y_Leon	Central
ESP0590	K1a+195	C_Valenciana	East
ESP0591	HV9	Catalunya	East
ESP0592	T1a2	Andalucia	South
ESP0593	J1c3	Catalunya	East
ESP0594	K1a4a	Andalucia	South
ESP0596	H1e2c	Castilla_LaMancha	Central
ESP0597	U5b2b4a	Extremadura	Central
ESP0598	J1c5a	Extremadura	Central
ESP0599	H1e1a6	Andalucia	South
ESP0600	H3k	Andalucia	South
ESP0601	H84	Catalunya	East
ESP0602	H	Cantabria	North
ESP0603	H3	Castilla_LaMancha	Central
ESP0604	H3+152	Catalunya	East
ESP0605	V+@72	Galicia	North
ESP0606	H1e1a8	C_Valenciana	East
ESP0607	J1c2	Canarias	Other
ESP0608	J1c1f	Madrid	Central
ESP0609	K1a1a	Catalunya	East
ESP0610	T2b3+151	Extremadura	Central
ESP0611	H1e1a6	Castilla_y_Leon	Central
ESP0612	J1c3	Madrid	Central
ESP0613	H+152	Baleares	East
ESP0614	V+150	Andalucia	South
ESP0615	U8a1a1a	Andalucia	South
ESP0616	H1j1a	Castilla_y_Leon	Central
ESP0617	K1a4a1	Aragon	North
ESP0618	H1c	Castilla_y_Leon	Central
ESP0619	X2c1	Extremadura	Central
ESP0621	T2a1b1a2	Castilla_y_Leon	Central
ESP0622	U4b1a4	Castilla_y_Leon	Central
ESP0623	H1ak2	Extremadura	Central
ESP0624	H5+709	Catalunya	East
ESP0625	H	Andalucia	South
ESP0626	U2e1	Baleares	East
ESP0627	V3a	Castilla_y_Leon	Central
ESP0628	W1	Galicia	North
ESP0629	H1e1a1	Castilla_LaMancha	Central
ESP0630	H3av	Murcia	South

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0631	R0a2	Castilla_y_Leon	Central
ESP0632	U4a3a	Madrid	Central
ESP0633	T2b33	Andalucia	South
ESP0634	H1bw	Galicia	North
ESP0635	H3c2	Africa	Other
ESP0636	K1a3a	Galicia	North
ESP0637	H1e1a2	Basque_Country	North
ESP0638	U5b1i	Madrid	Central
ESP0639	T2c1d+152	Asturias	North
ESP0640	K1a1a	Castilla_y_Leon	Central
ESP0641	U2e1a1	Extremadura	Central
ESP0643	U4a3a	Castilla_LaMancha	Central
ESP0644	X2b+226	Andalucia	South
ESP0645	J1c1b	Aragon	Central
ESP0646	HV0+195	La_Rioja	Central
ESP0647	H45	Canarias	Other
ESP0648	T2a3	Andalucia	South
ESP0649	H3n	Castilla_y_Leon	Central
ESP0650	H5a	Castilla_y_Leon	Central
ESP0651	T2b21	Aragon	Central
ESP0652	H	Aragon	Central
ESP0653	H1t1a	Cantabria	North
ESP0654	J1b1a1	Galicia	North
ESP0655	HV0g	Catalunya	East
ESP0656	H3t	Castilla_LaMancha	Central
ESP0657	V1a	Catalunya	East
ESP0658	V+@72	Castilla_y_Leon	Central
ESP0659	T1a2	Andalucia	South
ESP0660	H	Galicia	North
ESP0661	U3a1	Andalucia	South
ESP0662	T2b3b	Murcia	South
ESP0663	U5b3	C_Valenciana	East
ESP0664	H58a	Castilla_y_Leon	Central
ESP0665	J1c5c1	Aragon	North
ESP0666	T2a1a	Andalucia	South
ESP0667	HV0d	Madrid	Central
ESP0668	J2b1a	Galicia	North
ESP0669	T2e1	Extremadura	Central
ESP0670	H1	Baleares	East
ESP0671	H1cf	Canarias	Other
ESP0672	H42a1	Extremadura	Central
ESP0673	J1c5c1	Castilla_y_Leon	Central
ESP0674	H1c	Madrid	Central
ESP0675	K1a17	Aragon	Central
ESP0676	T1a1+@152	Extremadura	Central
ESP0677	U2e1	Andalucia	South
ESP0678	U5a1+@16192	Castilla_y_Leon	Central
ESP0679	K1a1a	C_Valenciana	East

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0680	H1c	Andalucia	South
ESP0681	H4a1a	Catalunya	East
ESP0682	HV24	Galicia	North
ESP0683	U4a	Extremadura	Central
ESP0684	H1c	Murcia	South
ESP0685	H+152	Extremadura	Central
ESP0686	U4a2f	Catalunya	East
ESP0687	U2e1e	Extremadura	Central
ESP0688	T2c1f	Basque_Country	North
ESP0689	H15a1a1	Andalucia	South
ESP0690	H1u2	Catalunya	East
ESP0691	T2b7a3	Madrid	Central
ESP0692	X2m'n	Galicia	North
ESP0693	H1	C_Valenciana	East
ESP0694	H5a3a1	Andalucia	South
ESP0695	U5b1+16189+@16192	Andalucia	South
ESP0696	U1a1a3	Castilla_y_Leon	Central
ESP0697	J1c1	Castilla_LaMancha	Central
ESP0698	T2b	Madrid	Central
ESP0699	T2b5	C_Valenciana	East
ESP0700	U5b2a1a2	Catalunya	East
ESP0701	H1+152	Asturias	North
ESP0702	L1b1a+189	Castilla_LaMancha	Central
ESP0703	I5a1b	Castilla_LaMancha	Central
ESP0704	H14b	Andalucia	South
ESP0705	J1c	Murcia	South
ESP0706	H6a1a	C_Valenciana	East
ESP0707	U5a2a1	C_Valenciana	East
ESP0708	H5a+152	Murcia	South
ESP0710	U5b1f1a	Andalucia	South
ESP0712	I2	Asturias	North
ESP0713	U5b1g	Catalunya	East
ESP0714	L3b1a+@16124	Castilla_y_Leon	Central
ESP0715	U5b1f1a	Murcia	South
ESP0716	U6a3a1a	Aragon	Central
ESP0717	H1aq	Castilla_y_Leon	Central
ESP0718	H20a1a	Murcia	South
ESP0719	HV9	Catalunya	East
ESP0720	H3s	Castilla_LaMancha	Central
ESP0721	HV0b	Murcia	South
ESP0722	K1b1a2	Andalucia	South
ESP0723	K1c1	Andalucia	South
ESP0724	U2e1a1	Catalunya	East
ESP0725	J1c1c	Castilla_y_Leon	Central
ESP0726	K1a4a1e	Murcia	South
ESP0727	T2c1d+152	C_Valenciana	East
ESP0728	H1+16189	Castilla_y_Leon	Central
ESP0729	T2c1d+152	C_Valenciana	East

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0730	H1	Castilla_LaMancha	Central
ESP0732	U6b	Andalucia	South
ESP0733	H	Andalucia	South
ESP0735	T2	Catalunya	East
ESP0736	HV9	Andalucia	South
ESP0737	H1af1	Castilla_y_Leon	Central
ESP0738	H3+16189	Andalucia	South
ESP0739	H6a1a7	Castilla_LaMancha	Central
ESP0740	J1c5a	Canarias	Other
ESP0741	U5b1b1g	Andalucia	South
ESP0742	T2b25	Aragon	North
ESP0743	HV0b	Madrid	Central
ESP0744	J1c2	Castilla_y_Leon	Central
ESP0745	I	Castilla_y_Leon	Central
ESP0746	H1	Andalucia	South
ESP0747	H1e6	Catalunya	East
ESP0748	M1a3b1	Murcia	South
ESP0749	U6a7a1b	Andalucia	South
ESP0750	J1c1b1	Murcia	South
ESP0751	H3	Castilla_y_Leon	Central
ESP0752	H1j2a	Andalucia	South
ESP0753	HV0a	C_Valenciana	East
ESP0754	H1cf	Catalunya	East
ESP0755	J2a1a1	Castilla_LaMancha	Central
ESP0756	H24	Castilla_y_Leon	Central
ESP0757	R0a1a	Catalunya	East
ESP0758	H1j1a	Madrid	Central
ESP0759	J2a1a1e	Catalunya	East
ESP0760	U2e1a1	Asturias	North
ESP0761	R0a	Castilla_y_Leon	Central
ESP0762	J1c5	Galicia	North
ESP0763	H3k	Andalucia	South
ESP0764	J1d6	Navarra	North
ESP0765	H3ah	Andalucia	South
ESP0767	V+@72	Andalucia	South
ESP0768	H17	Madrid	Central
ESP0769	V18a	Murcia	South
ESP0770	H1j	Murcia	South
ESP0771	H13b1+200	Galicia	North
ESP0772	M1b2	Cantabria	North
ESP0773	H1t1a	Andalucia	South
ESP0774	K1a1b2a1	Catalunya	East
ESP0775	H1	Castilla_y_Leon	Central
ESP0776	T2b33	Extremadura	Central
ESP0777	T2b25	Catalunya	East
ESP0778	K2b1a1a	Andalucia	South
ESP0779	H53	Andalucia	South
ESP0780	H1c4b	Murcia	South

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0781	U4d3	C_Valenciana	East
ESP0782	H27	Andalucia	South
ESP0783	T2b21	Aragon	Central
ESP0784	H1bx	Asturias	North
ESP0785	U5b1c1a1	Andalucia	South
ESP0786	H2a2b	Castilla_y_Leon	Central
ESP0787	H20a1a	Africa	Other
ESP0788	H5a1a	Basque_Country	North
ESP0789	T2b4	Catalunya	East
ESP0790	J1c1b1a	Madrid	Central
ESP0791	J2b1g	Castilla_LaMancha	Central
ESP0792	H1e1a6	Galicia	North
ESP0793	H1ah	Extremadura	Central
ESP0794	T2b3+151	C_Valenciana	East
ESP0795	U5a2+16294	Extremadura	Central
ESP0796	K1a4a1	Catalunya	East
ESP0797	H3c	Castilla_y_Leon	Central
ESP0798	H1j	Castilla_y_Leon	Central
ESP0799	H1j2a	Cantabria	North
ESP0800	HV0+195	Andalucia	South
ESP0801	H1c	Andalucia	South
ESP0802	U5a1+@16192	Castilla_y_Leon	Central
ESP0803	J2b1a2	Catalunya	East
ESP0804	H	Basque_Country	North
ESP0805	H1ag1	Galicia	North
ESP0806	H40	Galicia	North
ESP0807	T2b33	Basque_Country	North
ESP0808	U4a	Murcia	South
ESP0809	H1t	Castilla_LaMancha	Central
ESP0810	K1a4	Catalunya	East
ESP0811	H105a	Castilla_y_Leon	Central
ESP0812	U5a2c	C_Valenciana	East
ESP0813	L1b1a6	Andalucia	South
ESP0814	T2b	Castilla_LaMancha	Central
ESP0815	H1aq	Catalunya	East
ESP0816	H20a1a	Extremadura	Central
ESP0817	T2a1a	Castilla_LaMancha	Central
ESP0818	U5a1b1e	C_Valenciana	East
ESP0819	H27	Baleares	East
ESP0820	H6a1b2	Andalucia	South
ESP0821	H56	Catalunya	East
ESP0822	J1c1b	Castilla_LaMancha	Central
ESP0823	I1c	Catalunya	East
ESP0824	T2b+152	Murcia	South
ESP0825	J2b1a	Murcia	South
ESP0826	K1a+195	Asturias	North
ESP0827	H1r	Aragon	North
ESP0828	H14a2	Castilla_LaMancha	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0830	H1v	Catalunya	East
ESP0831	H1e6	Andalucia	South
ESP0832	H1t	C_Valenciana	East
ESP0833	H1	Murcia	South
ESP0834	H4a1a	Aragon	North
ESP0835	H2a2b	Andalucia	South
ESP0836	T2b11	Murcia	South
ESP0837	H5a1f	Castilla_y_Leon	Central
ESP0838	H3au	Castilla_y_Leon	Central
ESP0839	U3c	Castilla_LaMancha	Central
ESP0840	T2b4	Castilla_LaMancha	Central
ESP0841	J1c2e2	Murcia	South
ESP0842	HV19	Castilla_LaMancha	Central
ESP0843	K1a1b1	Asturias	North
ESP0844	T2c1d+152	Castilla_y_Leon	Central
ESP0845	H1i	Aragon	North
ESP0846	H1j	Extremadura	Central
ESP0847	T1a1	Madrid	Central
ESP0848	K1a13a	Andalucia	South
ESP0849	H1e1a	Murcia	South
ESP0850	H3y	Galicia	North
ESP0851	U5b2c	C_Valenciana	East
ESP0852	H27e	Andalucia	South
ESP0853	N1b1a2	Andalucia	South
ESP0854	H1	Castilla_y_Leon	Central
ESP0855	H1	Baleares	East
ESP0856	H1q	C_Valenciana	East
ESP0857	H5s	C_Valenciana	East
ESP0858	H3av	Madrid	Central
ESP0859	T2b	Catalunya	East
ESP0861	T2a1a	Castilla_LaMancha	Central
ESP0862	H5a1a	Andalucia	South
ESP0863	H1ai	Galicia	North
ESP0864	T2b3+151	Extremadura	Central
ESP0865	H1j1b	Basque_Country	North
ESP0866	H13a2b3	C_Valenciana	East
ESP0867	H	Castilla_LaMancha	Central
ESP0868	T2c1d+152	Asturias	North
ESP0869	H5b1	C_Valenciana	East
ESP0870	H1j1	Andalucia	South
ESP0871	T1a5	Catalunya	East
ESP0872	H4a1d	Castilla_y_Leon	Central
ESP0873	J1c3	Murcia	South
ESP0874	V22	Andalucia	South
ESP0875	U8a1a4	Murcia	South
ESP0876	H13a1a1	Catalunya	East
ESP0877	H1af1	Castilla_y_Leon	Central
ESP0879	I2	Navarra	North

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0880	H1	Castilla_y_Leon	Central
ESP0881	H1e2	Catalunya	East
ESP0882	H10+(16093)	Cantabria	North
ESP0883	H1c7	C_Valenciana	East
ESP0884	H14b	Castilla_y_Leon	Central
ESP0885	H20a1a	Andalucia	South
ESP0886	H3c	Asturias	North
ESP0888	H	Castilla_LaMancha	Central
ESP0889	U3a1	Catalunya	East
ESP0890	H5	Asturias	North
ESP0891	H83	Asturias	North
ESP0892	K1a4c	Andalucia	South
ESP0893	U5b2b5	Andalucia	South
ESP0894	H1b1+16362	Castilla_LaMancha	Central
ESP0895	K2b1a1a	Catalunya	East
ESP0896	H6a1	Navarra	North
ESP0897	H	Madrid	Central
ESP0898	R	Andalucia	South
ESP0899	J1c5c1	Madrid	Central
ESP0900	H1j1	Andalucia	South
ESP0901	U3a1	C_Valenciana	East
ESP0902	J1c3	Andalucia	South
ESP0903	H5a	Andalucia	South
ESP0904	J1c2e2	Castilla_y_Leon	Central
ESP0905	U5b2b5	Madrid	Central
ESP0906	H1ao	Castilla_y_Leon	Central
ESP0907	K2b1a	Extremadura	Central
ESP0908	U5b1f1a	Castilla_y_Leon	Central
ESP0909	V	Castilla_y_Leon	Central
ESP0910	T2c1d+152	Andalucia	South
ESP0911	H	Andalucia	South
ESP0912	K1b1a1c	Andalucia	South
ESP0913	K1a4a1	Andalucia	South
ESP0914	K1b2b	Andalucia	South
ESP0915	H1b	Andalucia	South
ESP0916	V3a	Asturias	North
ESP0917	H2a2a1	Castilla_LaMancha	Central
ESP0918	U5b1e	Castilla_y_Leon	Central
ESP0919	U6a1a1	Castilla_y_Leon	Central
ESP0920	J1c1b	Castilla_LaMancha	Central
ESP0921	H1q	Andalucia	South
ESP0922	L3b1b1	Aragon	Central
ESP0923	HV24	Castilla_y_Leon	Central
ESP0924	H1bs	Andalucia	South
ESP0925	J2b1a2	Andalucia	South
ESP0926	L2c5	Andalucia	South
ESP0927	U4a	Andalucia	South
ESP0928	L2a1c+16129	Andalucia	South

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0929	I5a2+16086	Extremadura	Central
ESP0930	HV0b	Andalucia	South
ESP0931	H17c	Murcia	South
ESP0932	N1b1a2	Castilla_y_Leon	Central
ESP0933	U6b	Andalucia	South
ESP0934	H1e1a	Murcia	South
ESP0935	K1a3a1	Castilla_y_Leon	Central
ESP0936	J1c2e2	Andalucia	South
ESP0937	H4a1a	Andalucia	South
ESP0938	H5a1a	Asturias	North
ESP0939	H1	Castilla_LaMancha	Central
ESP0940	K1a+195	Andalucia	South
ESP0941	H2a1	Catalunya	East
ESP0942	H3au	Aragon	Central
ESP0943	J1c2c1	Andalucia	South
ESP0944	K1a4a1	Baleares	East
ESP0945	H1+16189	Cantabria	North
ESP0946	H1j1	Castilla_LaMancha	Central
ESP0947	H1a3	Murcia	South
ESP0948	H5k	Castilla_y_Leon	Central
ESP0949	J1c5c1	Andalucia	South
ESP0950	K1a4a1	Asturias	North
ESP0952	HV	Castilla_LaMancha	Central
ESP0953	X1c	Aragon	North
ESP0954	L1b1a14	Africa	Other
ESP0955	H3c2a	Andalucia	South
ESP0956	H10	C_Valenciana	East
ESP0957	J1c3a	C_Valenciana	East
ESP0958	H6a1a7	Cantabria	North
ESP0959	H1c	Murcia	South
ESP0960	H1j1	Castilla_y_Leon	Central
ESP0961	H64	C_Valenciana	East
ESP0962	H1j1	Cantabria	North
ESP0963	H35	Aragon	North
ESP0964	H1q	Castilla_LaMancha	Central
ESP0965	U6a3a1	Asturias	North
ESP0966	U6a1b3	Andalucia	South
ESP0967	X2b+226	Aragon	Central
ESP0968	H1a3	Castilla_y_Leon	Central
ESP0969	L1b1a6	Murcia	South
ESP0970	H1	Andalucia	South
ESP0971	T2a1b1a1	Castilla_LaMancha	Central
ESP0972	J1c1g	Madrid	Central
ESP0973	H5g	Asturias	North
ESP0974	H5a1	Castilla_LaMancha	Central
ESP0975	L3e2b	Galicia	North
ESP0976	H5a3a	Castilla_y_Leon	Central
ESP0977	U5a1c1a	Aragon	Central

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP0978	J1c2e2	Castilla_y_Leon	Central
ESP0980	H2a2a1	C_Valenciana	East
ESP0981	H1b3	Aragon	Central
ESP0982	M1b1a	Aragon	Central
ESP0983	A8a	Murcia	South
ESP0984	U5b2a1a2	Cantabria	North
ESP0985	T2c1d+152	Andalucia	South
ESP0986	U5b1f1	Asturias	North
ESP0987	K1a+195	Murcia	South
ESP0988	H1cf	Catalunya	East
ESP0989	H+152	C_Valenciana	East
ESP0990	J1c3	C_Valenciana	East
ESP0991	U1a1b	Andalucia	South
ESP0992	H+16129	Andalucia	South
ESP0993	H64	Andalucia	South
ESP0994	U5b1c2	Andalucia	South
ESP0995	H1t1a	Castilla_y_Leon	Central
ESP0996	V14	Andalucia	South
ESP0997	U5a2a2	Andalucia	South
ESP0998	V1a	Castilla_y_Leon	Central
ESP0999	U5b2b3	Andalucia	South
ESP1000	U5b1f1a	Andalucia	South
ESP1001	J1c3g	Murcia	South
ESP1002	H39	Catalunya	East
ESP1003	U3b2a1	Castilla_LaMancha	Central
ESP1004	HV+16311	Asturias	North
ESP1005	H1cf	Canarias	Other
ESP1006	K1a+195	Catalunya	East
ESP1007	J1c3	Castilla_y_Leon	Central
ESP1008	T1a	Catalunya	East
ESP1009	U5a2	Castilla_y_Leon	Central
ESP1010	T2b33	Aragon	Central
ESP1011	U4a1	Andalucia	South
ESP1012	HV0+195	Andalucia	South
ESP1013	H1ak1	Catalunya	East
ESP1014	J2a2d	Extremadura	Central
ESP1015	X1c	Andalucia	South
ESP1016	H1e2	Castilla_y_Leon	Central
ESP1017	H1+152	Castilla_y_Leon	Central
ESP1018	K2b1a	Castilla_y_Leon	Central
ESP1019	I1	Murcia	South
ESP1020	H1j1b	Aragon	Central
ESP1021	H13a2b2a	Cantabria	North
ESP1022	H3c	Cantabria	North
ESP1023	J1c3g	Extremadura	Central
ESP1024	L3e1f	Andalucia	South
ESP1025	H1a3	Asturias	North
ESP1026	V3a1	Andalucia	South

Table S2 continued from previous page

ID	*Haplogroup*	*Region*	*Area*
ESP1027	J1c1	Catalunya	East
ESP1028	HV0f	Catalunya	East
ESP1029	U5b1c	Castilla_y_Leon	Central
ESP1030	H1+16189	Aragon	North
ESP1031	I2'3	Castilla_LaMancha	Central
ESP1032	H13a2b2a	Andalucia	South
ESP1033	H1bv1	Andalucia	South
ESP1034	T2b	Asturias	North
ESP1035	H13a1a1	Catalunya	East
ESP1036	H105	Andalucia	South
ESP1037	U5b1f1	Asturias	North
ESP1038	U5b1f1a	Andalucia	South
ESP1039	H3b5	Castilla_LaMancha	Central
ESP1040	H1+16189	Andalucia	South
ESP1041	H1e2	Andalucia	South
ESP1042	M1a2a	Galicia	North
ESP1043	H3b+16129	Aragon	Central

6.2 Supplement Chapter II

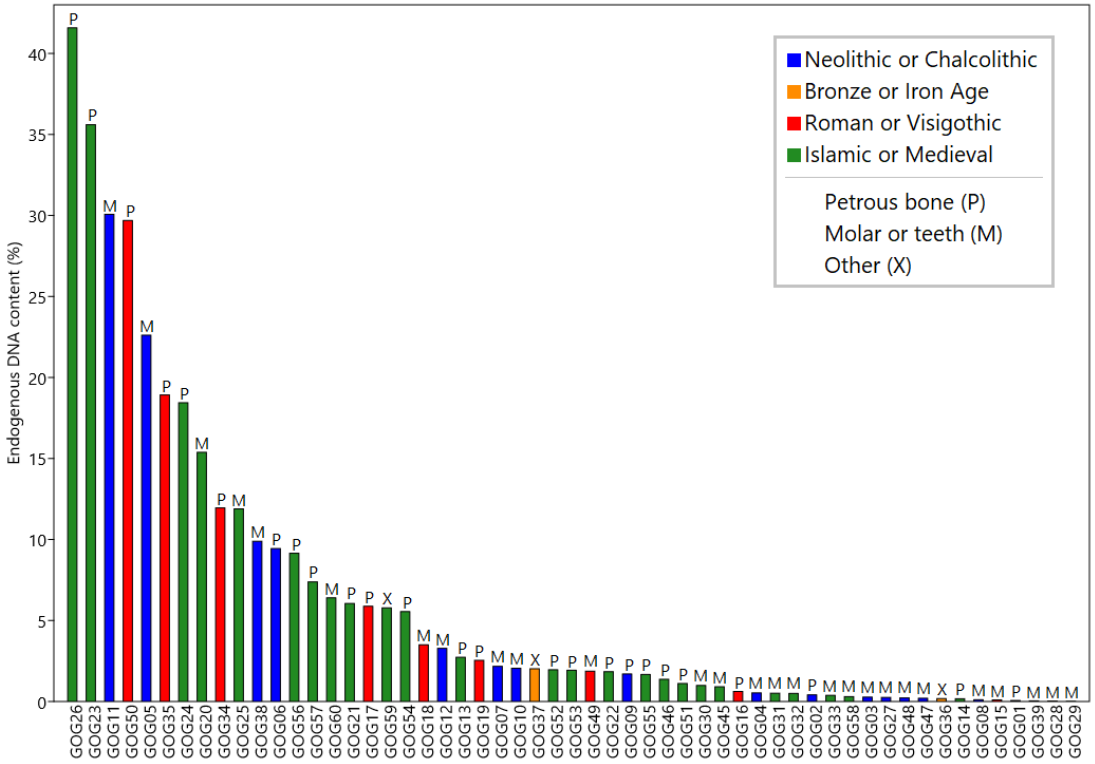


Figure S8: Endogenous content calculated with the MiSeq sequencing results for all 55 GOG samples which were processed in the ancient laboratory by me.

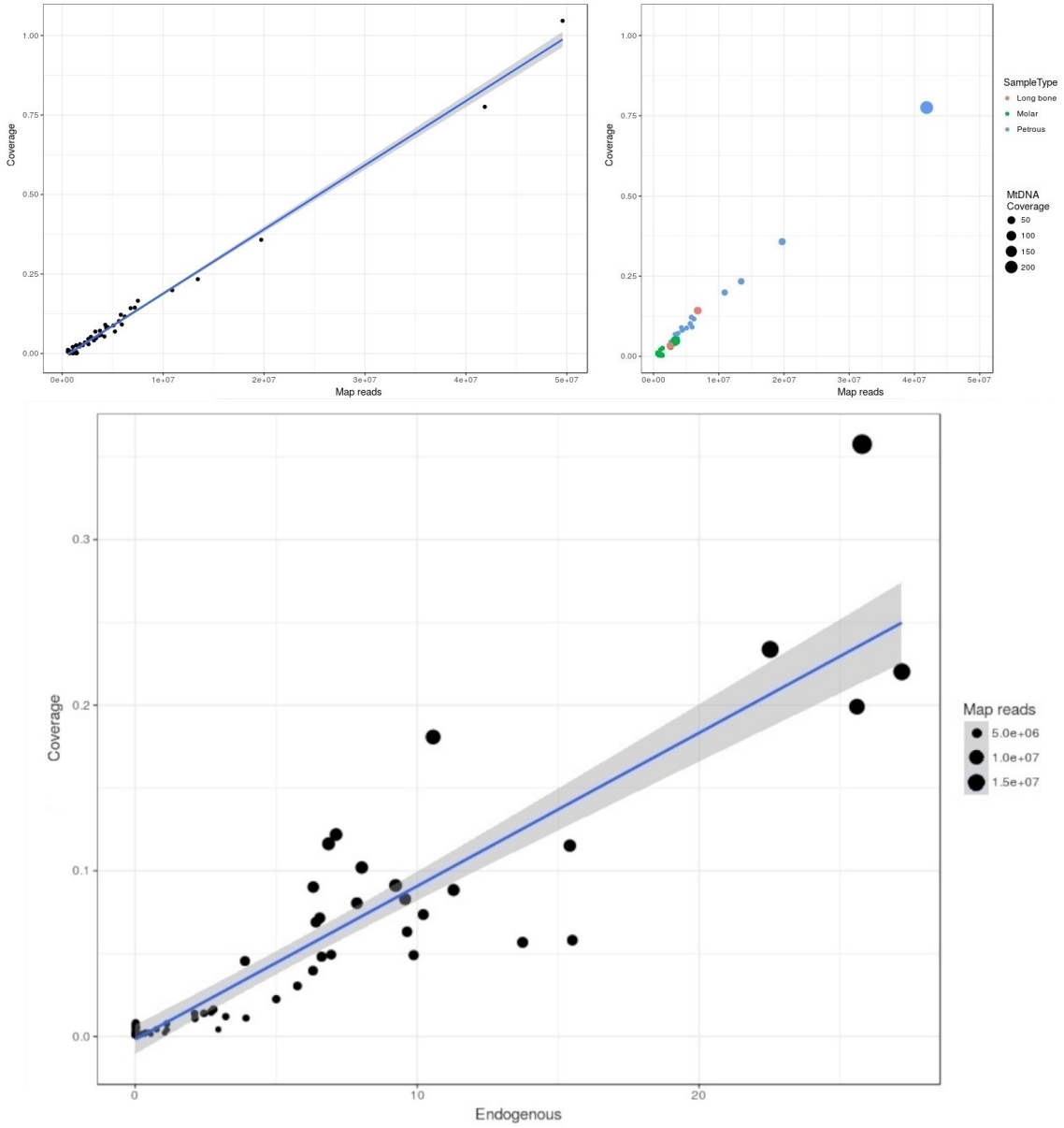


Figure S9: Correlation between number of map reads and endogenous DNA with resulting coverage from shotgun sequencing. This information was used to decide how to distribute the best samples across eight Illumina HiSeq4000 lanes in order to maximize coverage based on endogenous DNA and expected mapped reads. Combining the information provided by endogenous DNA and the known output of 300 millions of sequences generated by an Illumina HiSeq4000, I was able to predict the amounf of mapped reads obtained by each sample in each lane even when samples are multiplexed in the same lane. This data allowed infering that 60 million reads roughly correlate with 1X coverage and helped decision making and budget strategies.

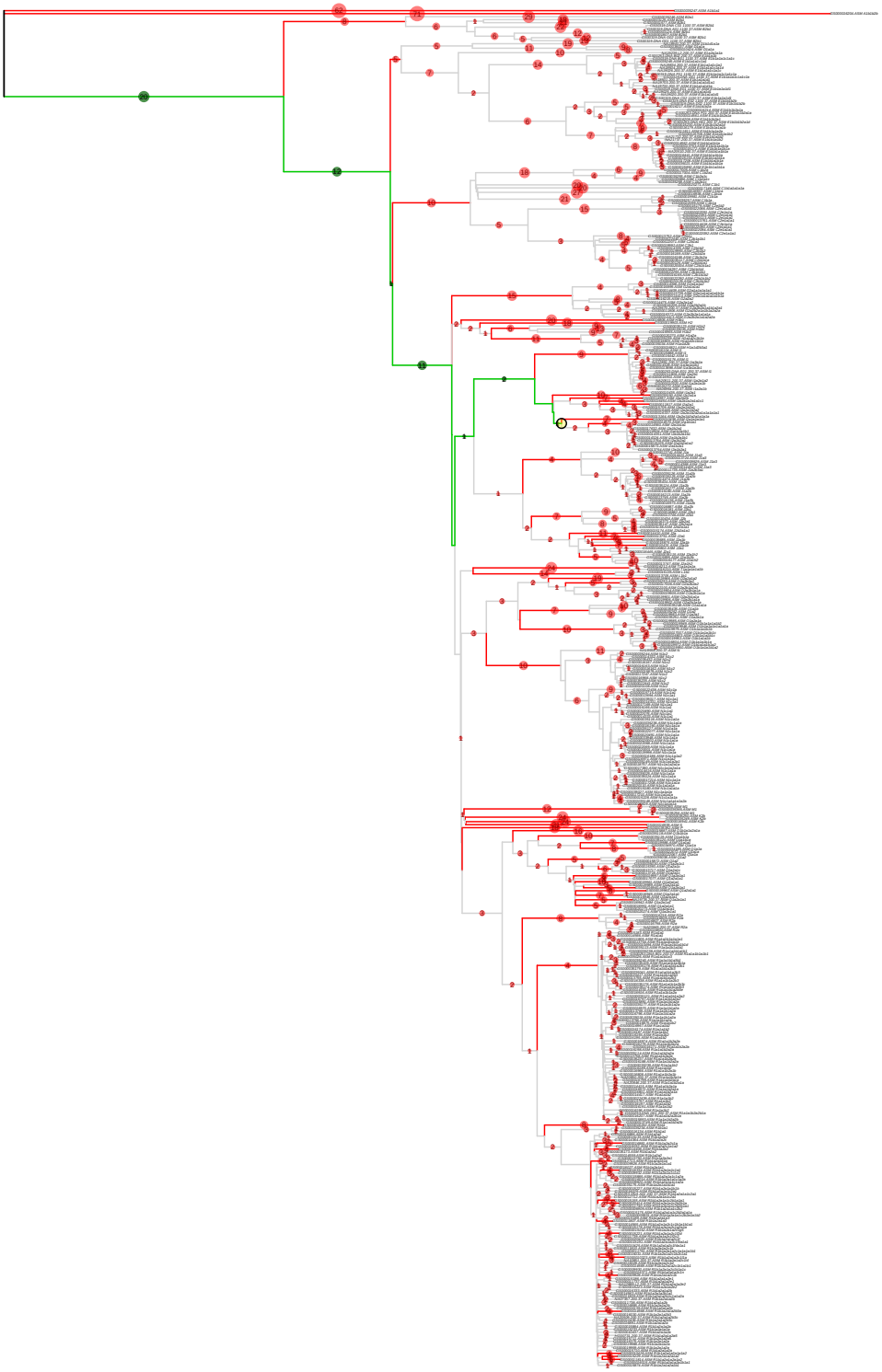


Figure S10: PathPhynder result for the classification of Chromosome Y haplogroup of sample GOG38 whose Yleaf classification was doubtful. The best resolution of the Yleaf output classified the haplogorup as I1, however it is very rare in Iberia, pathPhynder on the other hand traces all the mutations to a sub-branch of I2a which is very common in Chalcolithic Iberian males and makes more sense.

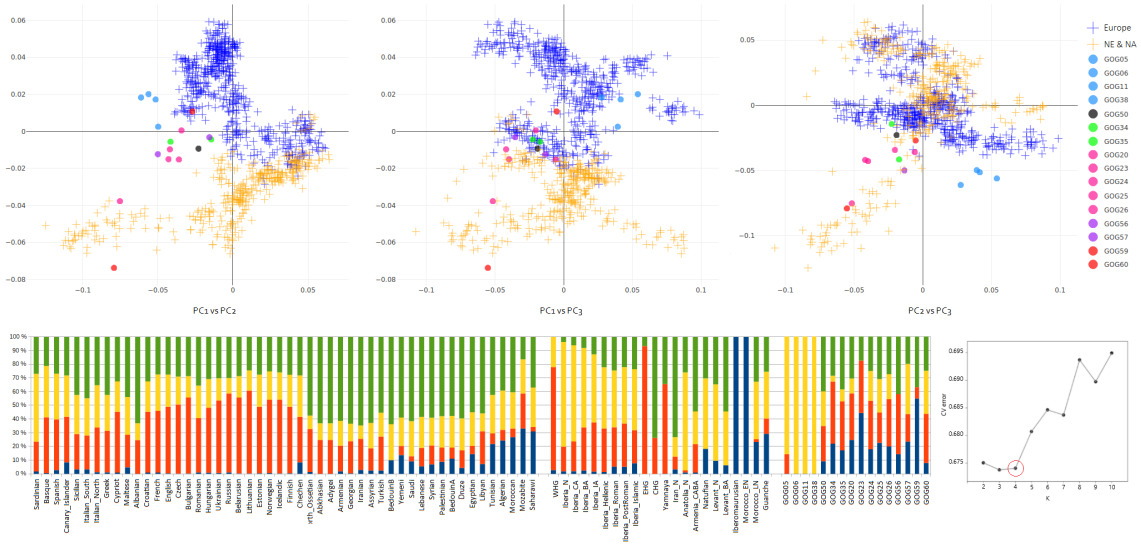


Figure S11: PCA made only with 25k transversions extracted from the 1000Genomes panel present in the modern populations from the Human Origins dataset and GOG samples projected (top). Unsupervised ADMIXTURE at K=4 with its cross-validation errors of each K performed with the same 25k transvrsions only dataset. Each bar represents the averaged ADMIXTURE profiles of all individuals within each population except GOG labels which are individuals (bottom).

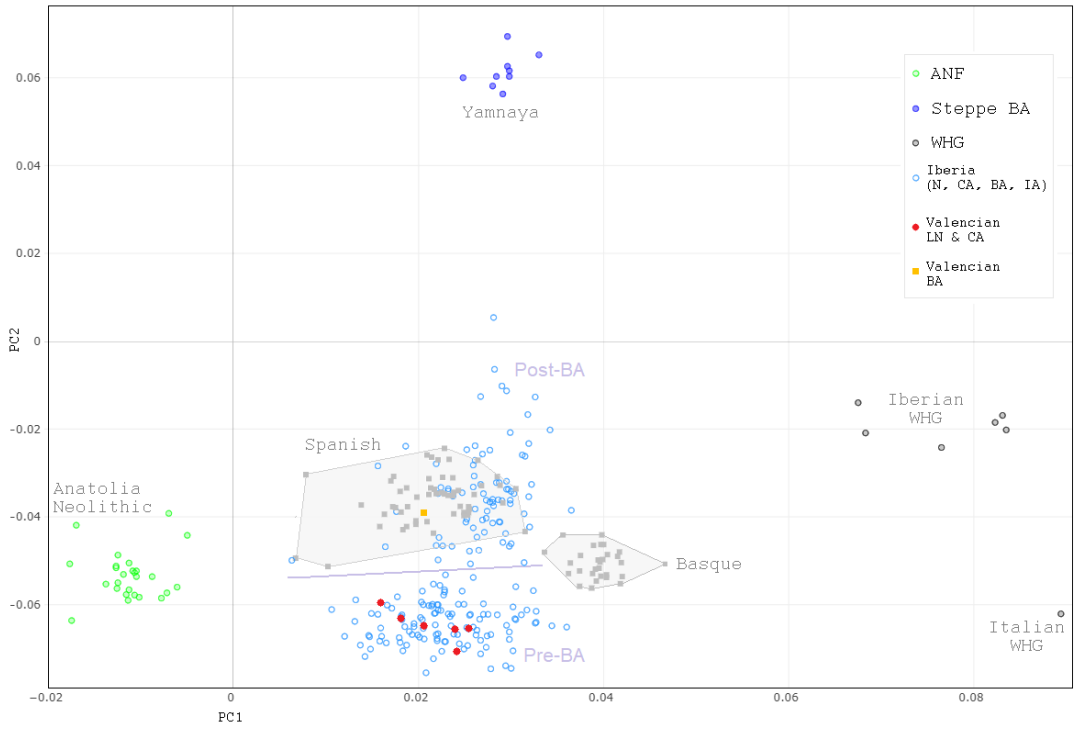


Figure S12: PCA with three poles of ancestry of Prehistorical samples. Fundamentally the same as the prehistorical PCA in the main text, but includes the cluster of modern Spaniards, the MS samples under 10k SNPs and one Italian WHG (Continenza) for comparison.

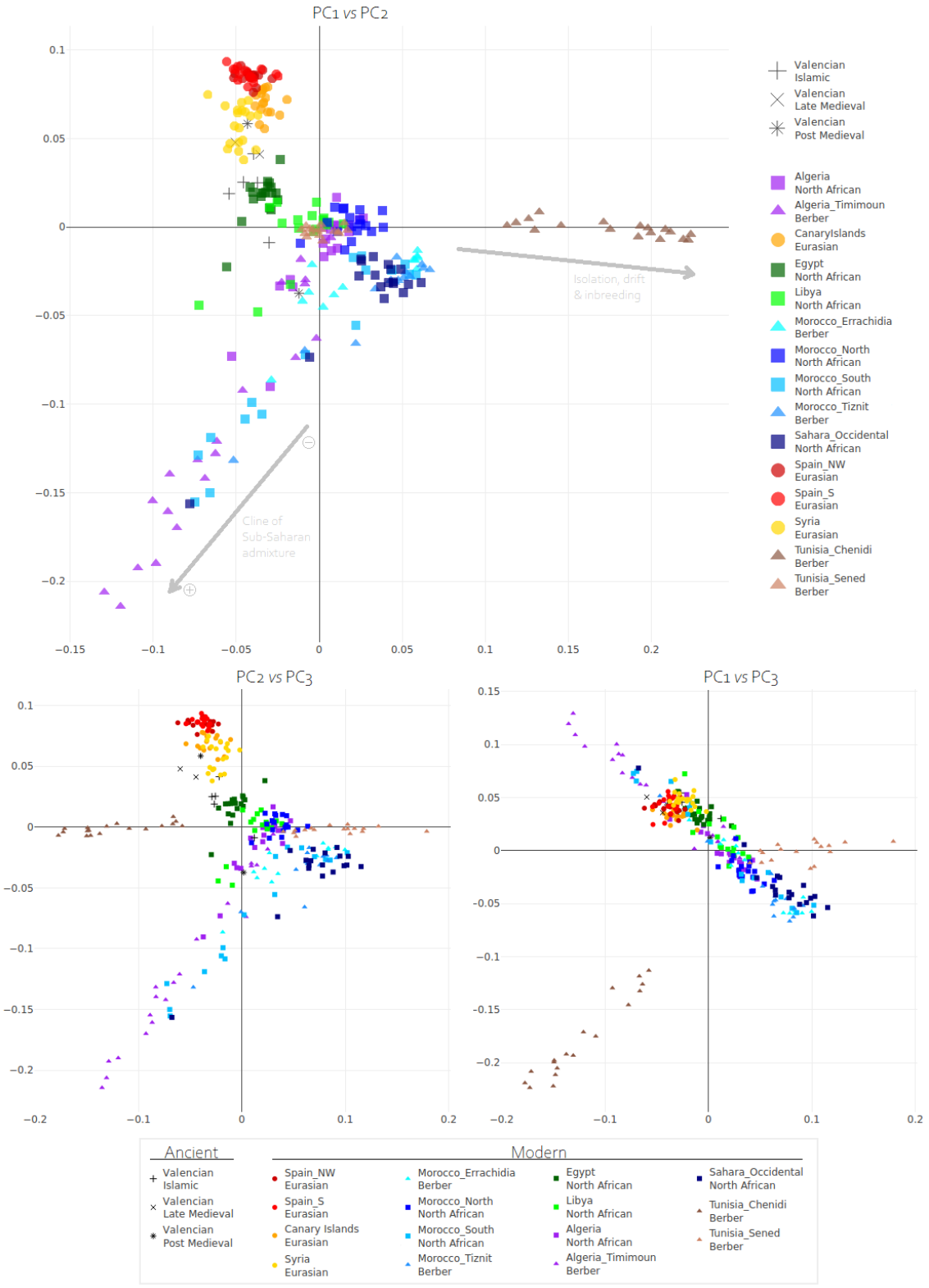


Figure S13: PCA plots (PC1vsPC2, PC1vsPC3 and PC2vsPC3) merging and projecting GOG Islamic and Medieval samples onto the dataset with the North African Affymetrix 6.0 from David Comas Lab alone(Henn et al., 2012; Botigué et al., 2013; Arauna et al., 2017).

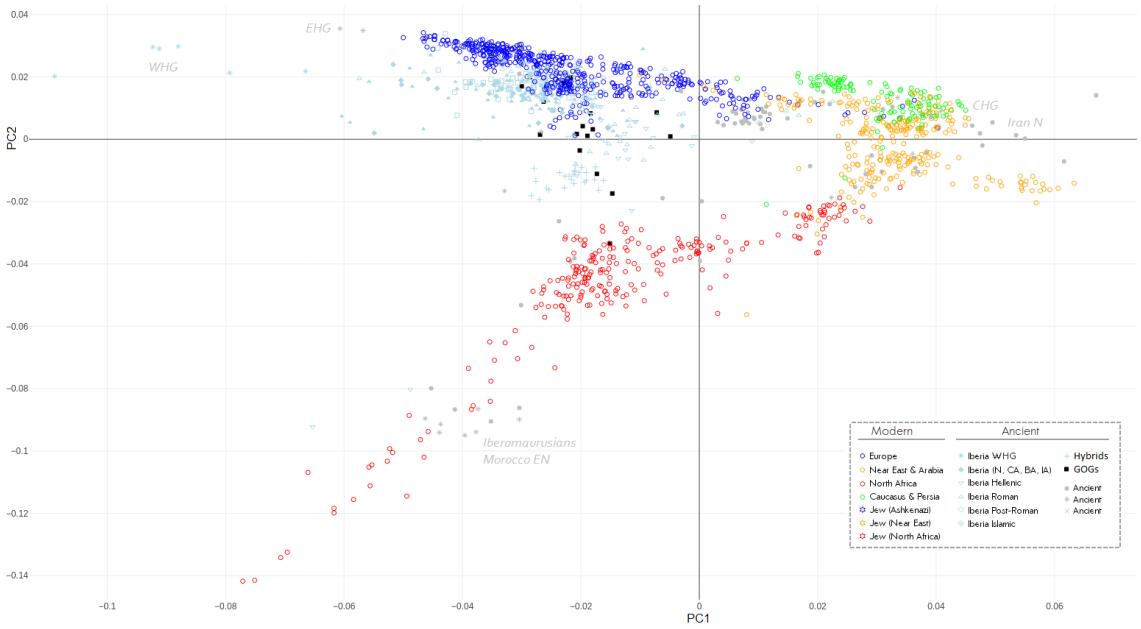


Figure S14: PCA with the GOG samples only (does not include the MS samples) merged with the full North African dataset from David Comas Lab (Arauna et al., 2017) and the HO 600k SNPs dataset. The overlap of SNPs between the two SNP chips is just over 20,000 positions. A subset of populations from this combined dataset were used to make the Medieval PCA from the main text.

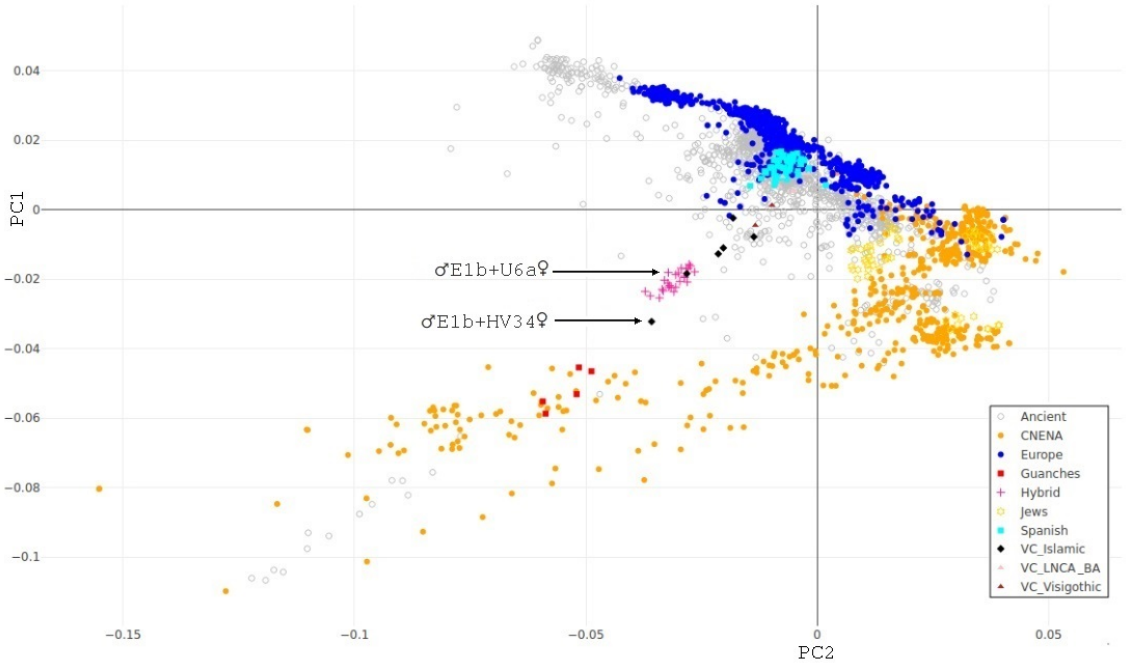


Figure S15: PCA with Guanche+Spanish simulated hybrids projected onto the Human Origins dataset of 1240k SNPs which includes several more ancient populations (in grey), instead of the 620k SNPs dataset with only modern populations used in the main text. There are less ancient popuations in the figures from the main body because they were especifically re-processed to reduce bias.

Figure S16: Unsupervised ADMIXTURE using only only 434 ancient samples, the majority of them from Spain.

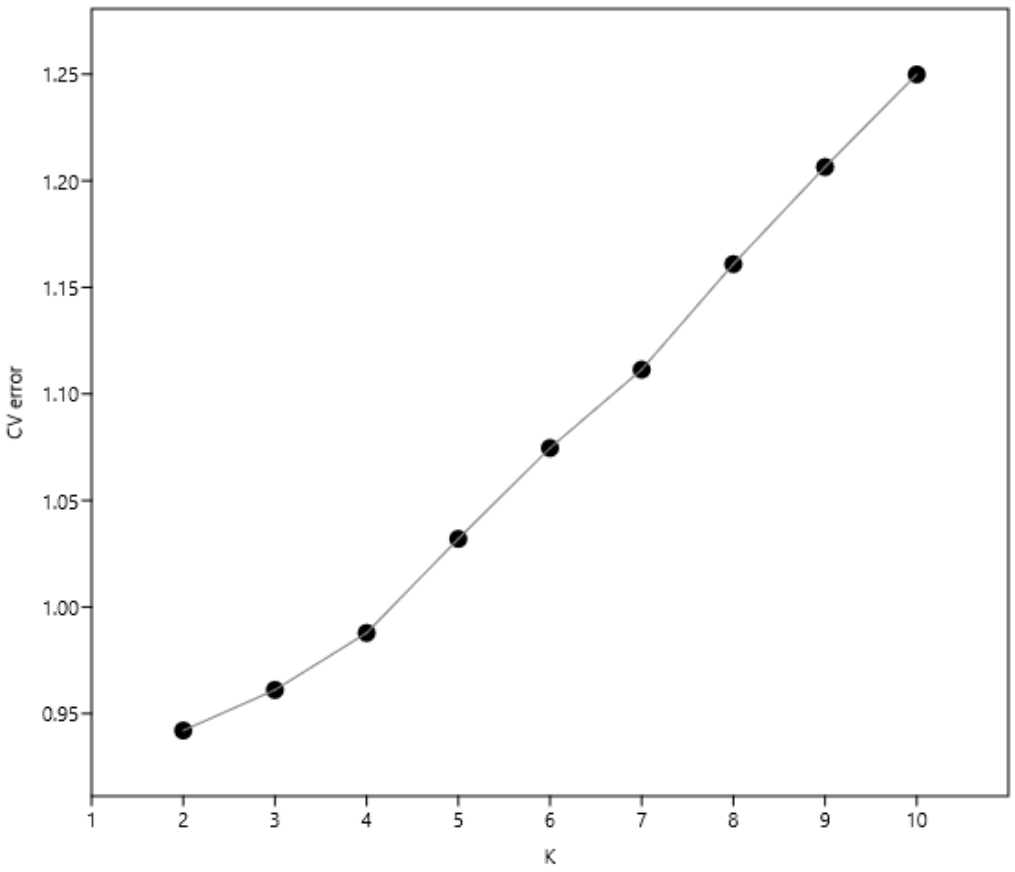


Figure S17: Cross validation errors of the best run for each K in the ADMIXTURE plot made with 434 ancient genomes above 10k SNPs covered (see previous figure).

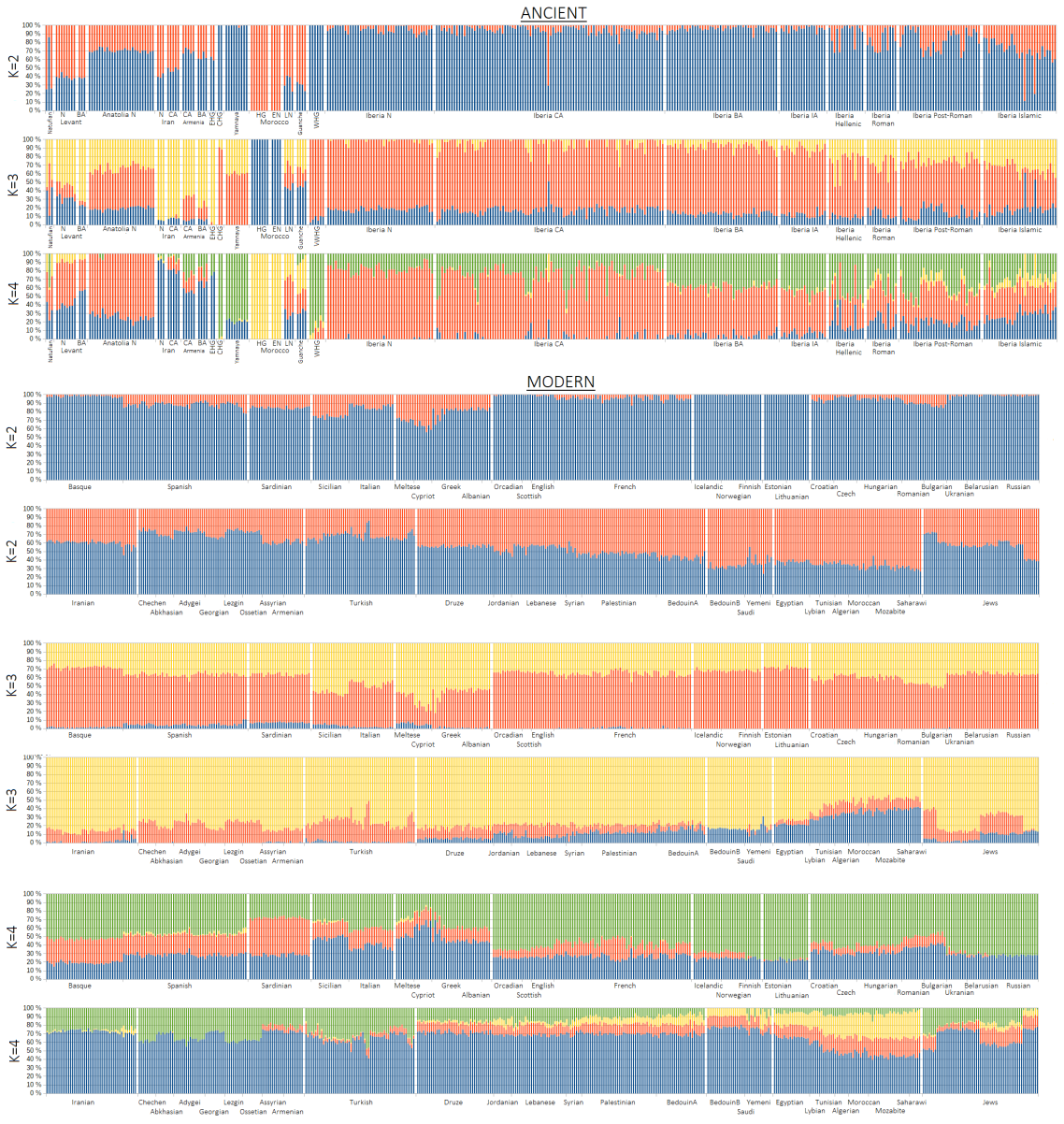


Figure S18: Unsupervised ADMIXTURE of 434 ancient samples and modern Human Origins populations. For simplicity, plots here only for the best K values before CV error rates start increasing after K=4.

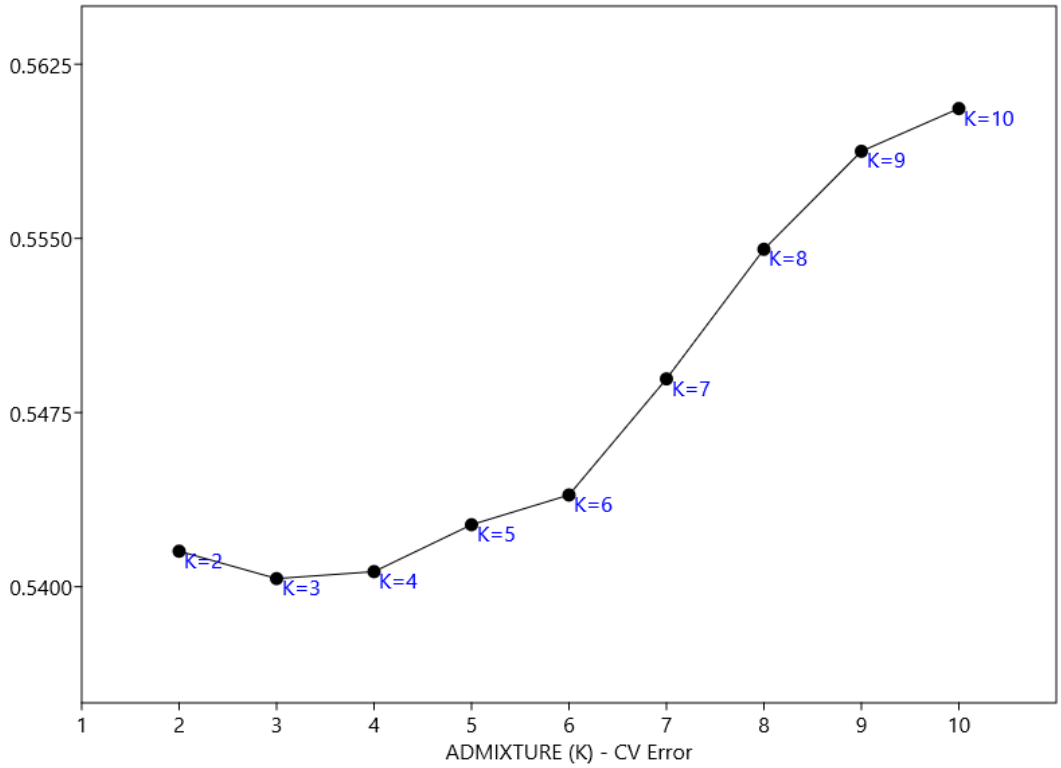


Figure S19: Cross validation errors of the best run for each K in the ADMIXTURE plot made with 434 ancient samples and modern European, Near Eastern and North African populations from the HO dataset (see previous figure).

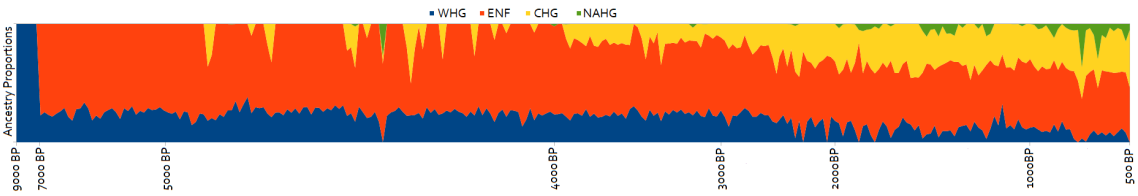


Figure S20: Supervised ADMIXTURE at K=4 with all public Spanish ancient samples ancient samples. Individuals have been ordered chronologically according to their archaeological dating. With transversions only and samples with minimum 10k SNPs.

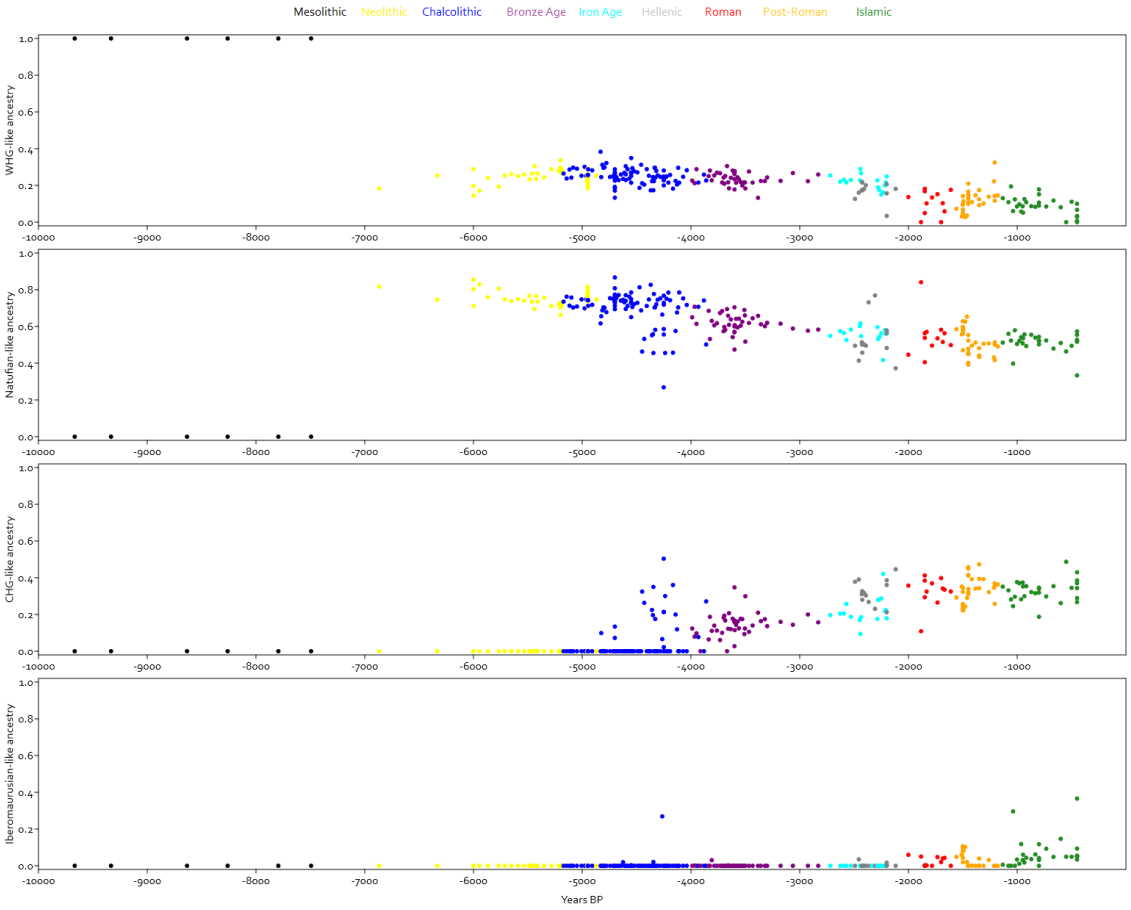


Figure S21: Evolution of ancestry frequency through time in Iberia with a 3-point rolling average to smooth the uncertainty of the error in the dating of the individuals and reduce the effect of individual variability.

$Rank$	RSS	$\%P_1$	P_1	$\%P_2$	P_2	$\%P_3$	P_3
1	4.0 10^{-4}	25.4	Iberomaurusian	67.8	Iberia-CA	6.8	CHG
2	5.6 10^{-4}	42.1	Guanches	51.4	Sardinian	6.5	CHG
3	7.5 10^{-4}	30.0	Guanches	62.5	Iberia-CA	7.5	CHG
4	7.8 10^{-4}	23.9	Iberomaurusian	49.8	Spanish	26.3	Natufian
5	8.5 10^{-4}	42.2	Spanish	26.0	Natufian	31.8	Mozabite
6	8.8 10^{-4}	43.5	Guanches	43.1	Iberia-CA	13.3	Yamnaya
7	9.4 10^{-4}	56.0	Iberia-CA	38.5	Mozabite	5.5	CHG
8	9.6 10^{-4}	45.4	Guanches	48.5	Iberia-BB	6.1	CHG
9	9.9 10^{-4}	37.5	Guanches	45.6	Spanish	16.9	Natufian
10	0.00102	50.8	Spanish	42.8	Natufian	6.4	Yoruba

Table S3: ADMIXTURE Top-10 triplets for *Segorbe*. Contains the top-10 triplets for *Segorbe*. In this case, 189 triplets give one negative coefficient out of the 286 possible combinations of size m = 3.

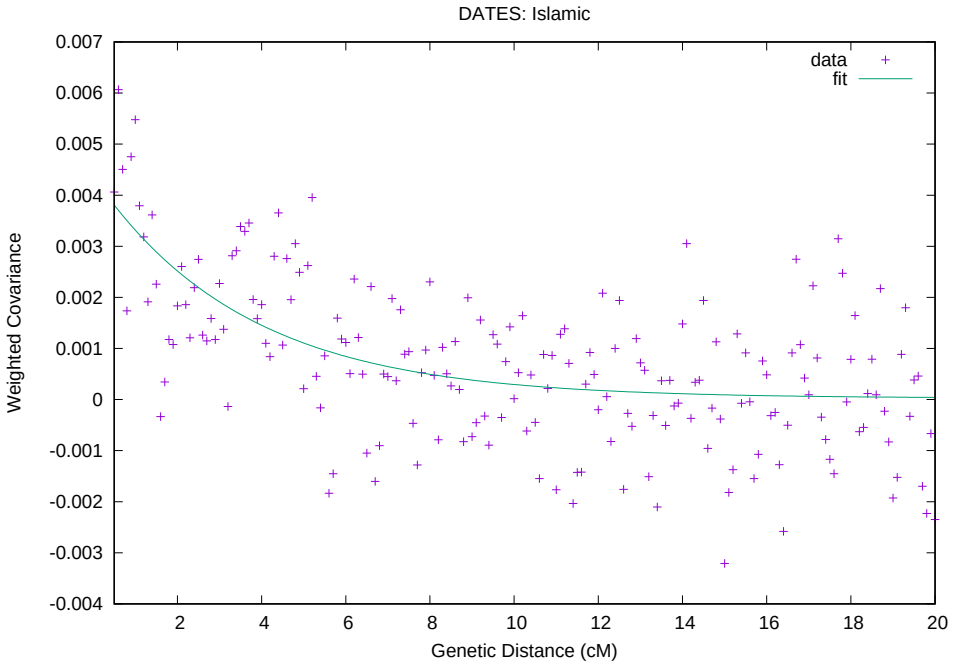


Figure S22: DATES for modern Spanish+Moroccan as sources for the Islamic GOG samples (Admixture event 785 years before age of ancient samples (11-13th Cent. CE) with a standard error of 194 years. Mixture estimated between 1800-1400 years before present (3rd to 7th Century CE).

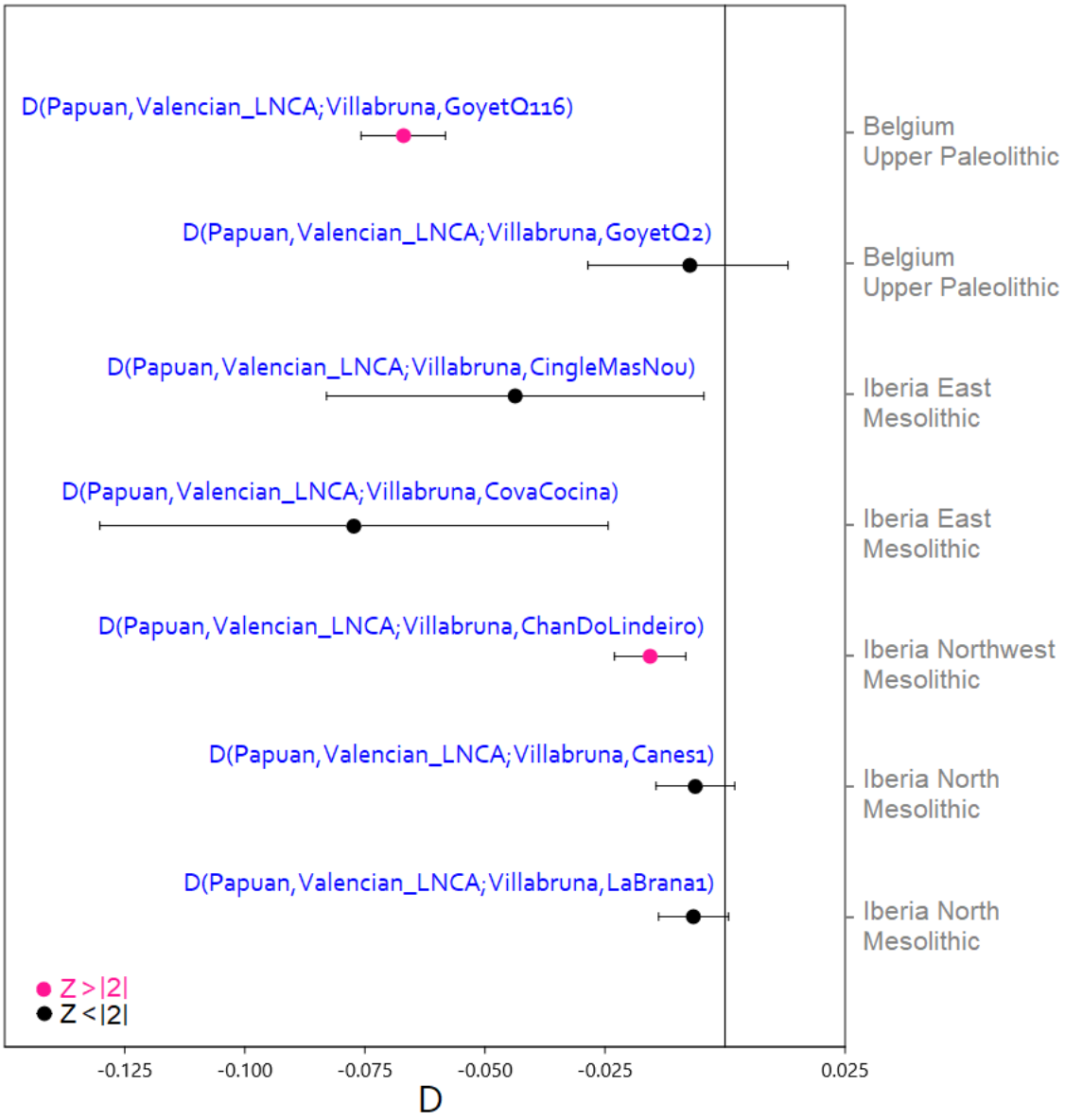


Figure S23: D-statistic tests to check wheter the Valencina Late Neolithic and Chalcolirhic populations are more similar to hunter-gatherers from Goyet or Villabruna clusters.

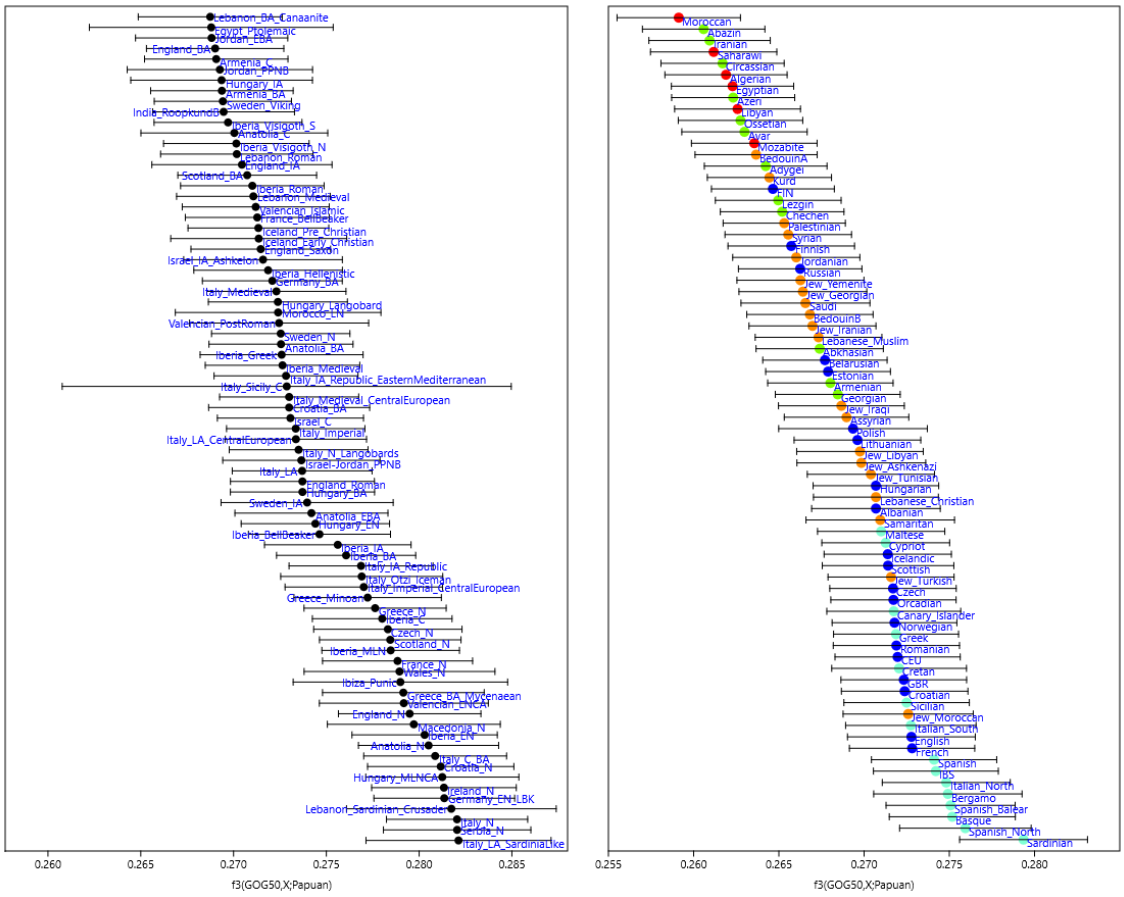


Figure S24: Outgroup-f3 statistic tests for Dafne (GOG50), a Late Roman sample, compared with ancient (left and) modern (right) populations.

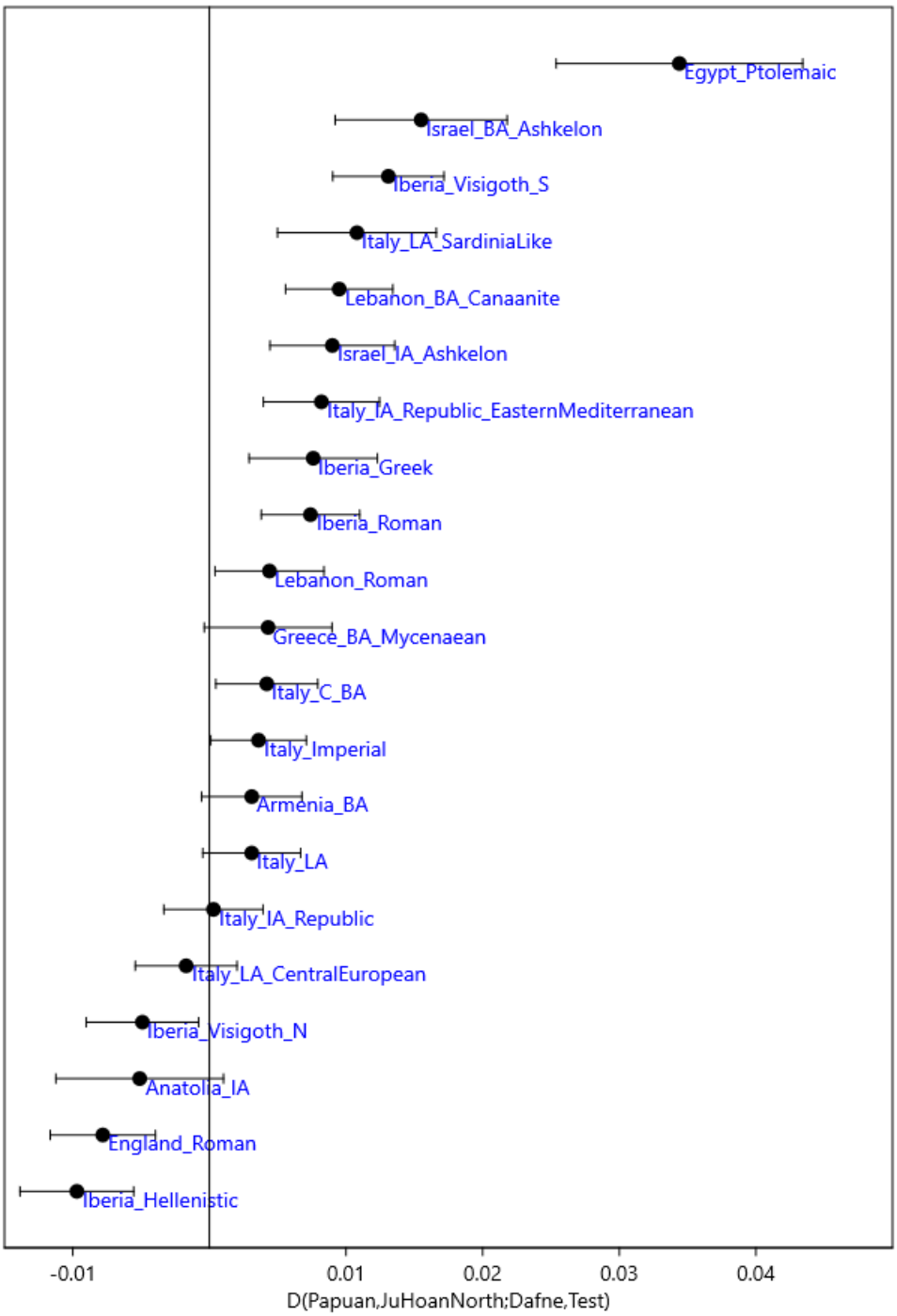


Figure S25: D-statistic tests for Dafne (GOG50), a Late Roman sample, compared with other ancient populations that could have been in a higher or lower degree contemporary to her.

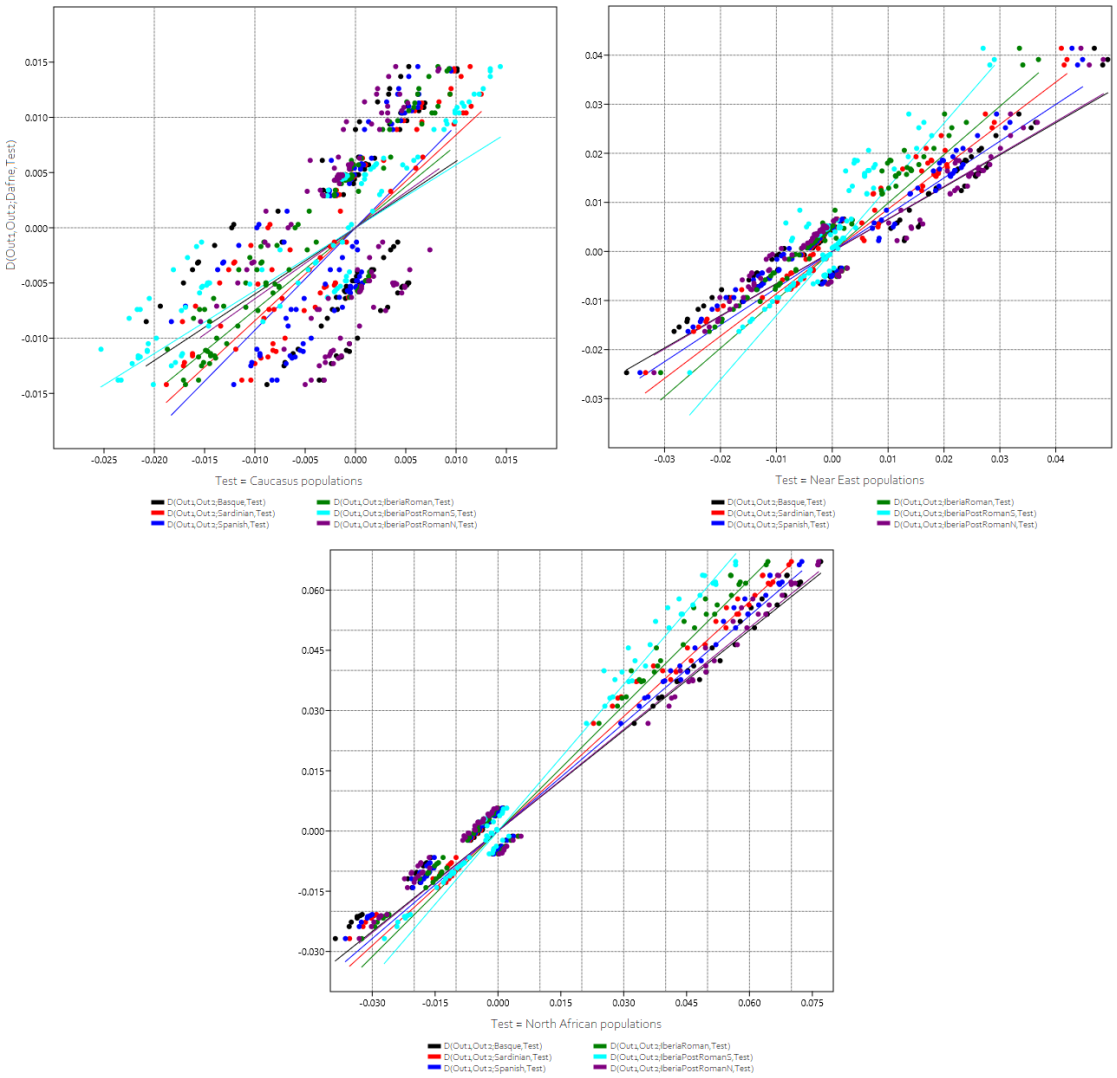


Figure S26: These three plots represents the fitted linear regressions that the f4-ratio tests use to calculate contribution coefficient from designated sources (alpha and 1-alpha). These are the linear regression that correspong to Figure 65 in the main text but grouping the secondary populations (1-alpha) under Caucasus, Near East an North Africa. The primary sources (alpha) are indicated by the colours in the legends (Basque, Sardinian, Spanish, Iberia Roman, Iberia Visigothic South and Iberia Visigothic North). Ten different combination of outgroups were used for each individual population to adjust the regression.

Table S4: Admixture proportion α for Dafne=α·Pop$_1$+(1-α)·Pop$_2$, where Pop$_1$ stands for A=Basque, B= Sardinian, C=Spanish, D=Iberia-Roman, E=Iberia-Visigoth-S, F=Iberia-Visigoth-N. NaN: Not-a-Number. Formula: $\alpha = D(Out_1, Out_2; Dafne, Pop_2)/D(Out_1, Out_2; Pop_1, Pop_2)$.

			Pop$_1$					
Out$_1$–Out$_2$	Pop$_2$	metaPop$_2$	A	B	C	D	E	F
Papuan-Australian	Adygei	Caucasus-N	6.11	-13.75	11.00	-13.75	3.06	-9.17
Papuan-Australian	Adygei	Caucasus-N	6.11	-13.75	11.00	-13.75	3.06	-9.17
Papuan-Han	Adygei	Caucasus-N	1.64	1.71	1.79	1.92	1.07	2.17
Papuan-Ju-hoan-North	Adygei	Caucasus-N	3.22	0.86	1.16	0.48	0.34	2.06
Mbuti-Ju-hoan-North	Adygei	Caucasus-N	-2.56	-7.67	-3.83	-6.57	-3.54	-2.56
Papuan-Karitiana	Adygei	Caucasus-N	2.76	0.96	2.61	1.42	1.03	5.88
Papuan-Mbuti	Adygei	Caucasus-N	12.33	1.35	2.02	0.74	0.54	5.05
Papuan-Saami	Adygei	Caucasus-N	0.20	0.26	0.24	0.33	0.79	0.31
Papuan-Somali	Adygei	Caucasus-N	1.74	0.72	1.18	0.81	0.68	4.07
Papuan-Surui	Adygei	Caucasus-N	5.18	1.37	3.68	2.24	1.01	12.67
Papuan-Yoruba	Adygei	Caucasus-N	6.11	1.22	1.78	0.83	0.54	8.29
Average:	————-	————-	0.20	0.70	0.24	0.64	0.58	0.31
Papuan-Australian	Chechen	Caucasus-N	5.89	-13.25	10.60	-10.60	3.53	-8.83
Papuan-Han	Chechen	Caucasus-N	1.68	1.80	1.89	2.00	1.07	2.30
Papuan-Ju-hoan-North	Chechen	Caucasus-N	4.11	0.91	1.25	0.49	0.35	1.95
Mbuti-Ju-hoan-North	Chechen	Caucasus-N	-3.20	-16.00	-4.80	-12.00	-4.36	-2.82
Papuan-Karitiana	Chechen	Caucasus-N	1.95	0.97	1.92	1.29	1.03	2.75
Papuan-Mbuti	Chechen	Caucasus-N	18.67	1.42	2.15	0.76	0.54	4.67
Papuan-Saami	Chechen	Caucasus-N	-0.02	-0.04	-0.03	-0.05	-0.75	-0.05
Papuan-Somali	Chechen	Caucasus-N	1.87	0.72	1.22	0.81	0.67	4.46
Papuan-Surui	Chechen	Caucasus-N	3.71	1.33	2.93	2.00	0.99	5.73
Papuan-Yoruba	Chechen	Caucasus-N	6.94	1.26	1.87	0.85	0.55	6.94

Table S4 continued from previous page

Average:	—————	—————	NaN	0.87	NaN	0.73	0.62	NaN
Papuan-Australian	Circassian	Caucasus-N	-20.00	-2.67	-6.67	-2.86	6.67	-2.50
Papuan-Han	Circassian	Caucasus-N	1.43	1.47	1.53	1.55	1.07	1.73
Papuan-Ju-hoan-North	Circassian	Caucasus-N	1.93	0.92	1.12	0.59	0.43	1.55
Mbuti-Ju-hoan-North	Circassian	Caucasus-N	-1.06	-1.79	-1.31	-1.70	-1.26	-0.97
Papuan-Karitiana	Circassian	Caucasus-N	2.31	0.97	2.26	1.37	1.05	4.00
Papuan-Mbuti	Circassian	Caucasus-N	4.31	1.33	1.77	0.81	0.59	3.14
Papuan-Saami	Circassian	Caucasus-N	0.23	0.30	0.28	0.39	0.81	0.35
Papuan-Somali	Circassian	Caucasus-N	1.61	0.76	1.17	0.84	0.71	2.84
Papuan-Surui	Circassian	Caucasus-N	3.04	1.30	2.58	1.88	1.02	4.15
Papuan-Yoruba	Circassian	Caucasus-N	3.83	1.22	1.66	0.88	0.58	4.45
Average:	—————	—————	0.23	0.74	0.28	0.70	0.63	0.35
Papuan-Australian	Lezgin	Caucasus-N	16.00	-4.80	-48.00	-4.80	5.33	-3.43
Papuan-Han	Lezgin	Caucasus-N	2.02	2.17	2.41	2.62	1.14	3.71
Papuan-Ju-hoan-North	Lezgin	Caucasus-N	6.40	0.88	1.25	0.45	0.31	2.37
Mbuti-Ju-hoan-North	Lezgin	Caucasus-N	-3.00	-15.00	-5.00	-11.25	-4.50	-2.81
Papuan-Karitiana	Lezgin	Caucasus-N	2.12	0.96	2.08	1.34	1.02	3.44
Papuan-Mbuti	Lezgin	Caucasus-N	-50.00	1.43	2.27	0.72	0.51	6.67
Papuan-Saami	Lezgin	Caucasus-N	0.00	0.00	0.00	0.00	0.00	0.00
Papuan-Somali	Lezgin	Caucasus-N	1.84	0.70	1.20	0.79	0.66	4.56
Papuan-Surui	Lezgin	Caucasus-N	2.75	1.28	2.35	1.78	1.01	3.65
Papuan-Yoruba	Lezgin	Caucasus-N	17.50	1.28	2.02	0.82	0.51	17.50
Average:	—————	—————	0.00	0.63	0.00	0.56	0.40	0.00
Papuan-Australian	Ossetian	Caucasus-N	48.00	-4.36	-24.00	-4.00	4.80	-3.69
Papuan-Han	Ossetian	Caucasus-N	1.41	1.45	1.49	1.56	1.06	1.73
Papuan-Ju-hoan-North	Ossetian	Caucasus-N	2.93	0.90	1.19	0.51	0.36	1.86
Mbuti-Ju-hoan-North	Ossetian	Caucasus-N	-2.44	-7.33	-3.67	-5.50	-2.93	-2.44
Papuan-Karitiana	Ossetian	Caucasus-N	2.21	0.95	2.12	1.32	1.01	3.25
Papuan-Mbuti	Ossetian	Caucasus-N	8.36	1.34	1.95	0.75	0.55	3.90

Table S4 continued from previous page

Papuan-Saami	Ossetian	Caucasus-N	0.29	0.37	0.34	0.46	0.88	0.43
Papuan-Somali	Ossetian	Caucasus-N	1.76	0.73	1.20	0.82	0.68	3.47
Papuan-Surui	Ossetian	Caucasus-N	3.67	1.29	2.88	1.95	0.98	5.76
Papuan-Yoruba	Ossetian	Caucasus-N	5.12	1.23	1.78	0.85	0.55	5.59
Average:	————	————	0.29	0.74	0.34	0.68	0.67	0.43
Papuan-Australian	Abkhasian	Caucasus-S	5.80	-19.33	9.67	-19.33	2.76	-14.50
Papuan-Han	Abkhasian	Caucasus-S	1.90	2.02	2.16	2.31	1.08	3.03
Papuan-Ju-hoan-North	Abkhasian	Caucasus-S	-0.96	0.77	2.08	0.26	0.16	-1.93
Mbuti-Ju-hoan-North	Abkhasian	Caucasus-S	-0.94	-1.53	-1.16	-1.38	-1.04	-0.91
Papuan-Karitiana	Abkhasian	Caucasus-S	21.33	0.96	12.80	1.83	1.03	-3.76
Papuan-Mbuti	Abkhasian	Caucasus-S	-0.96	2.50	-7.14	0.56	0.34	-1.28
Papuan-Saami	Abkhasian	Caucasus-S	0.11	0.15	0.14	0.21	0.73	0.18
Papuan-Somali	Abkhasian	Caucasus-S	-12.75	0.53	1.70	0.64	0.49	-1.24
Papuan-Surui	Abkhasian	Caucasus-S	24.00	1.45	7.38	2.82	0.97	-10.67
Papuan-Yoruba	Abkhasian	Caucasus-S	-1.40	1.69	11.80	0.72	0.38	-1.28
Average:	————	————	0.11	0.60	0.14	0.48	0.51	0.18
Papuan-Australian	Armenian	Caucasus-S	-44.00	-3.14	-11.00	-3.38	4.89	-2.44
Papuan-Han	Armenian	Caucasus-S	-2.50	-1.88	-1.58	-1.43	1.25	-0.91
Papuan-Ju-hoan-North	Armenian	Caucasus-S	-0.45	0.68	5.67	0.18	0.11	-0.74
Mbuti-Ju-hoan-North	Armenian	Caucasus-S	-2.59	-8.80	-4.00	-7.33	-3.38	-2.32
Papuan-Karitiana	Armenian	Caucasus-S	0.18	1.44	0.18	0.33	1.00	0.14
Papuan-Mbuti	Armenian	Caucasus-S	-1.02	2.48	-8.67	0.58	0.35	-1.37
Papuan-Saami	Armenian	Caucasus-S	0.37	0.45	0.42	0.54	0.95	0.51
Papuan-Somali	Armenian	Caucasus-S	-0.57	0.31	-10.00	0.42	0.27	-0.27
Papuan-Surui	Armenian	Caucasus-S	-0.19	-1.00	-0.22	-0.33	0.88	-0.17
Papuan-Yoruba	Armenian	Caucasus-S	-1.17	1.74	NaN	0.71	0.35	-1.10
Average:	————	————	0.27	0.48	NaN	0.46	0.56	0.33
Papuan-Australian	Azeri	Caucasus-S	3.32	9.00	3.94	10.50	2.25	15.75
Papuan-Han	Azeri	Caucasus-S	1.60	1.68	1.76	1.88	1.08	2.27

Table S4 continued from previous page

Papuan-Ju-hoan-North	Azeri	Caucasus-S	-2.38	0.83	1.52	0.33	0.21	-19.00
Mbuti-Ju-hoan-North	Azeri	Caucasus-S	-2.15	-5.38	-2.87	-4.30	-2.69	-2.05
Papuan-Karitiana	Azeri	Caucasus-S	-18.00	0.90	-54.00	1.93	1.06	-2.57
Papuan-Mbuti	Azeri	Caucasus-S	-2.25	1.80	5.54	0.66	0.43	-4.00
Papuan-Saami	Azeri	Caucasus-S	0.41	0.49	0.46	0.59	0.92	0.55
Papuan-Somali	Azeri	Caucasus-S	2.62	0.66	1.33	0.76	0.61	-22.25
Papuan-Surui	Azeri	Caucasus-S	NaN	1.46	9.89	3.30	1.01	-7.42
Papuan-Yoruba	Azeri	Caucasus-S	-3.00	1.47	3.57	0.77	0.43	-2.50
Average:	————-	————-	NaN	0.72	0.46	0.62	0.52	0.55
Papuan-Australian	Georgian	Caucasus-S	5.18	-28.50	7.12	-28.50	2.85	-11.40
Papuan-Han	Georgian	Caucasus-S	4.38	5.70	9.50	19.00	1.21	-8.14
Papuan-Ju-hoan-North	Georgian	Caucasus-S	-0.31	0.62	-13.00	0.14	0.08	-0.48
Mbuti-Ju-hoan-North	Georgian	Caucasus-S	-3.62	-47.00	-6.71	-15.67	-5.88	-3.62
Papuan-Karitiana	Georgian	Caucasus-S	-1.18	0.94	-1.27	11.00	1.27	-0.69
Papuan-Mbuti	Georgian	Caucasus-S	-0.94	2.50	-7.14	0.56	0.34	-1.35
Papuan-Saami	Georgian	Caucasus-S	0.20	0.26	0.24	0.34	0.80	0.31
Papuan-Somali	Georgian	Caucasus-S	-2.11	0.46	2.38	0.58	0.41	-0.70
Papuan-Surui	Georgian	Caucasus-S	-1.85	2.18	-2.54	-15.25	1.05	-1.36
Papuan-Yoruba	Georgian	Caucasus-S	-1.13	1.83	-53.00	0.70	0.35	-1.08
Average:	————-	————-	0.20	0.57	0.24	0.46	0.40	0.31
Papuan-Australian	Iranian	Caucasus-SE	16.00	-5.33	NaN	-4.80	3.43	-4.36
Papuan-Han	Iranian	Caucasus-SE	3.32	4.20	5.25	7.88	1.17	-31.50
Papuan-Ju-hoan-North	Iranian	Caucasus-SE	-1.12	0.76	1.87	0.26	0.17	-3.50
Mbuti-Ju-hoan-North	Iranian	Caucasus-SE	-1.77	-3.90	-2.44	-3.55	-2.17	-1.56
Papuan-Karitiana	Iranian	Caucasus-SE	-0.82	0.90	-0.87	-13.50	1.17	-0.54
Papuan-Mbuti	Iranian	Caucasus-SE	-1.37	2.03	29.50	0.60	0.38	-2.11
Papuan-Saami	Iranian	Caucasus-SE	0.52	0.61	0.58	0.69	0.95	0.66
Papuan-Somali	Iranian	Caucasus-SE	2.59	0.66	1.31	0.75	0.61	-88.00
Papuan-Surui	Iranian	Caucasus-SE	-1.11	2.88	-1.40	-3.06	1.04	-0.88

Table S4 continued from previous page

Papuan-Yoruba	Iranian	Caucasus-SE	-1.70	1.62	7.00	0.73	0.39	-1.62
Average:	————-	————-	0.52	0.73	NaN	0.61	0.50	0.66
Papuan-Australian	Assyrian	Caucasus-SE	-44.00	-3.14	-11.00	-2.93	5.50	-2.75
Papuan-Han	Assyrian	Caucasus-SE	-21.00	-8.40	-5.25	-3.50	1.24	-2.00
Papuan-Ju-hoan-North	Assyrian	Caucasus-SE	0.11	-7.00	0.33	-0.10	-0.05	0.15
Mbuti-Ju-hoan-North	Assyrian	Caucasus-SE	-1.95	-4.56	-2.56	-4.10	-2.41	-2.05
Papuan-Karitiana	Assyrian	Caucasus-SE	0.15	1.57	0.16	0.28	0.79	0.13
Papuan-Mbuti	Assyrian	Caucasus-SE	-0.32	-4.17	-0.76	0.39	0.21	-0.38
Papuan-Saami	Assyrian	Caucasus-SE	0.37	0.45	0.42	0.54	0.96	0.51
Papuan-Somali	Assyrian	Caucasus-SE	-0.04	0.04	-0.10	0.07	0.04	-0.02
Papuan-Surui	Assyrian	Caucasus-SE	-0.16	-0.68	-0.18	-0.25	1.08	-0.14
Papuan-Yoruba	Assyrian	Caucasus-SE	-0.41	5.80	-1.16	0.57	0.23	-0.39
Average:	————-	————-	0.21	0.25	0.31	0.37	0.44	0.26
Papuan-Australian	Kurd	Caucasus-SE	5.00	-27.50	7.86	-18.33	2.75	-7.86
Papuan-Han	Kurd	Caucasus-SE	11.25	NaN	-15.00	-5.62	1.12	-2.37
Papuan-Ju-hoan-North	Kurd	Caucasus-SE	0.57	1.09	0.83	-14.60	-1.11	0.65
Mbuti-Ju-hoan-North	Kurd	Caucasus-SE	-1.44	-2.77	-1.89	-2.57	-1.80	-1.38
Papuan-Karitiana	Kurd	Caucasus-SE	0.00	0.00	0.00	0.00	NaN	0.00
Papuan-Mbuti	Kurd	Caucasus-SE	0.30	0.58	0.43	7.50	-0.90	0.34
Papuan-Saami	Kurd	Caucasus-SE	0.49	0.58	0.55	0.67	0.98	0.64
Papuan-Somali	Kurd	Caucasus-SE	0.14	-0.26	0.27	-0.47	-0.20	0.09
Papuan-Surui	Kurd	Caucasus-SE	-0.44	NaN	-0.52	-0.82	0.87	-0.36
Papuan-Yoruba	Kurd	Caucasus-SE	0.32	0.65	0.46	1.96	-0.92	0.32
Average:	————-	————-	0.30	NaN	0.42	0.34	NaN	0.34
Papuan-Australian	Turkish	Caucasus-W	24.00	-4.80	-48.00	-4.80	3.69	-4.00
Papuan-Han	Turkish	Caucasus-W	1.62	1.70	1.81	1.91	1.06	2.33
Papuan-Ju-hoan-North	Turkish	Caucasus-W	19.67	0.89	1.34	0.44	0.30	3.69
Mbuti-Ju-hoan-North	Turkish	Caucasus-W	-3.07	-15.33	-5.11	-11.50	-3.83	-3.07
Papuan-Karitiana	Turkish	Caucasus-W	-7.29	0.91	-8.50	2.04	1.04	-1.96

Table S4 continued from previous page

Papuan-Mbuti	Turkish	Caucasus-W	-10.56	1.48	2.57	0.72	0.50	23.75
Papuan-Saami	Turkish	Caucasus-W	0.39	0.48	0.45	0.57	0.94	0.53
Papuan-Somali	Turkish	Caucasus-W	2.63	0.68	1.35	0.78	0.63	-46.00
Papuan-Surui	Turkish	Caucasus-W	-3.45	1.68	-5.75	7.67	0.97	-2.03
Papuan-Yoruba	Turkish	Caucasus-W	-19.60	1.36	2.39	0.84	0.51	-10.89
Average:	————-	————-	0.39	0.74	0.45	0.67	0.64	0.53
Papuan-Australian	Italian-North	Europe-S	-14.33	-2.69	-7.17	-2.69	7.17	-2.26
Papuan-Han	Italian-North	Europe-S	4.14	5.27	7.25	14.50	1.16	-7.25
Papuan-Ju-hoan-North	Italian-North	Europe-S	-1.08	0.74	1.75	0.26	0.16	-2.80
Mbuti-Ju-hoan-North	Italian-North	Europe-S	-14.75	7.38	59.00	9.83	-59.00	-11.80
Papuan-Karitiana	Italian-North	Europe-S	-13.75	0.90	-27.50	1.96	1.04	-2.50
Papuan-Mbuti	Italian-North	Europe-S	-2.59	1.70	4.41	0.66	0.43	-5.36
Papuan-Saami	Italian-North	Europe-S	-3.43	10.67	-32.00	2.82	1.10	4.00
Papuan-Somali	Italian-North	Europe-S	-0.73	0.35	24.00	0.47	0.30	-0.35
Papuan-Surui	Italian-North	Europe-S	-8.20	1.55	-82.00	4.32	0.99	-3.42
Papuan-Yoruba	Italian-North	Europe-S	-2.26	1.52	4.38	0.75	0.41	-2.12
Average:	————-	————-	NaN	0.66	NaN	0.53	0.46	NaN
Papuan-Australian	Italian-South	Europe-S	8.50	-7.29	17.00	-6.38	3.19	-12.75
Papuan-Han	Italian-South	Europe-S	4.33	6.50	10.40	26.00	1.08	-5.78
Papuan-Ju-hoan-North	Italian-South	Europe-S	0.31	1.71	0.67	-0.41	-0.20	0.35
Mbuti-Ju-hoan-North	Italian-South	Europe-S	-3.64	-25.50	-6.38	-12.75	-4.64	-3.64
Papuan-Karitiana	Italian-South	Europe-S	-1.28	0.82	-1.39	5.33	1.03	-0.73
Papuan-Mbuti	Italian-South	Europe-S	-0.18	-1.07	-0.38	0.29	0.15	-0.20
Papuan-Saami	Italian-South	Europe-S	-0.36	-0.60	-0.50	-1.23	1.03	-1.19
Papuan-Somali	Italian-South	Europe-S	0.27	-1.05	0.47	-5.50	-0.67	0.18
Papuan-Surui	Italian-South	Europe-S	-0.78	3.25	-0.95	-1.86	0.87	-0.66
Papuan-Yoruba	Italian-South	Europe-S	-0.13	-0.92	-0.28	0.33	0.11	-0.12
Average:	————-	————-	0.29	0.82	0.57	0.31	0.37	0.27
Papuan-Australian	Sicilian	Europe-S	5.70	-19.00	9.50	-14.25	3.00	-11.40

Table S4 continued from previous page

Papuan-Han	Sicilian	Europe-S	12.00	48.00	-24.00	-8.00	1.14	-3.69
Papuan-Ju-hoan-North	Sicilian	Europe-S	0.31	1.47	0.64	-0.46	-0.22	0.38
Mbuti-Ju-hoan-North	Sicilian	Europe-S	-2.82	-9.60	-4.00	-6.86	-3.69	-2.53
Papuan-Karitiana	Sicilian	Europe-S	-0.27	0.81	-0.28	-0.81	1.18	-0.20
Papuan-Mbuti	Sicilian	Europe-S	-0.14	-0.62	-0.27	0.27	0.12	-0.16
Papuan-Saami	Sicilian	Europe-S	-0.25	-0.40	-0.34	-0.69	1.19	-0.61
Papuan-Somali	Sicilian	Europe-S	0.27	-0.91	0.48	-3.00	-0.64	0.18
Papuan-Surui	Sicilian	Europe-S	-0.40	-5.40	-0.47	-0.73	0.96	-0.35
Papuan-Yoruba	Sicilian	Europe-S	-0.03	-0.13	-0.06	0.12	0.03	-0.03
Average:	————	————	0.29	0.81	0.56	0.19	0.37	0.28
Papuan-Australian	Cretan	Europe-SE	4.47	22.33	6.09	NaN	2.68	Infinity
Papuan-Han	Cretan	Europe-SE	2.92	3.33	4.12	5.83	1.25	-14.00
Papuan-Ju-hoan-North	Cretan	Europe-SE	-0.11	0.60	-0.50	0.08	0.04	-0.14
Mbuti-Ju-hoan-North	Cretan	Europe-SE	-2.26	-5.38	-3.07	-5.38	-2.26	-1.59
Papuan-Karitiana	Cretan	Europe-SE	-14.00	0.93	-28.00	2.15	1.17	-2.24
Papuan-Mbuti	Cretan	Europe-SE	-0.59	10.00	-1.74	0.55	0.30	-0.63
Papuan-Saami	Cretan	Europe-SE	0.04	0.06	0.05	0.08	0.26	0.07
Papuan-Somali	Cretan	Europe-SE	-0.04	0.04	-0.10	0.06	0.03	-0.02
Papuan-Surui	Cretan	Europe-SE	-1.37	2.55	-1.81	-4.67	1.24	-1.00
Papuan-Yoruba	Cretan	Europe-SE	-0.74	3.00	-3.00	0.71	0.32	-0.66
Average:	————	————	0.04	0.41	0.05	NaN	0.19	0.07
Papuan-Australian	Cypriot	Europe-SE	5.50	-18.33	9.17	-13.75	2.89	-18.33
Papuan-Han	Cypriot	Europe-SE	15.00	NaN	-15.00	-7.50	1.12	-3.21
Papuan-Ju-hoan-North	Cypriot	Europe-SE	0.37	1.13	0.66	-0.88	-0.33	0.45
Mbuti-Ju-hoan-North	Cypriot	Europe-SE	-4.08	NaN	-7.00	-16.33	-4.45	-2.88
Papuan-Karitiana	Cypriot	Europe-SE	0.03	-0.67	0.03	0.07	1.00	0.03
Papuan-Mbuti	Cypriot	Europe-SE	-0.03	-0.09	-0.05	0.08	0.03	-0.03
Papuan-Saami	Cypriot	Europe-SE	0.20	0.26	0.24	0.35	0.91	0.32
Papuan-Somali	Cypriot	Europe-SE	0.45	5.56	0.66	1.92	-16.67	0.34

Table S4 continued from previous page

Papuan-Surui	Cypriot	Europe-SE	-0.28	-2.00	-0.32	-0.49	0.95	-0.24
Papuan-Yoruba	Cypriot	Europe-SE	0.01	0.03	0.02	-0.06	-0.01	0.01
Average:	————	————	0.21	NaN	0.32	0.17	0.72	0.23
Papuan-Australian	Greek	Europe-SE	15.67	-4.70	-47.00	-4.27	3.92	-3.92
Papuan-Han	Greek	Europe-SE	5.00	6.88	13.75	55.00	1.15	-6.11
Papuan-Ju-hoan-North	Greek	Europe-SE	-0.60	0.72	3.00	0.21	0.13	-1.00
Mbuti-Ju-hoan-North	Greek	Europe-SE	-5.20	26.00	-10.40	NaN	-8.67	-4.33
Papuan-Karitiana	Greek	Europe-SE	-1.61	0.90	-1.76	3.70	1.03	-0.92
Papuan-Mbuti	Greek	Europe-SE	-1.44	2.07	20.67	0.63	0.39	-2.00
Papuan-Saami	Greek	Europe-SE	-0.56	-1.05	-0.81	-2.59	1.13	-1.76
Papuan-Somali	Greek	Europe-SE	-0.49	0.30	-3.80	0.42	0.26	-0.25
Papuan-Surui	Greek	Europe-SE	-2.17	1.97	-3.00	63.00	0.97	-1.54
Papuan-Yoruba	Greek	Europe-SE	-1.40	1.76	15.00	0.75	0.38	-1.28
Average:	————	————	NaN	0.64	NaN	NaN	0.43	NaN
Papuan-Australian	Maltese	Europe-SE	-2.91	-1.33	-2.29	-1.28	-10.67	-1.19
Papuan-Han	Maltese	Europe-SE	-2.50	-2.00	-1.67	-1.36	1.25	-0.97
Papuan-Ju-hoan-North	Maltese	Europe-SE	0.52	1.09	0.79	-4.13	-0.77	0.58
Mbuti-Ju-hoan-North	Maltese	Europe-SE	NaN	5.17	10.33	6.20	20.67	-31.00
Papuan-Karitiana	Maltese	Europe-SE	-0.11	0.60	-0.12	-0.27	2.00	-0.09
Papuan-Mbuti	Maltese	Europe-SE	0.11	0.27	0.18	-0.54	-0.16	0.12
Papuan-Saami	Maltese	Europe-SE	0.08	0.11	0.09	0.14	0.77	0.13
Papuan-Somali	Maltese	Europe-SE	0.39	-7.40	0.59	3.36	-2.18	0.27
Papuan-Surui	Maltese	Europe-SE	-0.54	16.00	-0.64	-1.07	0.97	-0.46
Papuan-Yoruba	Maltese	Europe-SE	0.18	0.45	0.28	7.67	-0.31	0.17
Average:	————	————	NaN	0.36	0.39	0.14	0.87	0.26
Papuan-Australian	Canary-Islander	Europe-SW	-26.00	-3.47	-10.40	-4.00	4.73	-3.06
Papuan-Han	Canary-Islander	Europe-SW	-20.50	-6.83	-4.56	-4.10	1.11	-1.64
Papuan-Ju-hoan-North	Canary-Islander	Europe-SW	0.63	1.09	0.87	5.11	-2.20	0.71
Mbuti-Ju-hoan-North	Canary-Islander	Europe-SW	-3.57	-25.00	-5.56	-16.67	-12.50	-4.17

Table S4 continued from previous page

Papuan-Karitiana	Canary-Islander	Europe-SW	-1.50	0.88	-1.64	3.27	0.92	-1.00
Papuan-Mbuti	Canary-Islander	Europe-SW	0.35	0.63	0.48	2.71	-1.39	0.39
Papuan-Saami	Canary-Islander	Europe-SW	-0.73	-1.53	-1.13	-8.67	1.11	-2.74
Papuan-Somali	Canary-Islander	Europe-SW	0.48	4.91	0.69	1.93	-54.00	0.37
Papuan-Surui	Canary-Islander	Europe-SW	-4.47	1.69	-9.50	5.07	0.90	-2.92
Papuan-Yoruba	Canary-Islander	Europe-SW	0.40	0.72	0.55	1.41	-2.56	0.40
Average:	————-	————-	0.47	0.74	0.65	NaN	0.91	0.47
Papuan-Australian	Spanish-Balear	Europe-SW	-3.09	-1.48	-2.43	-1.70	Infinity	-1.89
Papuan-Han	Spanish-Balear	Europe-SW	-14.00	-7.00	-3.82	-3.00	1.31	-2.33
Papuan-Ju-hoan-North	Spanish-Balear	Europe-SW	0.14	-1.33	0.50	-0.11	-0.06	0.19
Mbuti-Ju-hoan-North	Spanish-Balear	Europe-SW	-16.00	9.14	64.00	10.67	Infinity	-8.00
Papuan-Karitiana	Spanish-Balear	Europe-SW	-1.00	0.94	-1.03	-15.50	1.29	-0.61
Papuan-Mbuti	Spanish-Balear	Europe-SW	-0.69	3.91	-2.69	0.53	0.31	-0.86
Papuan-Saami	Spanish-Balear	Europe-SW	-1.81	-13.00	-4.11	3.39	1.08	5.57
Papuan-Somali	Spanish-Balear	Europe-SW	-0.31	0.21	-1.62	0.31	0.18	-0.16
Papuan-Surui	Spanish-Balear	Europe-SW	-2.59	2.00	-3.89	70.00	1.08	-2.19
Papuan-Yoruba	Spanish-Balear	Europe-SW	-0.73	1.86	-4.56	0.60	0.28	-0.69
Average:	————-	————-	0.14	0.58	0.50	0.48	0.26	0.19
Papuan-Australian	Spanish-North	Europe-SW	8.83	-7.57	17.67	-10.60	2.65	-10.60
Papuan-Han	Spanish-North	Europe-SW	4.58	6.88	11.00	NaN	1.17	-3.93
Papuan-Ju-hoan-North	Spanish-North	Europe-SW	-9.80	0.83	1.32	0.37	0.25	3.77
Mbuti-Ju-hoan-North	Spanish-North	Europe-SW	-7.12	14.25	-28.50	57.00	-9.50	-4.75
Papuan-Karitiana	Spanish-North	Europe-SW	3.67	0.90	3.35	1.51	0.97	-19.25
Papuan-Mbuti	Spanish-North	Europe-SW	-7.83	1.52	2.69	0.71	0.49	31.33
Papuan-Saami	Spanish-North	Europe-SW	5.50	2.37	2.91	1.71	1.06	1.92
Papuan-Somali	Spanish-North	Europe-SW	-2.11	0.48	2.67	0.61	0.42	-0.74
Papuan-Surui	Spanish-North	Europe-SW	5.45	1.33	3.76	2.27	0.94	-54.50
Papuan-Yoruba	Spanish-North	Europe-SW	-6.29	1.38	2.67	0.79	0.47	-6.29
Average:	————-	————-	NaN	0.73	NaN	NaN	0.59	NaN

Table S4 continued from previous page

Papuan-Australian	BedouinB	NearEast-E	9.80	-6.12	24.50	-5.44	3.27	-4.90
Papuan-Han	BedouinB	NearEast-E	-0.40	-0.35	-0.32	-0.30	2.00	-0.24
Papuan-Ju-hoan-North	BedouinB	NearEast-E	0.78	1.02	0.93	1.60	3.17	0.83
Mbuti-Ju-hoan-North	BedouinB	NearEast-E	-2.81	-11.25	-4.09	-9.00	-4.09	-2.50
Papuan-Karitiana	BedouinB	NearEast-E	0.54	1.01	0.55	0.71	0.96	0.48
Papuan-Mbuti	BedouinB	NearEast-E	0.62	0.83	0.74	1.28	2.31	0.65
Papuan-Saami	BedouinB	NearEast-E	0.56	0.64	0.62	0.73	0.96	0.70
Papuan-Somali	BedouinB	NearEast-E	0.79	1.25	0.90	1.15	1.39	0.70
Papuan-Surui	BedouinB	NearEast-E	0.39	0.65	0.41	0.48	1.02	0.36
Papuan-Yoruba	BedouinB	NearEast-E	0.63	0.86	0.76	1.14	2.30	0.63
Average:	————-	————-	0.62	0.75	0.70	0.64	0.96	0.62
Papuan-Australian	Saudi	NearEast-E	3.65	15.50	4.77	20.67	2.38	31.00
Papuan-Han	Saudi	NearEast-E	-1.05	-0.87	-0.78	-0.70	1.50	-0.55
Papuan-Ju-hoan-North	Saudi	NearEast-E	0.74	1.02	0.91	1.84	6.83	0.80
Mbuti-Ju-hoan-North	Saudi	NearEast-E	-5.20	26.00	-10.40	52.00	-6.50	-3.25
Papuan-Karitiana	Saudi	NearEast-E	0.57	1.04	0.57	0.75	1.00	0.53
Papuan-Mbuti	Saudi	NearEast-E	0.54	0.77	0.67	1.41	4.13	0.57
Papuan-Saami	Saudi	NearEast-E	0.55	0.63	0.61	0.72	0.97	0.70
Papuan-Somali	Saudi	NearEast-E	0.75	1.33	0.87	1.20	1.51	0.64
Papuan-Surui	Saudi	NearEast-E	0.35	0.62	0.38	0.45	1.00	0.34
Papuan-Yoruba	Saudi	NearEast-E	0.56	0.82	0.70	1.17	3.69	0.55
Average:	————-	————-	0.58	0.71	0.67	0.64	0.99	0.59
Papuan-Australian	Yemeni	NearEast-E	-2.43	-1.31	-2.00	-1.36	-11.33	-1.26
Papuan-Han	Yemeni	NearEast-E	0.14	0.13	0.12	0.12	0.41	0.10
Papuan-Ju-hoan-North	Yemeni	NearEast-E	0.88	1.01	0.97	1.24	1.53	0.93
Mbuti-Ju-hoan-North	Yemeni	NearEast-E	-2.81	-9.00	-4.09	-7.50	-3.21	-2.81
Papuan-Karitiana	Yemeni	NearEast-E	0.63	1.03	0.64	0.79	0.95	0.55
Papuan-Mbuti	Yemeni	NearEast-E	0.79	0.92	0.87	1.11	1.35	0.83
Papuan-Saami	Yemeni	NearEast-E	0.67	0.74	0.72	0.80	0.97	0.77

Table S4 continued from previous page

Papuan-Somali	Yemeni	NearEast-E	0.81	1.21	0.91	1.13	1.33	0.74
Papuan-Surui	Yemeni	NearEast-E	0.51	0.75	0.53	0.61	0.95	0.47
Papuan-Yoruba	Yemeni	NearEast-E	0.79	0.93	0.87	1.06	1.35	0.81
Average:	————-	————-	0.65	0.69	0.70	0.58	0.82	0.65
Papuan-Australian	Druze	NearEast-N	-10.00	-2.35	-5.00	-2.35	8.00	-2.50
Papuan-Han	Druze	NearEast-N	-0.71	-0.61	-0.55	-0.50	1.55	-0.39
Papuan-Ju-hoan-North	Druze	NearEast-N	0.50	1.09	0.77	-3.41	-0.70	0.57
Mbuti-Ju-hoan-North	Druze	NearEast-N	-2.88	-11.50	-4.18	-7.67	-3.54	-2.56
Papuan-Karitiana	Druze	NearEast-N	0.40	1.15	0.41	0.60	0.97	0.34
Papuan-Mbuti	Druze	NearEast-N	0.17	0.39	0.27	-1.69	-0.30	0.19
Papuan-Saami	Druze	NearEast-N	0.39	0.48	0.45	0.58	0.96	0.54
Papuan-Somali	Druze	NearEast-N	0.45	9.40	0.65	2.24	-5.88	0.33
Papuan-Surui	Druze	NearEast-N	0.06	0.18	0.07	0.09	1.50	0.06
Papuan-Yoruba	Druze	NearEast-N	0.21	0.50	0.32	3.86	-0.38	0.20
Average:	————-	————-	0.31	0.39	0.42	0.42	0.97	0.32
Papuan-Australian	Lebanese	NearEast-N	16.00	-4.80	-48.00	-4.80	4.00	-4.80
Papuan-Han	Lebanese	NearEast-N	-1.69	-1.42	-1.23	-1.04	1.29	-0.77
Papuan-Ju-hoan-North	Lebanese	NearEast-N	0.61	1.11	0.87	16.80	-1.45	0.68
Mbuti-Ju-hoan-North	Lebanese	NearEast-N	-2.15	-5.38	-2.87	-4.30	-2.69	-1.95
Papuan-Karitiana	Lebanese	NearEast-N	0.38	1.09	0.39	0.57	0.95	0.33
Papuan-Mbuti	Lebanese	NearEast-N	0.32	0.61	0.46	4.17	-1.09	0.35
Papuan-Saami	Lebanese	NearEast-N	0.44	0.52	0.49	0.62	0.98	0.59
Papuan-Somali	Lebanese	NearEast-N	0.53	3.05	0.73	1.78	7.11	0.41
Papuan-Surui	Lebanese	NearEast-N	0.09	0.23	0.10	0.13	1.12	0.08
Papuan-Yoruba	Lebanese	NearEast-N	0.35	0.67	0.50	1.60	-1.33	0.35
Average:	————-	————-	0.39	0.51	0.51	0.44	0.96	0.40
Papuan-Australian	BedouinA	NearEast-S	16.33	-5.44	NaN	-4.90	3.50	-4.90
Papuan-Han	BedouinA	NearEast-S	-0.20	-0.18	-0.17	-0.16	Infinity	-0.13
Papuan-Ju-hoan-North	BedouinA	NearEast-S	0.84	1.02	0.95	1.39	2.04	0.88

Table S4 continued from previous page

Mbuti-Ju-hoan-North	BedouinA	NearEast-S	-1.38	-2.57	-1.80	-2.40	-1.71	-1.29
Papuan-Karitiana	BedouinA	NearEast-S	0.52	1.05	0.53	0.70	0.94	0.46
Papuan-Mbuti	BedouinA	NearEast-S	0.71	0.88	0.81	1.18	1.63	0.74
Papuan-Saami	BedouinA	NearEast-S	0.58	0.66	0.63	0.74	0.98	0.71
Papuan-Somali	BedouinA	NearEast-S	0.78	1.29	0.89	1.18	1.42	0.68
Papuan-Surui	BedouinA	NearEast-S	0.33	0.60	0.35	0.42	1.05	0.31
Papuan-Yoruba	BedouinA	NearEast-S	0.72	0.91	0.83	1.10	1.60	0.72
Average:	———————-	———————-	0.64	0.76	NaN	0.62	0.96	0.64
Papuan-Australian	Jordanian	NearEast-S	12.75	-6.38	51.00	-5.67	3.92	-4.25
Papuan-Han	Jordanian	NearEast-S	-1.69	-1.42	-1.23	-1.04	1.42	-0.77
Papuan-Ju-hoan-North	Jordanian	NearEast-S	0.78	1.05	0.94	1.81	4.65	0.84
Mbuti-Ju-hoan-North	Jordanian	NearEast-S	-1.25	-2.06	-1.52	-1.84	-1.52	-1.09
Papuan-Karitiana	Jordanian	NearEast-S	0.48	1.08	0.49	0.68	0.98	0.43
Papuan-Mbuti	Jordanian	NearEast-S	0.61	0.84	0.74	1.34	2.74	0.64
Papuan-Saami	Jordanian	NearEast-S	0.49	0.58	0.55	0.67	0.98	0.64
Papuan-Somali	Jordanian	NearEast-S	0.68	1.59	0.84	1.31	1.90	0.56
Papuan-Surui	Jordanian	NearEast-S	0.15	0.34	0.16	0.21	1.00	0.14
Papuan-Yoruba	Jordanian	NearEast-S	0.62	0.87	0.75	1.17	2.58	0.61
Average:	———————-	———————-	0.55	0.66	0.64	0.52	0.99	0.55
Papuan-Australian	Palestinian	NearEast-S	12.00	-5.33	NaN	-4.36	3.69	-4.36
Papuan-Han	Palestinian	NearEast-S	-0.26	-0.24	-0.22	-0.20	4.50	-0.18
Papuan-Ju-hoan-North	Palestinian	NearEast-S	0.77	1.04	0.93	1.76	4.11	0.81
Mbuti-Ju-hoan-North	Palestinian	NearEast-S	-1.95	-4.10	-2.56	-3.73	-2.41	-1.86
Papuan-Karitiana	Palestinian	NearEast-S	0.54	1.05	0.54	0.72	0.99	0.49
Papuan-Mbuti	Palestinian	NearEast-S	0.60	0.82	0.72	1.34	2.64	0.62
Papuan-Saami	Palestinian	NearEast-S	0.53	0.62	0.59	0.70	0.97	0.68
Papuan-Somali	Palestinian	NearEast-S	0.69	1.48	0.84	1.28	1.74	0.58
Papuan-Surui	Palestinian	NearEast-S	0.33	0.59	0.35	0.42	1.05	0.31
Papuan-Yoruba	Palestinian	NearEast-S	0.62	0.86	0.75	1.16	2.50	0.61

Table S4 continued from previous page

Average:	————-	————-	0.58	0.72	NaN	0.61	0.98	0.59
Papuan-Australian	Syrian	NearEast-S	3.14	7.33	3.67	9.43	1.94	9.43
Papuan-Han	Syrian	NearEast-S	9.60	24.00	-48.00	-9.60	1.14	-4.80
Papuan-Ju-hoan-North	Syrian	NearEast-S	0.77	1.05	0.94	1.82	4.62	0.82
Mbuti-Ju-hoan-North	Syrian	NearEast-S	-1.18	-2.06	-1.43	-1.94	-1.37	-1.14
Papuan-Karitiana	Syrian	NearEast-S	0.20	1.36	0.21	0.37	1.25	0.18
Papuan-Mbuti	Syrian	NearEast-S	0.60	0.83	0.73	1.36	2.66	0.63
Papuan-Saami	Syrian	NearEast-S	0.48	0.57	0.54	0.66	0.99	0.65
Papuan-Somali	Syrian	NearEast-S	0.68	1.60	0.84	1.31	1.89	0.55
Papuan-Surui	Syrian	NearEast-S	-0.07	-0.24	-0.08	-0.11	0.60	-0.07
Papuan-Yoruba	Syrian	NearEast-S	0.61	0.86	0.74	1.18	2.71	0.60
Average:	————-	————-	0.56	0.75	0.67	0.52	0.80	0.57
Papuan-Australian	Algerian	NorthAfrica	-0.38	-0.28	-0.35	-0.27	-0.50	-0.24
Papuan-Han	Algerian	NorthAfrica	0.20	0.19	0.18	0.17	0.44	0.15
Papuan-Ju-hoan-North	Algerian	NorthAfrica	0.92	1.02	0.99	1.19	1.37	0.96
Mbuti-Ju-hoan-North	Algerian	NorthAfrica	-0.48	-0.65	-0.55	-0.61	-0.52	-0.46
Papuan-Karitiana	Algerian	NorthAfrica	0.62	0.99	0.63	0.76	0.94	0.57
Papuan-Mbuti	Algerian	NorthAfrica	0.84	0.94	0.90	1.09	1.23	0.87
Papuan-Saami	Algerian	NorthAfrica	0.65	0.72	0.70	0.79	0.95	0.77
Papuan-Somali	Algerian	NorthAfrica	0.86	1.16	0.95	1.12	1.24	0.80
Papuan-Surui	Algerian	NorthAfrica	0.49	0.73	0.51	0.58	0.97	0.46
Papuan-Yoruba	Algerian	NorthAfrica	0.86	0.96	0.92	1.05	1.22	0.87
Average:	————-	————-	0.68	0.75	0.72	0.57	0.82	0.68
Papuan-Australian	Egyptian	NorthAfrica-a	NaN	-3.38	-14.67	-3.38	4.40	-2.93
Papuan-Han	Egyptian	NorthAfrica-a	0.12	0.12	0.11	0.10	0.43	0.09
Papuan-Ju-hoan-North	Egyptian	NorthAfrica-a	0.88	1.02	0.97	1.25	1.58	0.92
Mbuti-Ju-hoan-North	Egyptian	NorthAfrica-a	-0.88	-1.36	-1.07	-1.30	-0.97	-0.88
Papuan-Karitiana	Egyptian	NorthAfrica-a	0.65	1.02	0.66	0.80	0.97	0.59
Papuan-Mbuti	Egyptian	NorthAfrica-a	0.78	0.92	0.86	1.12	1.35	0.81

Table S4 continued from previous page

Papuan-Saami	Egyptian	NorthAfrica-a	0.65	0.72	0.70	0.79	0.98	0.76
Papuan-Somali	Egyptian	NorthAfrica-a	0.82	1.18	0.91	1.11	1.27	0.75
Papuan-Surui	Egyptian	NorthAfrica-a	0.46	0.71	0.49	0.56	0.99	0.43
Papuan-Yoruba	Egyptian	NorthAfrica-a	0.80	0.93	0.88	1.06	1.34	0.80
Average:	————-	————-	NaN	0.68	0.70	0.56	0.84	0.64
Papuan-Australian	Libyan	NorthAfrica-b	-6.33	-2.00	-4.22	-2.11	9.50	-1.81
Papuan-Han	Libyan	NorthAfrica-b	-0.05	-0.05	-0.04	-0.04	-1.00	-0.04
Papuan-Ju-hoan-North	Libyan	NorthAfrica-b	0.89	1.01	0.97	1.21	1.47	0.92
Mbuti-Ju-hoan-North	Libyan	NorthAfrica-b	-2.00	-4.20	-2.62	-3.82	-2.21	-1.75
Papuan-Karitiana	Libyan	NorthAfrica-b	0.59	0.99	0.60	0.75	0.95	0.56
Papuan-Mbuti	Libyan	NorthAfrica-b	0.80	0.92	0.87	1.09	1.30	0.82
Papuan-Saami	Libyan	NorthAfrica-b	0.64	0.72	0.69	0.80	1.00	0.78
Papuan-Somali	Libyan	NorthAfrica-b	0.84	1.14	0.92	1.09	1.22	0.76
Papuan-Surui	Libyan	NorthAfrica-b	0.41	0.66	0.43	0.50	0.97	0.40
Papuan-Yoruba	Libyan	NorthAfrica-b	0.82	0.94	0.89	1.05	1.28	0.81
Average:	————-	————-	0.71	0.85	0.77	0.69	0.97	0.72
Papuan-Australian	Tunisian	NorthAfrica-c	12.00	-6.00	48.00	-4.80	4.36	-3.43
Papuan-Han	Tunisian	NorthAfrica-c	-0.11	-0.10	-0.09	-0.08	-1.33	-0.07
Papuan-Ju-hoan-North	Tunisian	NorthAfrica-c	0.90	1.00	0.97	1.18	1.39	0.93
Mbuti-Ju-hoan-North	Tunisian	NorthAfrica-c	-0.52	-0.72	-0.60	-0.70	-0.55	-0.52
Papuan-Karitiana	Tunisian	NorthAfrica-c	0.63	1.00	0.64	0.78	0.95	0.58
Papuan-Mbuti	Tunisian	NorthAfrica-c	0.83	0.93	0.89	1.08	1.24	0.85
Papuan-Saami	Tunisian	NorthAfrica-c	0.63	0.71	0.68	0.78	0.96	0.75
Papuan-Somali	Tunisian	NorthAfrica-c	0.85	1.14	0.93	1.09	1.22	0.79
Papuan-Surui	Tunisian	NorthAfrica-c	0.47	0.72	0.50	0.56	0.98	0.44
Papuan-Yoruba	Tunisian	NorthAfrica-c	0.84	0.95	0.91	1.04	1.24	0.85
Average:	————-	————-	0.74	0.86	0.79	0.71	0.96	0.74
Papuan-Australian	Moroccan	NorthAfrica-d	-1.21	-0.74	-1.05	-0.74	-2.09	-0.64
Papuan-Han	Moroccan	NorthAfrica-d	0.35	0.34	0.32	0.32	0.79	0.27

Table S4 continued from previous page

Papuan-Ju-hoan-North	Moroccan	NorthAfrica-d	0.92	1.01	0.98	1.14	1.29	0.95
Mbuti-Ju-hoan-North	Moroccan	NorthAfrica-d	-0.88	-1.32	-1.07	-1.32	-0.94	-0.83
Papuan-Karitiana	Moroccan	NorthAfrica-d	0.69	0.99	0.69	0.82	0.97	0.65
Papuan-Mbuti	Moroccan	NorthAfrica-d	0.86	0.94	0.91	1.07	1.18	0.88
Papuan-Saami	Moroccan	NorthAfrica-d	0.69	0.76	0.74	0.82	0.99	0.81
Papuan-Somali	Moroccan	NorthAfrica-d	0.87	1.12	0.94	1.08	1.17	0.80
Papuan-Surui	Moroccan	NorthAfrica-d	0.55	0.77	0.58	0.64	0.95	0.53
Papuan-Yoruba	Moroccan	NorthAfrica-d	0.87	0.96	0.92	1.04	1.17	0.87
Average:	————-	————-	0.73	0.79	0.76	0.65	0.92	0.72
Papuan-Australian	Mozabite	NorthAfrica-d	6.11	-18.33	9.17	-11.00	4.23	-5.50
Papuan-Han	Mozabite	NorthAfrica-d	0.10	0.10	0.09	0.09	0.31	0.07
Papuan-Ju-hoan-North	Mozabite	NorthAfrica-d	0.92	1.01	0.98	1.17	1.34	0.95
Mbuti-Ju-hoan-North	Mozabite	NorthAfrica-d	-0.47	-0.62	-0.53	-0.59	-0.51	-0.44
Papuan-Karitiana	Mozabite	NorthAfrica-d	0.66	1.02	0.67	0.80	0.95	0.60
Papuan-Mbuti	Mozabite	NorthAfrica-d	0.85	0.94	0.91	1.08	1.21	0.87
Papuan-Saami	Mozabite	NorthAfrica-d	0.64	0.71	0.69	0.79	0.98	0.76
Papuan-Somali	Mozabite	NorthAfrica-d	0.87	1.13	0.94	1.09	1.19	0.82
Papuan-Surui	Mozabite	NorthAfrica-d	0.53	0.76	0.55	0.62	0.98	0.50
Papuan-Yoruba	Mozabite	NorthAfrica-d	0.86	0.96	0.92	1.04	1.20	0.87
Average:	————-	————-	0.68	0.69	0.72	0.58	0.81	0.68
Papuan-Australian	Saharawi	NorthAfrica-d	4.38	NaN	5.70	-57.00	2.71	-19.00
Papuan-Han	Saharawi	NorthAfrica-d	0.28	0.26	0.25	0.24	0.67	0.21
Papuan-Ju-hoan-North	Saharawi	NorthAfrica-d	0.92	1.01	0.98	1.15	1.31	0.96
Mbuti-Ju-hoan-North	Saharawi	NorthAfrica-d	-0.63	-0.89	-0.73	-0.89	-0.65	-0.65
Papuan-Karitiana	Saharawi	NorthAfrica-d	0.69	1.03	0.70	0.83	0.98	0.64
Papuan-Mbuti	Saharawi	NorthAfrica-d	0.86	0.95	0.91	1.07	1.19	0.89
Papuan-Saami	Saharawi	NorthAfrica-d	0.67	0.74	0.72	0.81	0.99	0.79
Papuan-Somali	Saharawi	NorthAfrica-d	0.88	1.11	0.95	1.08	1.17	0.83
Papuan-Surui	Saharawi	NorthAfrica-d	0.54	0.77	0.57	0.63	0.99	0.51

Table S4 continued from previous page

Papuan-Yoruba	Saharawi	NorthAfrica-d	0.87	0.96	0.93	1.04	1.18	0.88
Average:	————--	————--	0.71	NaN	0.75	0.63	0.91	0.71

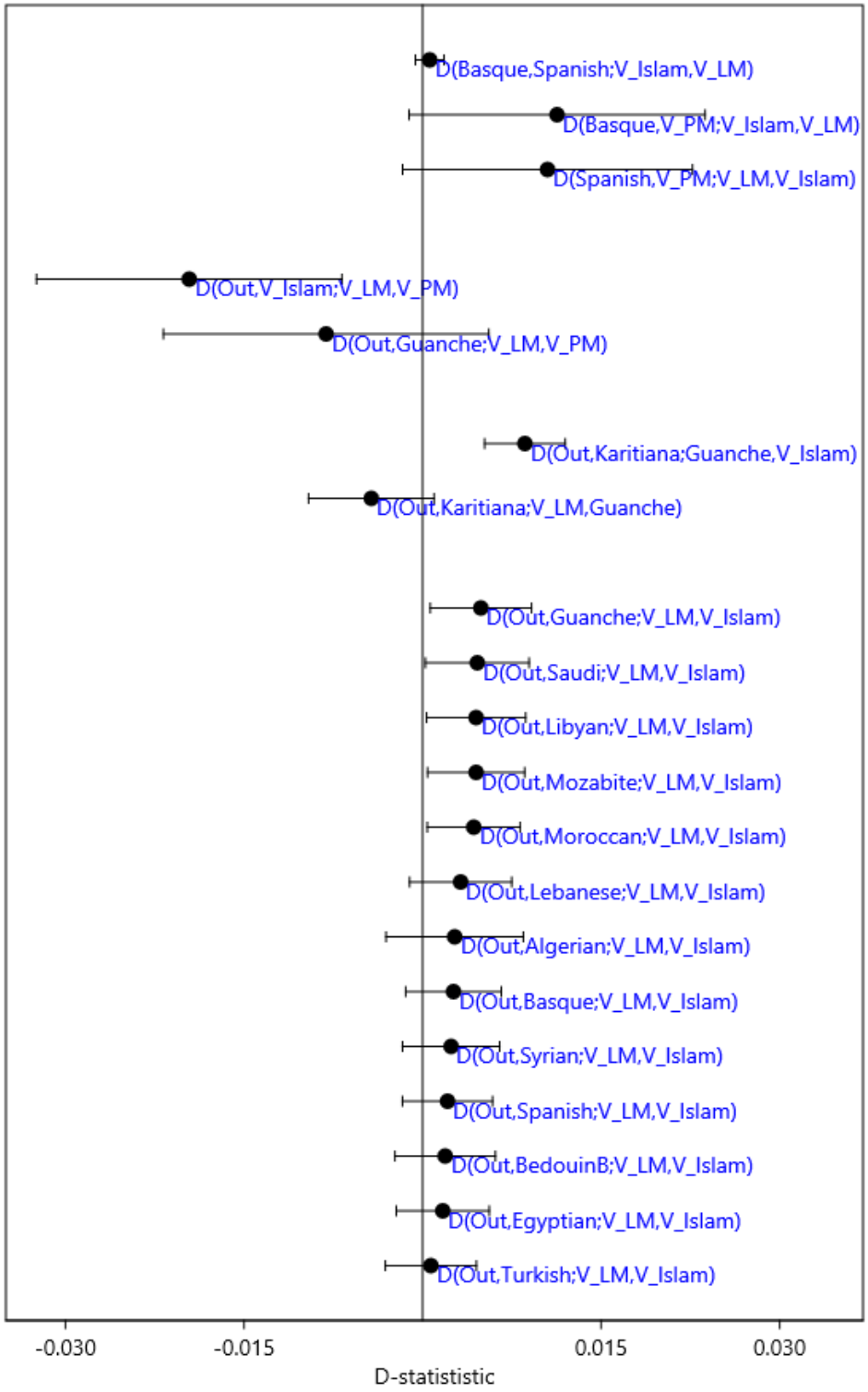


Figure S27: D-statistic for cross-comparison of Valencian Islamic (V_Islam), Late Medieval (V_LM) and post-Medieval (V_PM) groups and with some other modern and ancient populations.

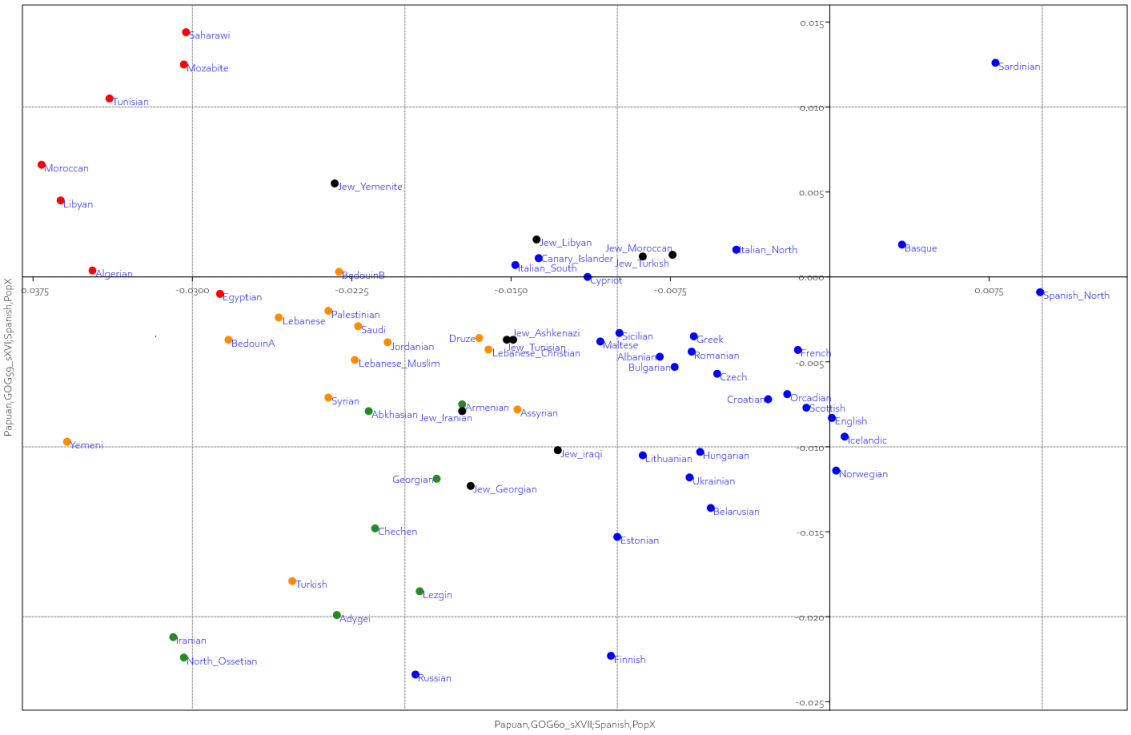


Figure S28: Crossed D-statistics for the two GOG Post-Medieval samples, dated to the 16th and 17th century. This should not be confused with an f4-ratio test because neither the slope indicates a valid alpha values, nor the numbers used correpond with f4 values.

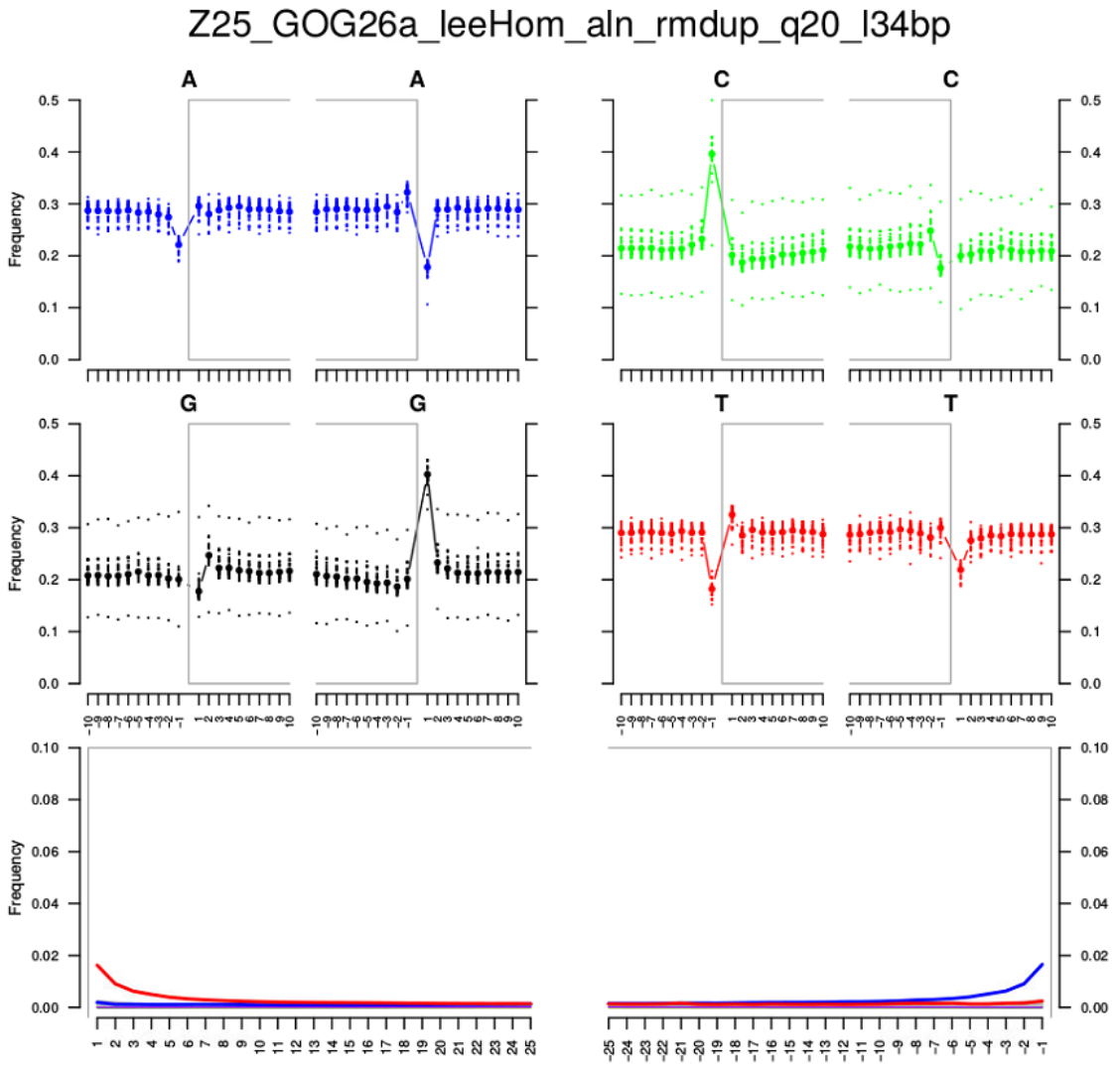


Figure S29: Example of the post-mortem damage detected at the end of the UDG-treated reads in the ancient DNA of one of the libraries sequenced for sample GOG26; after duplicate removal, and applying quality and length filters. All libraries of all samples displayed a similar pattern. Based on this observation I soft-clipped three bases at both ends to reduce numbers of false mutations introduced in further analyses.

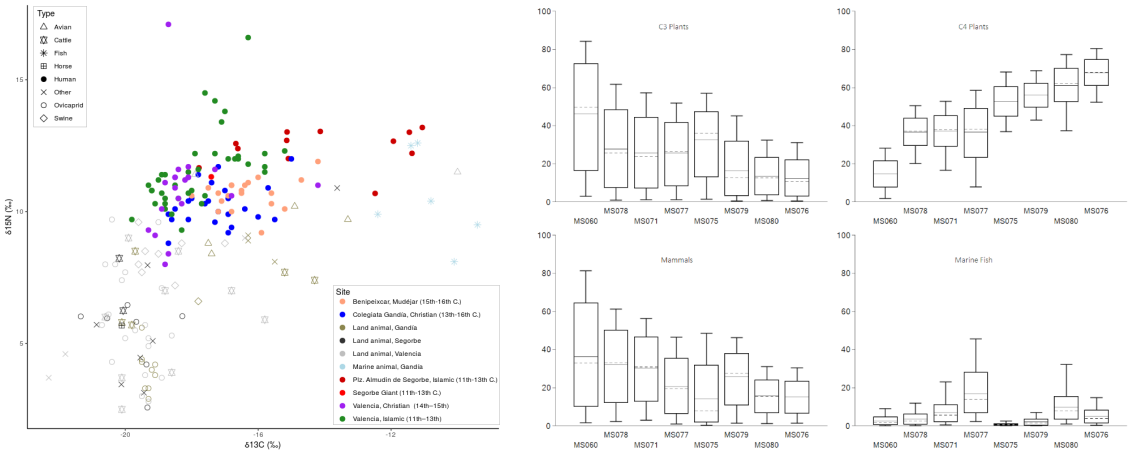


Figure S30: Dietary isotopes from the studied individuals from the Islamic necropolis of Segorbe. Only MS060 (Segorbe Giant) yielded DNA data as well as dietary information.

Figure S31: One of the four inhumations from Costamar, this individual was buried with a great amount of elaborated grave goods. The reconstructed pot belongs to one of the other inhumations at Costamar, the style is likely Cardium Pottery which indicates an Early Neolithic affiliation. Source: Enric Flors, head archaeologist.

Figure S32: Photos of samples taken from Visigothic Necropolis of La Union (Vall d'Uixo) (A), Islamic samples from Vall d'Uixo (B and C), and the Visigothic/Byzantine double inhumation of GOG34 and GOG35 from Gandia (D).

Figure S33: Remains of the golden threads found in some of the tombs of La Rauda from la Plaza de l'Almoina in Valencia. This was probably sewed to the funerary fabric that covered the bodies. The fabric has disappeared but the pattern of the golden thread can still be appreciated. Source: Gonzalo Oteo Garcia (Left). Individuals recovered during the excavation works in l'Almoina in 2002. The threeskeletons likely belonged to women of the royal family that ruled the city around the 11th CenturyAD. They were found buried with golden threads. Source: SIAM archive (Right).

Figure S34: Burial of sample GOG60 from the 16th century, evidence of slavery in Valencian archaeology.

6.3 Supplement Chapter III

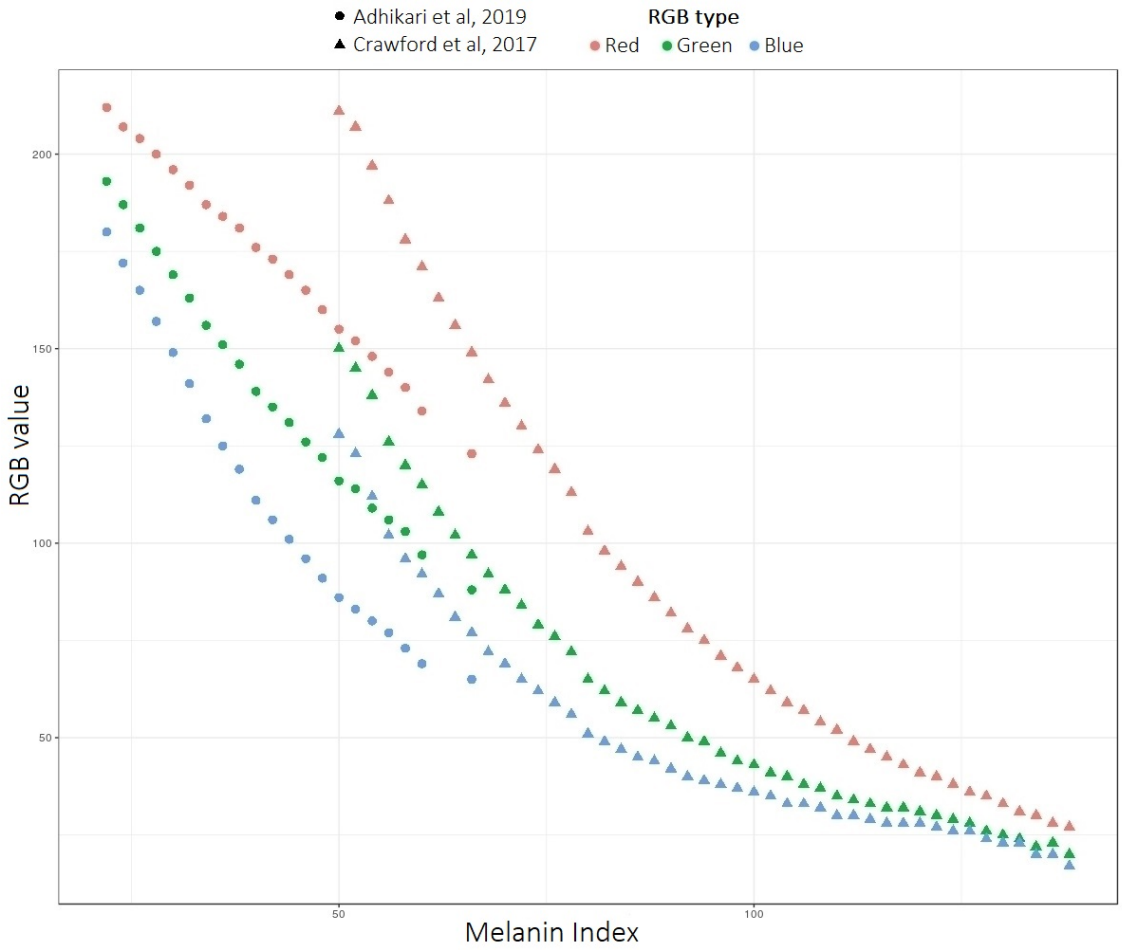


Figure S35: Correlation of RGB and MI from two different studies used (Crawford et al., 2017; Adhikari et al., 2019).

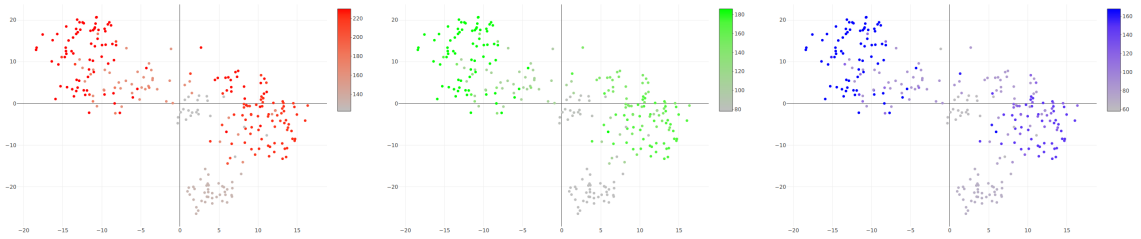


Figure S36: 2D tSNE and RGB values of each layer of colour.

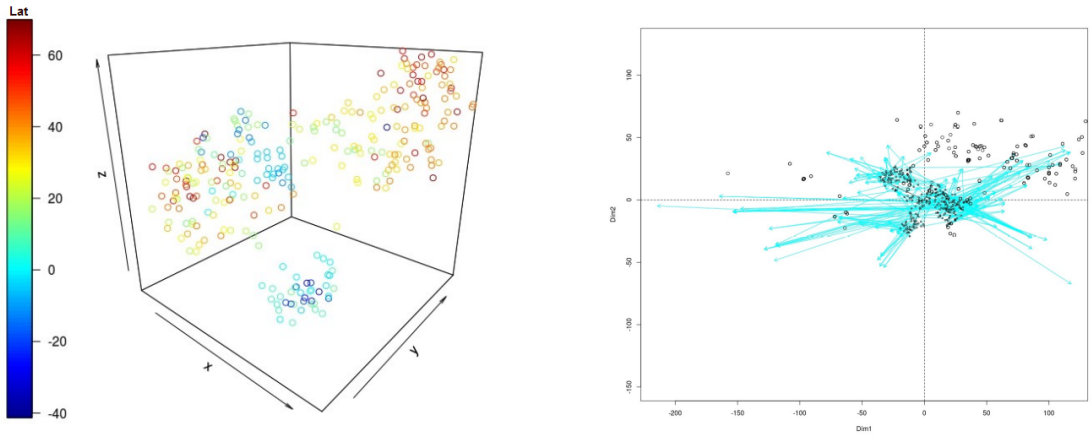


Figure S37: Left) 3D tSNE made using 36 SNPs genotypes linked to skin pigmentation. Dots are all 279 public samples of the Simons Genome Diversity Project (SGDP), coloured by their latitude. Right) Procrustes transformation of a 3D tSNE made using 36 SNPs genotypes linked to skin pigmentation and forced to target the 2D geographical locations of provenence of the SGDP samples. Small open dots connected to blue arrow are the tsne coordinates. Big open dots are the target geographical coordinates.

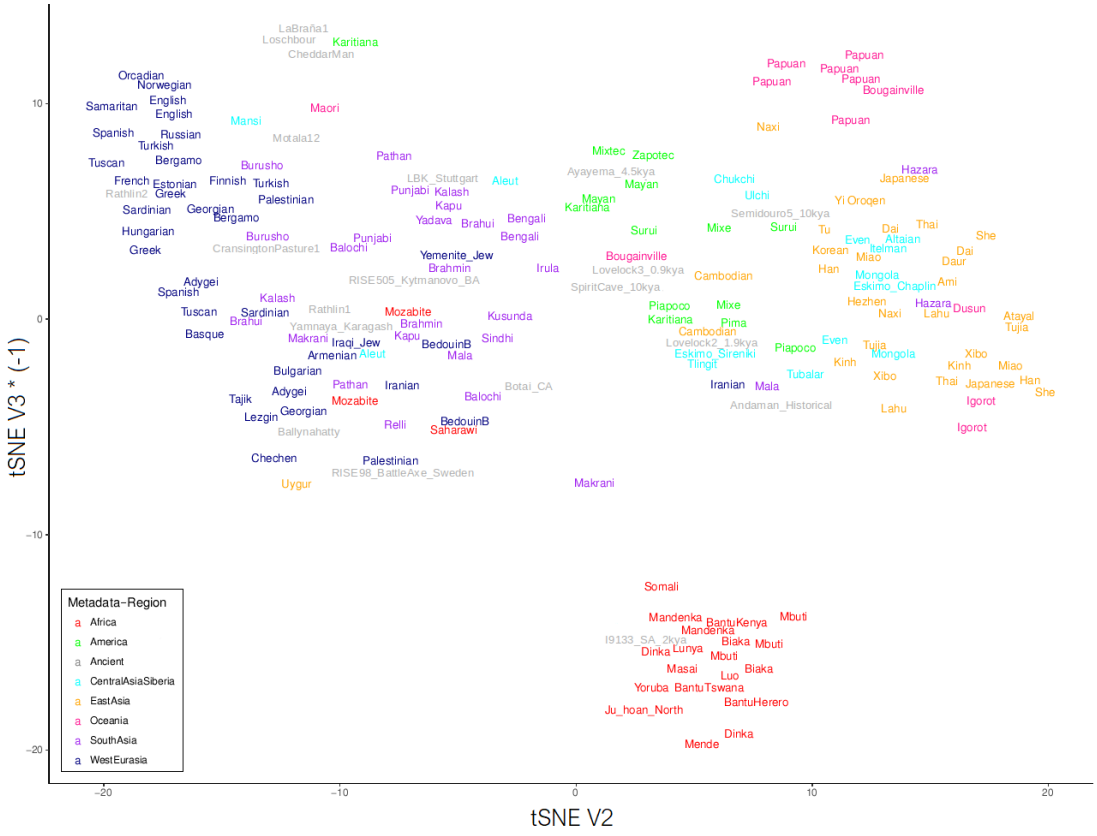


Figure S38: Detail of tSNE made with 36 SNPs and SGDP samples together with some high coverage ancient samples. Overlapping labels were removed for clearer view.

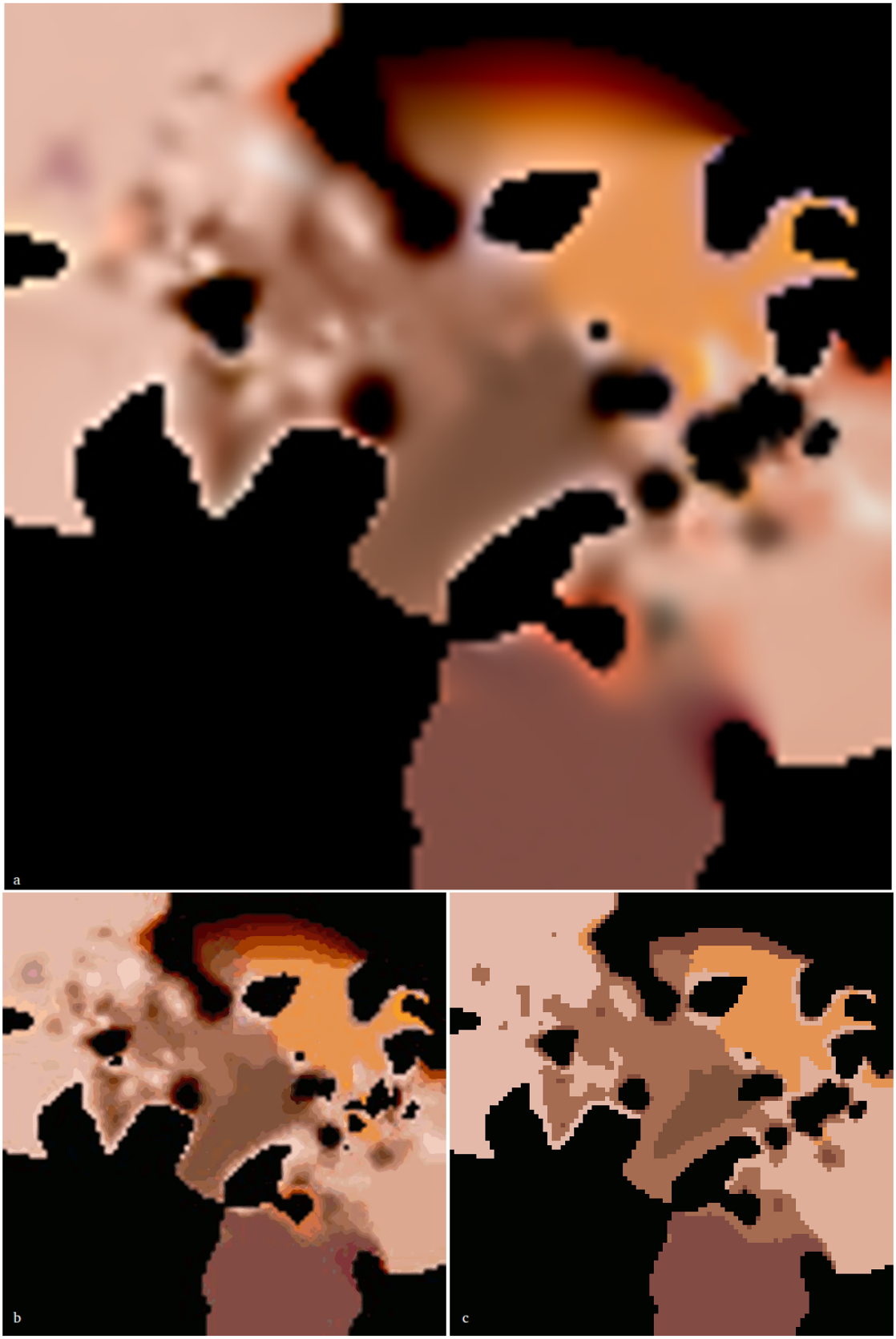


Figure S39: Colour map of human skin pigmentation reconstructed using Shepard 2D interpolation algorithm. Coordinates of the points used for interpolation were the dimensional reduction genomic coordinates of each of the 279 SGDP individuals obtained from the tSNE run using the 36 skin pigmentation related SNPs. The three interpolation Colour Maps were built combining each of the R, G and B values associated to the SGDP individuals. Map with 256 colours (a), map with 128 colours (b), and map with 6 colours (c).

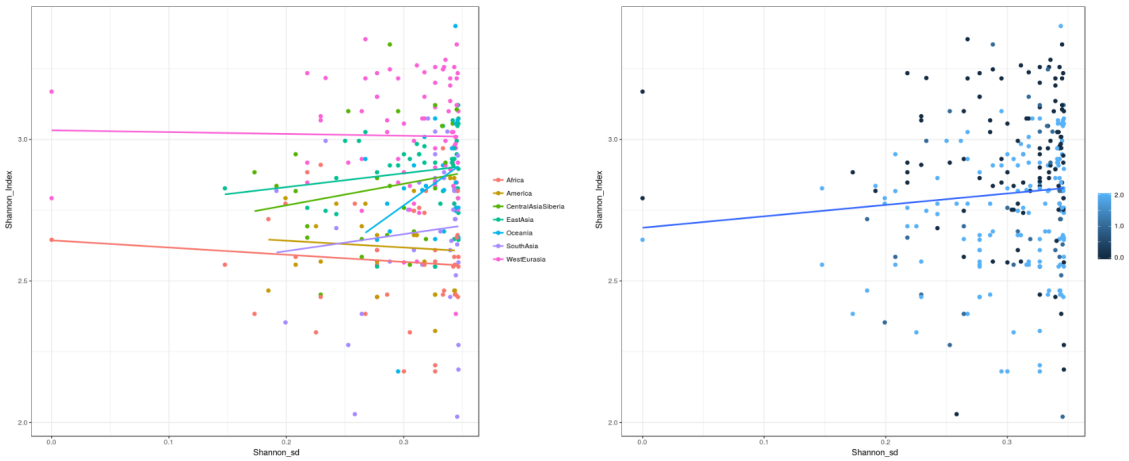


Figure S40: Left) Shannon indeces for skin SGDP genotypes by region. Right) Shannon indeces for skin SGDP genotypes and coloured by the state of their two alleles for SNP rs1426654 (in gene SLC24A5, chromosome 15). Most evident discrete differece when forced to simplify between *Darker vs Lighter* groups, but still no clear boundary.

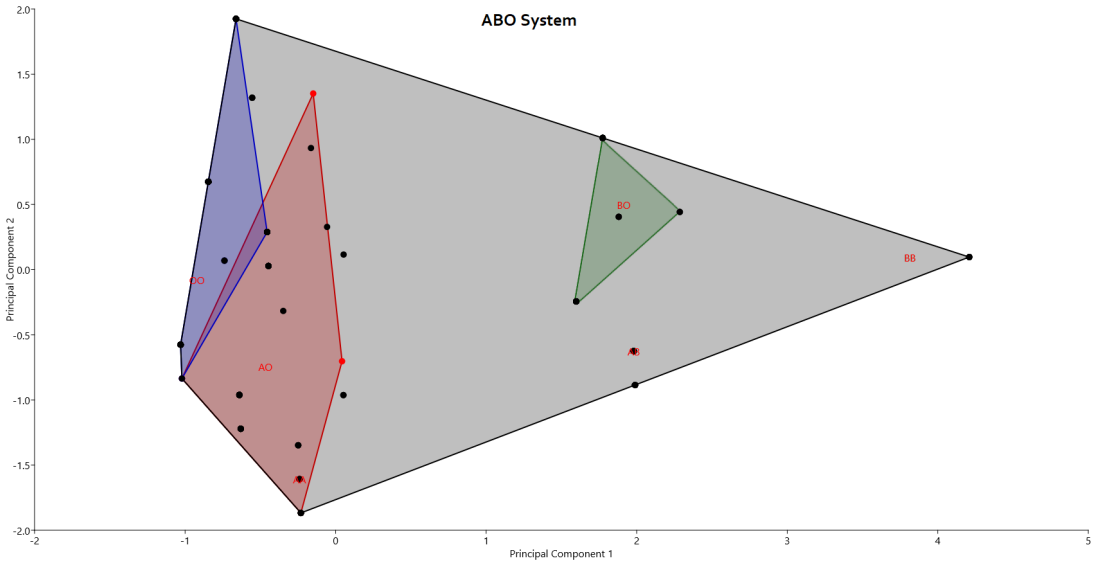


Figure S41: Allele frequency space resulting from dimension reduction using 12 SNPs related to the AB0 system. The outcome classified 300 samples into about 30 genotyped that correspond to the four blood types (A, B, 0 and AB).

rsID	Chr:Position	Reference	Derived
rs16891982	5:33951693	C	G
rs28777	5:33958959	C	A
rs12203592	6:396321	C	T
rs683	9:12709305	C	A
rs10756819	9:16858084	G	A
rs1042602	11:88911696	C	A
rs1393350	11:89011046	G	A
rs1126809	11:89017961	G	A
rs12821256	12:89328335	T	C
rs12896399	14:92773663	G	T
rs2402130	14:92801203	G	A
rs17128291	14:92882826	A	G
rs1545397	15:28187772	A	T
rs1800414	15:28197037	T	C
rs1800407	15:28230318	C	T
rs12441727	15:28271775	G	A
rs1470608	15:28288121	G	T
rs1129038	15:28356859	C	T
rs12913832	15:28365618	A	G
rs2238289	15:28453215	A	G
rs6497292	15:28496195	A	G
rs1667394	15:28530182	C	T
rs1426654	15:48426484	A	G
rs3114908	16:89383725	T	C
rs3212355	16:89984378	C	T
rs1805006	16:89985918	C	A
rs2228479	16:89985940	G	A
rs11547464	16:89986091	G	A
rs1805007	16:89986117	C	T
rs1110400	16:89986130	T	C
rs1805008	16:89986144	C	T
rs885479	16:89986154	G	A
rs8051733	16:90024206	A	G
rs6059655	20:32665748	A	G
rs6119471	20:32785212	C	G
rs2378249	20:33218090	G	A

Table S5: List of SNPs used to genotype the Simons Genome Diversity Project samples, from the publicly available VCFs. Reference allele according to the hg19 human reference genome version.

Sample ID	Context	Missing SNPs
Cheddar Man	British Mesolithic (9kya)	3 (8%)
La Brana 1 (Shotgun)	Iberian Mesolithic (7kya)	6 (17%)
La Brana 1 (I0585)	Iberian Mesolithic (7kya)	3 (8%)
Motala12	Scandinavian Mesolithic	9 (25%)
Loschbour	Luxembourg Mesolithic (8kya)	0
Stuttgart LBK	German Neolithic (7kya)	0
RISE98-BattleAxe	Swedish Neolithic	1 (3%)
Carsington Pasture 1	British Neolithic	0
Ballynahatty-BA64	Irish Neolithic	0
Rathlin1-M127	Irish Bronze Age	0
Rathlin2-RSK1	Irish Bronze Age	3 (8%)
Ava (I5385)	British Bronze Age	13 (36%)
Yamna-Karagash	Steppe Bronze Age (5.5kya)	0
BOTAI2016	Steppe Chalcolithic	0
RISE505-Andronovo	Steppe Bronze Age (3kya)	0
6DriffeldTerrace3	Roman Britain	7 (19%)
NO3423-Norton	Anglo-Saxon Britain	13 (36%)
I9133-Faraoskop	South Africa Prehistory (2kya)	3 (8%)
MA756-Andaman	Andaman Islands (19th century)	0
AHUR2064-Spirit Cave	Native North American (10kya)	0
Lovelock2	Native North American (2kya)	0
Lovelock3	Native North American (1kya)	0
Sumidouro5	Native South American (10kya)	0
A460-Ayayema	Native South American (4.5kya)	0
Ust'-Ishim Man	Siberian Paleolithic (45kya)	0
SSG-A2 (Pre-Christian)	Medieval Iceland (1kya)	0

Table S6: List of publicly available medium to high-coverage ancient genomes used in this study. Samples marked with * did not participate in the tSNE from Figure 4 but were included in the tSNE coordinates used in the interpolation of Figure 5. This accounts for the differeces in the relative position of the American and Papuan clusters in a 2D display.

References

Abascal Palazón, J. M. and Cebrián Fernández, R. (2015), 'Inscripciones romanas de Paterna, Valencia y Riba-roja de Túria (Territorium de Valentia, Hispania Citerior)', *SAGVNTVM: Papeles del Laboratorio de Arqueología de Valencia* **46**(18).

Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., Moral, P., Dugoujon, J. M., Roostalu, U., Loogväli, E. L., Kivisild, T., Bandelt, H. J., Richards, M., Villems, R., Silvana Santachiara-Benerecetti, A., Semino, O. and Torroni, A. (2004), 'The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool', *American Journal of Human Genetics* **75**(5), 910–918.

Adams, J. (2007), *The Regional Diversification of Latin 200 BC - AD 600*, Cambridge University Press.

Adhikari, K. and et al (2016*a*), 'A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation', *Nature Communications* **7**, 11616.

Adhikari, K. and et al (2016*b*), 'A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features', *Nature Communications* **7**, 10815.

Adhikari, K., Mendoza-Revilla, J., Sohail, A., Fuentes-Guajardo, M., Lampert, J., Chacón-Duque, J. C., Hurtado, M., Villegas, V., Granja, V., Acuña-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R. B., Everardo, P., Gómez-Valdés, J., Villamil-Ramírez, H., Silva de Cerqueira, C. C., Hunemeier, T., Ramallo, V., Schuler-Faccini, L., Salzano, F. M., Gonzalez-José, R., Bortolini, M. C., Canizales-Quinteros, S., Gallo, C., Poletti, G., Bedoya, G., Rothhammer, F., Tobin, D. J., Fumagalli, M., Balding, D. and Ruiz-Linares, A. (2019), 'A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia', *Nature Communications* **10**(1), 358.

Adler, C. J., Haak, W., Donlon, D. and Cooper, A. (2011), 'Survival and recovery of DNA from ancient teeth and bones', *Journal of Archaeological Science* **38**(5), 956–964.

Agranat-Tamir, L., Waldman, S., Martin, M. A., Gokhman, D., Mishol, N., Eshel, T., Cheronet, O., Rohland, N., Mallick, S., Adamski, N., Lawson, A. M., Mah, M., Michel, M., Oppenheimer, J., Stewardson, K. and Reich, D. (2020), 'The Genomic History of the Bronze Age Southern Levant', *Cell* **181**, 1146–1157.

Aguilella Arzo, G., Olaria Puyoles, C. and Gusi Jener, F. (1999), 'El jaciment prehistòric de La Cova dels Diablets (Alcalà de Xivert, Castelló)', *Quaderns de prehistòria i arqueologia de Castelló* **20**, 7–36.

Akey, J. M., Eberle, M. A., Rieder, M. J., Carlson, C. S., Shriver, M. D., Nickerson, D. A. and Kruglyak, L. (2004), 'Population history and natural selection shape patterns of genetic variation in 132 genes', *PLoS Biology* **2**(10), e286.

Alapont Martín, L., Calvo Gálvez, M. and Ribera i Lacomba, A. (2009), *La destruccion de Valentia por Pompeyo (75 a.C.)*, Quaderns de difusioó arqueològica (Ajuntament de València).

Albelda Borrás, V. (2015), 'Ruaya: los iberos junto a laciudad de València (el sucronensis sinus en época ibérica))', *SAGVNTVM: Papeles del Laboratorio de Arqueología de Valencia* **17**, 101–106.

Alday Ruiz, A. (2009), 'El final del Mesolítico y los inicios del Neolítico en la Península Ibérica: cronología y fases', *MUNIBE Antropologia-Arkeologia,* **60**, 157–173.

Alexander, D. H., Novembre, J. and Lange, K. (2009), 'Fast model-based estimation of ancestry in unrelated individuals', *Genome Research* **19**(9), 1655–1664.

Allentoft, M. E., Sikora, M., Sjögren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A. S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. D., Dąbrowski, P., Duffy, P. R., Ebel, A. V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolář, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtchyan, R., Moiseyev, V., Paja, L., Pálfi, G., Pokutta, D., Pospieszny, Ł., Douglas Price, T., Saag, L., Sablin, M., Shishlina, N., Smrčka, V., Soenov, V. I., Szeverényi, V., Tóth, G., Trifanova, S. V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sicheritz-Pontén, T., Brunak, S., Nielsen, R., Kristiansen, K. and Willerslev, E. (2015), 'Population genomics of Bronze Age Eurasia', *Nature* **522**(7555), 167–172.

Almagro Gorbea, M. (2004), 'Inscripciones y grafitos tartésicos de la necrópolis orientalizante de Medellín', *Palaeohispánica: Revista sobre lenguas y culturas de la Hispania antigua* **4**, 13–44.

Alonso-Fernandez, C and Jimenez-Echevarria, J. (2015), 'La progresión del 'fenómeno' Campaniforme durante el Bronce Medio en el entorno del Sistema Ibérico: el poblado de Valdescusa (Hervías, La Rioja, España)', *MUNIBE Antropologia-Arkeologia* **66**, 147–162.

Amorim, A., Fernandes, T. and Taveira, N. (2019), 'Mitochondrial DNA in human identification: A review', *PeerJ* **7**, e7314.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. and Howell, N. (1999), 'Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA', *Nature Genetics* **23**(2), 147.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C. and Wingett, S. (2012), 'FastQC', Babraham Institute.

Anthony, D. W. (2010), *The horse, the wheel, and language: How Bronze-Age riders from the Eurasian steppes shaped the modern world*, Princeton University Press.

Anthony, D. W. and Ringe, D. (2015), 'The Indo-European Homeland from Linguistic and Archaeological Perspectives', *Annual Review of Linguistics* **1**(1), 199–219.

Antonio, M. L., Gao, Z., Moots, H. M., Lucci, M., Candilio, F., Sawyer, S., Oberreiter, V., Calderon, D., Devitofranceschi, K., Aikens, R. C., Aneli, S., Bartoli, F., Bedini, A., Cheronet, O., Cotter, D. J., Fernandes, D. M., Gasperetti, G., Grifoni, R., Guidi, A., Pastina, F. L., Loreti, E., Manacorda, D., Matullo, G., Morretta, S., Nava, A., Nicolai, V. F., Nomi, F., Pavolini, C., Pentiricci, M., Pergola, P., Piranomonte, M., Schmidt, R., Spinola, G., Sperduti, A., Rubini, M., Bondioli, L., Coppa, A., Pinhasi, R. and Pritchard, J. K. (2019), 'Ancient Rome: A genetic crossroads of Europe and the Mediterranean', *Science* **366**(6466), 708–714.

Arauna, L. R., Mendoza-Revilla, J., Mas-Sandoval, A., Izaabel, H., Bekada, A., Benhamamouch, S., Fadhlaoui-Zid, K., Zalloua, P., Hellenthal, G. and Comas, D. (2017), 'Recent Historical Migrations Have Shaped the Gene Pool of Arabs and Berbers in North Africa', *Molecular biology and evolution* **34**(2), 318–329.

Artola Gallego, M. (1993), *Enciclopedia de Historia de España (VI) Cronología. Mapas. Estadísticas*, Alianza Editorial.

Aubet, M. E. (2001), *The Phoenicians and the West: Politics, Colonies and Trade*, Cambridge University Press.

Augustyn, A., Zeidan, A., Zelazko, A., Eldridge, A., McKenna, A., Tikkanen, A. and Schreiber, B. A., eds (2020), *Goths*, Encyclopædia Britannica, inc.

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., Mardis, E. R., Marth, G. T., McVean, G. A., Nickerson, D. A., Schmidt, J. P., Sherry, S. T., Wang, J., Wilson, R. K., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J. G., Zhu, Y., Chang, Y., Feng, Q., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Lan, T., Li, G., Li, J., Li, Y., Liu, S., Liu, X., Lu, Y., Ma, X., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Xu, X., Yin, Y., Zhang, D., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Gupta, N., Gharani, N., Toji, L. H., Gerry, N. P., Resch, A. M., Barker, J., Clarke, L., Gil, L., Hunt, S. E., Kelman, G., Kulesha, E., Leinonen, R., McLaren, W. M., Radhakrishnan, R., Roa, A., Smirnov, D., Smith, R. E., Streeter, I., Thormann, A., Toneva, I., Vaughan, B., Zheng-Bradley, X., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Sudbrak, R., Albrecht, M. W., Amstislavskiy, V. S., Borodina, T. A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M. L., Fulton, L., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Church, D., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., O'Sullivan, C., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T. M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Davies, C. J., Gollub, J., Webster, T., Wong, B., Zhan, Y., Campbell, C. L., Kong, Y., Marcketta, A., Yu, F., Antunes, L., Bainbridge, M., Sabo, A., Huang, Z., Coin, L. J., Fang, L., Li, Q., Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Xie, Y., Ye, C., Yu, C., Zhang, F., Zheng, H., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Garrison, E. P., Kural, D., Lee, W. P., Leong, W. F., Stromberg, M., Ward, A. N., Wu, J., Zhang, M., Daly, M. J., DePristo, M. A., Handsaker, R. E., Banks, E., Bhatia, G., Del Angel, G., Genovese, G., Li, H., Kashin, S., McCarroll, S. A., Nemesh, J. C., Poplin, R. E., Yoon, S. C., Lihm, J., Makarov, V., Gottipati, S., Keinan, A., Rodriguez-Flores, J. L., Rausch, T., Fritz, M. H., Stütz, A. M., Beal, K., Datta, A., Herrero, J., Ritchie, G. R., Zerbino, D., Sabeti, P. C., Shlyakhter, I., Schaffner, S. F., Vitti, J., Cooper, D. N., Ball, E. V., Stenson, P. D., Barnes, B., Bauer, M., Cheetham, R. K., Cox, A., Eberle, M., Kahn, S., Murray, L., Peden, J., Shaw, R., Kenny, E. E., Batzer, M. A., Konkel, M. K., Walker, J. A., MacArthur, D. G., Lek, M., Herwig, R., Ding, L., Koboldt, D. C., Larson, D., Ye, K., Gravel, S., Swaroop, A., Chew, E., Lappalainen, T., Erlich, Y., Gymrek, M., Willems, T. F., Simpson, J. T., Shriver, M. D., Rosenfeld, J. A., Bustamante, C. D., Montgomery, S. B., De La Vega, F. M., Byrnes, J. K., Carroll, A. W., DeGorter, M. K., Lacroute, P., Maples, B. K., Martin, A. R., Moreno-Estrada, A., Shringarpure, S. S., Zakharia, F., Halperin, E., Baran, Y., Cerveira, E., Hwang, J., Malhotra, A., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Hyland, F. C., Craig, D. W., Christoforides, A., Homer, N., Izatt, T., Kurdoglu, A. A., Sinari, S. A., Squire, K., Xiao, C., Sebat, J., Antaki, D., Gujral, M., Noor, A., Ye, K., Burchard, E. G., Hernandez, R. D., Gignoux, C. R., Haussler, D., Katzman, S. J., Kent, W. J., Howie, B., Ruiz-Linares, A., Dermitzakis, E. T., Devine, S. E., Kang, H. M., Kidd, J. M., Blackwell, T., Caron, S., Chen, W., Emery, S., Fritsche, L., Fuchsberger, C., Jun, G., Li, B., Lyons, R., Scheller, C., Sidore, C., Song, S., Sliwerska, E., Taliun, D., Tan, A., Welch, R., Wing, M. K., Zhan, X., Awadalla, P., Hodgkinson, A., Li, Y., Shi, X., Quitadamo, A., Lunter, G., Marchini, J. L., Myers, S., Churchhouse, C., Delaneau, O., Gupta-Hinch, A., Kretzschmar, W., Iqbal, Z., Mathieson, I., Menelaou, A., Rimmer, A., Xifara, D. K., Oleksyk, T. K., Fu, Y., Liu, X., Xiong, M., Jorde, L., Witherspoon, D., Xing, J., Browning, B. L., Browning, S. R., Hormozdiari, F., Sudmant, P. H., Khurana, E., Tyler-Smith, C., Albers, C. A., Ayub, Q., Chen, Y., Colonna, V., Jostins, L., Walter, K., Xue, Y., Gerstein, M. B., Abyzov, A., Balasubramanian, S., Chen, J., Clarke, D., Fu, Y., Harmanci, A. O., Jin, M., Lee, D., Liu, J., Mu, X. J., Zhang, J., Zhang, Y., Hartl, C., Shakir, K., Degenhardt, J., Meiers, S., Raeder, B., Casale, F. P., Stegle, O., Lameijer, E. W., Hall, I., Bafna, V., Michaelson, J., Gardner, E. J., Mills, R. E., Dayama, G., Chen, K., Fan, X.,

Chong, Z., Chen, T., Chaisson, M. J., Huddleston, J., Malig, M., Nelson, B. J., Parrish, N. F., Blackburne, B., Lindsay, S. J., Ning, Z., Zhang, Y., Lam, H., Sisu, C., Challis, D., Evani, U. S., Lu, J., Nagaswamy, U., Yu, J., Li, W., Habegger, L., Yu, H., Cunningham, F., Dunham, I., Lage, K., Jespersen, J. B., Horn, H., Kim, D., Desalle, R., Narechania, A., Sayres, M. A., Mendez, F. L., Poznik, G. D., Underhill, P. A., Mittelman, D., Banerjee, R., Cerezo, M., Fitzgerald, T. W., Louzada, S., Massaia, A., Yang, F., Kalra, D., Hale, W., Dan, X., Barnes, K. C., Beiswanger, C., Cai, H., Cao, H., Henn, B., Jones, D., Kaye, J. S., Kent, A., Kerasidou, A., Mathias, R., Ossorio, P. N., Parker, M., Rotimi, C. N., Royal, C. D., Sandoval, K., Su, Y., Tian, Z., Tishkoff, S., Via, M., Wang, Y., Yang, H., Yang, L., Zhu, J., Bodmer, W., Bedoya, G., Cai, Z., Gao, Y., Chu, J., Peltonen, L., Garcia-Montero, A., Orfao, A., Dutil, J., Martinez-Cruzado, J. C., Mathias, R. A., Hennis, A., Watson, H., McKenzie, C., Qadri, F., LaRocque, R., Deng, X., Asogun, D., Folarin, O., Happi, C., Omoniwa, O., Stremlau, M., Tariyal, R., Jallow, M., Joof, F. S., Corrah, T., Rockett, K., Kwiatkowski, D., Kooner, J., Hien, T. T., Dunstan, S. J., ThuyHang, N., Fonnie, R., Garry, R., Kanneh, L., Moses, L., Schieffelin, J., Grant, D. S., Gallo, C., Poletti, G., Saleheen, D., Rasheed, A., Brooks, L. D., Felsenfeld, A. L., McEwen, J. E., Vaydylevich, Y., Duncanson, A., Dunn, M. and Schloss, J. A. (2015), 'A global reference for human genetic variation', *Nature* **526**(7571), 68–74.

Bachtrog, D. and Charlesworth, B. (2001), 'Towards a complete sequence of the human Y chromosome'.

Balloux, F., Handley, L.-J. L., Jombart, T., Liu, H. and Manica, A. (2009), 'Climate shaped the worldwide distribution of human mitochondrial dna sequence variation', *Proceedings of the Royal Society B: Biological Sciences* **276**(1672), 3447–3455.

Bandelt, H. J., Kloss-Brandstätter, A., Richards, M. B., Yao, Y. G. and Logan, I. (2013), 'The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies', **59**(2), 66–77.

Barnosky, A. D., Koch, P. L., Feranec, R. S., Wing, S. L. and Shabel, A. B. (2004), 'Assessing the causes of late pleistocene extinctions on the continents', **306**(5693), 70–75.

Barrachina Ibáñez, E. (2005), 'La necròpolis islàmica de la plaça de l'Almudín, Sogorb (Alt Palància). Estudi antropològic i cronològic.', *Quaderns de prehistòria i arqueologia de Castelló* **24**, 281–294.

Barral-Arca, R., Pischedda, S., Gómez-Carballa, A., Pastoriza, A., Mosquera-Miguel, A., López-Soto, M., Martinón-Torres, F., Álvarez-Iglesias, V. and Salas, A. (2016), 'Meta-analysis of mitochondrial DNA variation in the Iberian Peninsula', *PLoS ONE* **11**(7), e0159735.

Barsh, G. S. (2003), 'What controls variation in human skin color?', **1**(1), e27.

Batini, C., Hallast, P., Vågene, Å. J., Zadik, D., Eriksen, H. A., Pamjav, H., Sajantila, A., Wetton, J. H. and Jobling, M. A. (2017), 'Population resequencing of European mitochondrial genomes highlights sex-bias in Bronze Age demographic expansions', *Scientific Reports* **7**, 12086.

Batini, C., Hallast, P., Zadik, D., Delser, P. M., Benazzo, A., Ghirotto, S., Arroyo-Pardo, E., Cavalleri, G. L., De Knijff, P., Dupuy, B. M., Eriksen, H. A., King, T. E., De Munain, A. L., López-Parra, A. M., Loutradis, A., Milasin, J., Novelletto, A., Pamjav, H., Sajantila, A., Tolun, A., Winney, B. and Jobling, M. A. (2015), 'Large-scale recent expansion of European patrilineages shown by population resequencing', *Nature Communications* **6**, 7152.

Battaglia, V., Fornarino, S., Al-Zahery, N., Olivieri, A., Pala, M., Myres, N. M., King, R. J., Rootsi, S., Marjanovic, D., Primorac, D., Hadziselimovic, R., Vidovic, S., Drobnic, K., Durmishi,

N., Torroni, A., Santachiara-Benerecetti, S. A., Underhill, P. A. and Semino, O. (2009), 'Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe', *European Journal of Human Genetics* **17**(6), 820–830.

Behar, D. M. and et al (2012*a*), 'A "copernican" reassessment of the human mitochondrial DNA tree from its root', *American Journal of Human Genetics* **90**(4), 675–684.

Behar, D. M. and et al (2012*b*), 'The Basque paradigm: Genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-neolithic times', *American Journal of Human Genetics* **90**(3), 486–493.

Belenguer Cebrià, E. (2008), *Jaime I y su reinado*, Editorial MILENIO.

Beltran Lloris, F. and Willi, A. (2012), 'El regadío en la Hispania romana. Estado de las cuestión', *Cuadernos de prehistoria y arqueología de la Universidad de Granada* **21**(21), 9–56.

Ben-Tor, D. (2011), Egyptian-Canaanite Relations In The Middle And Late Bronze Ages As Reflected By Scarabs, *in* 'Egypt, Canaan and Israel: History, Imperialism, Ideology and Literature', BRILL, chapter Egyptian-Canaanite Relations In The Middle And Late Bronze Ages As Reflected By Scarabs, pp. 23–43.

Benazzi, S., Douka, K., Fornai, C., Bauer, C. C., Kullmer, O., Svoboda, J., Pap, I., Mallegni, F., Bayle, P., Coquerelle, M., Condemi, S., Ronchitelli, A., Harvati, K. and Weber, G. W. (2011), 'Early dispersal of modern humans in Europe and implications for Neanderthal behaviour', *Nature* **479**(7374), 525–528.

Benazzi, S., Slon, V., Talamo, S., Negrino, F., Peresani, M., Bailey, S. E., Sawyer, S., Panetta, D., Vicino, G., Starnini, E., Mannino, M. A., Salvadori, P. A., Meyer, M., Pääbo, S. and Hublin, J. J. (2015), 'The makers of the Protoaurignacian and implications for Neandertal extinction', *Science* **348**(6236), 793–796.

Bernabeu-Auban, J. and Barton, C.M. and Pardo-Gordo, S. and Bergin, S.M. (2015), 'Modeling initial Neolithic dispersal. The first agricultural groups in West Mediterranean', *Ecological Modelling* **307**, 22–31.

Bertranpetit, J. and Cavalli-Sforza, L. L. (1991), 'A genetic reconstruction of the history of the population of the Iberian Peninsula', *Annals of Human Genetics* **55**(1), 51–67.

Bodmer, W. (2015), 'Genetic characterization of human populations: From abo to a genetic map of the british people', *Genetics* **199**(2), 267–279.

Bogucki, P. and Crabtree, P. (2004), *Ancient Europe 8000 B.C. to A.D. 1000: Encyclopedia of the Barbarian World*, Charles Scribner's Sons.

Botigué, L. R., Henn, B. M., Gravel, S., Maples, B. K., Gignoux, C. R., Corona, E., Atzmon, G., Burns, E., Ostrer, H., Flores, C., Bertranpetit, J., Comasa, D. and Bustamante, C. D. (2013), 'Gene flow from North Africa contributes to differential human genetic diversity in southern europe', *Proceedings of the National Academy of Sciences of the United States of America* **110**(29), 11791–11796.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., Gray, R. D., Suchard, M. A. and Atkinson, Q. D. (2012), 'Mapping the origins and expansion of the Indo-European language family', *Science* **337**(6097), 957–960.

Brace, S., Diekmann, Y., Booth, T. J., van Dorp, L., Faltyskova, Z., Rohland, N., Mallick, S., Olalde, I., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Stewardson,

K., Martiniano, R., Walsh, S., Kayser, M., Charlton, S., Hellenthal, G., Armit, I., Schulting, R., Craig, O. E., Sheridan, A., Parker Pearson, M., Stringer, C., Reich, D., Thomas, M. G. and Barnes, I. (2019), 'Ancient genomes indicate population replacement in Early Neolithic Britain', *Nature Ecology and Evolution* **3**(5), 765–771.

Bramanti, B., Thomas, M. G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M. N., Jankauskas, R., Kind, C. J., Lueth, F., Terberger, T., Hiller, J., Matsumura, S., Forster, P. and Burger, J. (2009), 'Genetic discontinuity between local hunter-gatherers and central Europe's first farmers', *Science* **326**(5949), 137–140.

Brandini, S., Bergamaschi, P., Fernando Cerna, M., Gandini, F., Bastaroli, F., Bertolini, E., Cereda, C., Ferretti, L., Gómez-Carballa, A., Battaglia, V., Salas, A., Semino, O., Achilli, A., Olivieri, A. and Torroni, A. (2018), 'The Paleo-Indian entry into South America according to mitogenomes', *Molecular Biology and Evolution* **35**(2), 299–311.

Brandt, G., Haak, W., Adler, C. J., Roth, C., Szécsényi-Nagy, A., Karimnia, S., Möller-Rieker, S., Meller, H., Ganslmeier, R., Friederich, S., Dresely, V., Nicklisch, N., Pickrell, J. K., Sirocko, F., Reich, D., Cooper, A. and Alt, K. W. (2013), 'Ancient DNA reveals key stages in the formation of Central European mitochondrial genetic diversity', *Science* **342**(6155), 257–261.

Brandt, G., Szécsényi-Nagy, A., Roth, C., Alt, K. W. and Haak, W. (2015), 'Human paleogenetics of Europe - The known knowns and the known unknowns', *Journal of Human Evolution* **79**, 73–92.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. and Pääbo, S. (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), 14616–14621.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M. and Pääbo, S. (2009), 'Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA', *Nucleic Acids Research* **38**(6), e87.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S. M., Sarkissian, C. D., Ganslmeier, R., Friederich, S., Dresely, V., Van Oven, M., Kenyon, R., Van Der Hoek, M. B., Korlach, J., Luong, K., Ho, S. Y., Quintana-Murci, L., Behar, D. M., Meller, H., Alt, K. W., Cooper, A., Adhikarla, S., Ganesh Prasad, A. K., Pitchappan, R., Varatharajan Santhakumari, A., Balanovska, E., Balanovsky, O., Bertranpetit, J., Comas, D., Martínez-Cruz, B., Melé, M., Clarke, A. C., Matisoo-Smith, E. A., Dulik, M. C., Gaieski, J. B., Owings, A. C., Schurr, T. G., Vilar, M. G., Hobbs, A., Soodyall, H., Javed, A., Parida, L., Platt, D. E., Royyuru, A. K., Jin, L., Li, S., Kaplan, M. E., Merchant, N. C., John Mitchell, R., Renfrew, C., Lacerda, D. R., Santos, F. R., Soria Hernanz, D. F., Spencer Wells, R., Swamikrishnan, P., Tyler-Smith, C., Paulo Vieira, P. and Ziegle, J. S. (2013), 'Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans', *Nature Communications* **4**, 1764.

Broushaki, F., Thomas, M. G., Link, V., López, S., van Dorp, L., Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L. M., Díez-del Molino, D., Kousathanas, A., Sell, C., Robson, H. K., Martiniano, R., Blöcher, J., Scheu, A., Kreutzer, S., Bollongino, R., Bobo, D., Davoudi, H., Munoz, O., Currat, M., Abdi, K., Biglari, F., Craig, O. E., Bradley, D. G., Shennan, S., Veeramah, K. R., Mashkour, M., Wegmann, D., Hellenthal, G. and Burger, J. (2016), 'Early Neolithic genomes from the eastern Fertile Crescent', *Science* **53**(6298), 499–503.

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), 1084–1097.

Burns, R. I. (1973), *Islam under the Crusaders: Colonial Survival in the Thirteenth-Century Kingdom of Valencia*, Princeton University Press.

Burns, R. I. (1975), *Medieval Colonialism: Postcrusade Exploitation of Islamic Valencia*, Princeton University Press.

Burns, R. I. (1991), *Diplomatarium of the Crusader Kingdom of Valencia*, Princeton University Press.

Butzer, K. W. (1988), 'Cattle and Sheep from Old to New Spain: Historical Antecedents', *Annals of the Association of American Geographers* **78**(1), 29–56.

Butzer, K. W., Mateu, J. F., Butzer, E. K. and Kraus, P. (1985), 'Irrigation Agrosystems in Eastern Spain: Roman or Islamic Origins?', *Annals of the Association of American Geographers* **75**(4), 479–509.

Bycroft, C., Fernández-Rozadilla, C., Ruiz-Ponte, C., Quintela-García, I., Carracedo, Á., Donnelly, P. and Myers, S. (2018), 'Patterns of genetic differentiation and the footprints of historical migrations in the Iberian Peninsula', *Nature Communications* **10**, 551.

Cabanes Pecourt, M. D. (1977), *El Repartiment de la ciudad de Valencia*, ANUBAR Ediciones.

Cadiou, F. (2008), *Hibera en terra miles: les armées romaines et la conquête de l'Hispanie sous la république (218-45 av. J.-C.)*, Casa de Velázquez, Madrid.

Cann, R. L., Stoneking, M. and Wilson, A. C. (1987), 'Mitochondrial DNA and human evolution', *Nature* **325**(6099), 31–36.

Cardoso, S., Alfonso-Sánchez, M. A., Valverde, L., Odriozola, A., Pérez-Miranda, A. M., Peña, J. A. and De Pancorbo, M. M. (2011), 'The maternal legacy of Basques in northern navarre: New insights into the mitochondrial DNA diversity of the Franco-Cantabrian area', *American Journal of Physical Anthropology* **145**(3), 480–488.

Cardoso, S., Valverde, L., Alfonso-Sánchez, M. A., Palencia-Madrid, L., Elcoroaristizabal, X., Algorta, J., Catarino, S., Arteta, D., Herrera, R. J., Zarrabeitia, M. T., Peña, J. A. and de Pancorbo, M. M. (2013), 'The expanded mtdna phylogeny of the franco-cantabrian region upholds the pre-neolithic genetic substrate of basques', *PLoS ONE* **8**(7), 1–9.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova, D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., Gilbert, M. T. P., Willerslev, E., Greenleaf, W. J. and Bustamante, C. D. (2013), 'Pulling out the 1%: Whole-Genome capture for the targeted enrichment of ancient dna sequencing libraries', *American Journal of Human Genetics* **93**(5), 852–864.

Carter, R. W. (2007), 'Mitochondrial diversity within modern human populations', *Nucleic Acids Research* **35**(9), 3039–3045.

Cassidy, L. M., Maoldúin, R. Ó., Kador, T., Lynch, A., Jones, C., Woodman, P. C., Murphy, E., Ramsey, G., Dowd, M., Noonan, A., Campbell, C., Jones, E. R., Mattiangeli, V. and Bradley, D. G. (2020), 'A dynastic elite in monumental Neolithic society', *Nature* **582**, 384–388.

Cassidy, L. M., Martiniano, R., Murphy, E. M., Teasdale, M. D., Mallory, J., Hartwell, B. and Bradley, D. G. (2016), 'Neolithic and bronze age migration to ireland and establishment of the insular atlantic genome', *Proceedings of the National Academy of Sciences* **113**(2), 368–373.

Casson, L. and Wilson, A. J. N. (1966), *Emigration from Italy in the Republican Age of Rome*, Vol. 72, Manchester University Press.

Catalano, G., Lo Vetro, D., Fabbri, P. F., Mallick, S., Reich, D., Rohland, N., Sineo, L., Mathieson, I. and Martini, F. (2020), 'Late Upper Palaeolithic hunter-gatherers in the Central Mediterranean: New archaeological and genetic data from the Late Epigravettian burial Oriente C (Favignana, Sicily)', *Quaternary International* **537**, 24–32.

Cavalli-Sforza, L. L. (1998), 'The DNA revolution in population genetics', *Trends in Genetics* **14**(2), 60–65.

Cavalli-Sforza, L. L. and Edwards, A. W. F. (1967), 'Phylogenetic Analysis Models and Estimation Procedures', *American journal of human genetics* **19**, 233–257.

Cavalli-Sforza, L. L. and Feldman, M. W. (2003), 'The application of molecular genetic approaches to the study of human evolution', **33**(3S), 266–275.

Cavalli-Sforza, L. L., Menozzi, P. and Piazza, A. (1994), *The History and Geography of Human Genes*, Princeton University Press.

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. and Wilson, J. F. (2018), 'Runs of homozygosity: Windows into population history and trait architecture', **19**(4), 220–234.

Cha, S., Lim, J. E., Park, A. Y., Do, J. H., Lee, S. W., Shin, C., Cho, N. H., Kang, J. O., Nam, J. M., Kim, J. S., Woo, K. M., Lee, S. H., Kim, J. Y. and Oh, B. (2018), 'Identification of five novel genetic loci related to facial morphology by genome-wide association studies', *BMC Genomics* **19**(1), 481.

Chacón-Duque, J. C., Adhikari, K., Fuentes-Guajardo, M., Mendoza-Revilla, J., Acuña-Alonzo, V., Barquera, R., Quinto-Sánchez, M., Gómez-Valdés, J., Everardo Martínez, P., Villamil-Ramírez, H., Hünemeier, T., Ramallo, V., Silva de Cerqueira, C. C., Hurtado, M., Villegas, V., Granja, V., Villena, M., Vásquez, R., Llop, E., Sandoval, J. R., Salazar-Granara, A. A., Parolin, M. L., Sandoval, K., Peñaloza-Espinosa, R. I., Rangel-Villalobos, H., Winkler, C. A., Klitz, W., Bravi, C., Molina, J., Corach, D., Barrantes, R., Gomes, V., Resende, C., Gusmão, L., Amorim, A., Xue, Y., Dugoujon, J. M., Moral, P., González-José, R., Schuler-Faccini, L., Salzano, F. M., Bortolini, M. C., Canizales-Quinteros, S., Poletti, G., Gallo, C., Bedoya, G., Rothhammer, F., Balding, D., Hellenthal, G. and Ruiz-Linares, A. (2018), 'Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance', *Nature Communications* **9**(1), 5388.

Cheng, K. C. and Canfield, V. A. (2006), 'The role of SLC24A5 in skin color', *Experimental Dermatology* **15**(10), 836–838.

Chyleński, M., Juras, A., Ehler, E., Malmström, H., Piontek, J., Jakobsson, M., Marciniak, A. and Dabert, M. (2017), 'Late Danubian mitochondrial genomes shed light into the Neolithisation of Central Europe in the 5th millennium BC', *BMC Evolutionary Biology* **17**, 80.

Claes, P., Liberton, D. K., Daniels, K., Rosana, K. M., Quillen, E. E., Pearson, L. N., McEvoy, B., Bauchet, M., Zaidi, A. A., Yao, W., Tang, H., Barsh, G. S., Absher, D. M., Puts, D. A., Rocha, J., Beleza, S., Pereira, R. W., Baynam, G., Suetens, P., Vandermeulen, D., Wagner, J. K., Boster, J. S. and Shriver, M. D. (2014), 'Modeling 3d facial shape from dna', *PLOS Genetics* **10**(3), 1–14.

Claes, P., Roosenboom, J., White, J. D., Swigut, T., Sero, D., Li, J., Lee, M. K., Zaidi, A., Mattern, B. C., Liebowitz, C., Pearson, L., González, T., Leslie, E. J., Carlson, J. C., Orlova, E., Suetens, P., Vandermeulen, D., Feingold, E., Marazita, M. L., Shaffer, J. R., Wysocka, J., Shriver, M. D.

and Weinberg, S. M. (2018), 'Genome-wide mapping of global-to-local genetic effects on human facial shape', *Nature Genetics* **50**(3), 414–423.

Cline, E. H. (2014), *1177 B.C.: The year civilization collapsed*, Princeton University Press.

Cole, J. B., Manyama, M., Larson, J. R., Liberton, D. K., Ferrara, T. M., Riccardi, S. L., Li, M., Mio, W., Klein, O. D., Santorico, S. A., Hallgrímsson, B. and Spritz, R. A. (2017), 'Human facial shape and size heritability and genetic correlations', *Genetics* **205**(2), 967–978.

Consortium, I. H. G. S. (2001), 'Initial sequencing and analysis of the human genome', *Nature* **409**, 860–921.

Cooper, A. (2000), 'Ancient DNA: Do It Right or Not at All', *Science* **289**(5482), 1139.

Corell, J. (2012), *Inscripcions romanes del País Valencià*, Universitat de Valencia.

Cortés Alonso, V. (1972), 'Procedencia de los esclavos negros en Valencia (1482 - 1516)', *Revista Española De Antropología Americana* **7**(1), 123–152.

Cortés López, J. L. (1989), *La esclavitud negra en la España peninsular del siglo XVI*, Universidad de Salamanca.

Cortés-Sánchez, M., Jiménez-Espejo, F. J., Simón-Vallejo, M. D., Stringer, C., Lozano Francisco, M. C., García-Alix, A., Vera Peláez, J. L., Odriozola, C. P., Riquelme-Cantal, J. A., Parrilla Giráldez, R., Maestro González, A., Ohkouchi, N. and Morales-Muñiz, A. (2019), 'An early Aurignacian arrival in southwestern Europe', *Nature Ecology and Evolution* **3**(2), 207–212.

Coscollá Sanz, V. (2003), *La Valencia musulmana*, Carena Editors.

Craig Venter, J., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Yuan Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M. H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S. C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Lai Cheng, M., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Ni Tint, N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak,

B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Deslattes Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001), 'The sequence of the human genome', *Science* **291**(5507), 1304–1351.

Crawford, N. G., Kelly, D. E., Hansen, M. E., Beltrame, M. H., Fan, S., Bowman, S. L., Jewett, E., Ranciaro, A., Thompson, S., Lo, Y., Pfeifer, S. P., Jensen, J. D., Campbell, M. C., Beggs, W., Hormozdiari, F., Mpoloka, S. W., Mokone, G. G., Nyambo, T., Meskel, D. W., Belay, G., Haut, J., Rothschild, H., Zon, L., Zhou, Y., Kovacs, M. A., Xu, M., Zhang, T., Bishop, K., Sinclair, J., Rivas, C., Elliot, E., Choi, J., Li, S. A., Hicks, B., Burgess, S., Abnet, C., Watkins-Chow, D. E., Oceana, E., Song, Y. S., Eskin, E., Brown, K. M., Marks, M. S., Loftus, S. K., Pavan, W. J., Yeager, M., Chanock, S. and Tishkoff, S. A. (2017), 'Loci associated with skin pigmentation identified in African populations', *Science* **358**(6365), eaan8433.

Currás, B. X. and Sastre, I. (2019), 'Egalitarianism and resistance: A theoretical proposal for Iron Age Northwestern Iberian archaeology', *Anthropological Theory* **20**(3), 300–309.

Currat, M. and Excoffier, L. (2005), 'The effect of the Neolithic expansion on European molecular diversity', *Proceedings of the Royal Society B: Biological Sciences* **272**, 679–688.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M., Weihmann, A., Nickel, B. and Valdiosera, C. (2013), 'Complete Mitochondrial Genome Sequence of a Middle Pleistocene Cave Bear Reconstructed from Ultrashort DNA Fragments', *Proceedings of the National Academy of Sciences of the United States of America* **110**(39), 15758–15763.

Damgaard, P. B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E. and Allentoft, M. E. (2015), 'Improving access to endogenous DNA in ancient bones and teeth', *Scientific Reports* **5**, 11184.

Darwin, C. (1871), *The descent of man and selection in relation to sex*, Vol. 2, John Murray.

Díaz Borrás, A. (1993), *Los orígenes de la piratería islámica en Valencia: la ofensiva musulmana trecentista y la reacción cristiana*, Consejo Superior de Investigaciones Científicas.

de Barros Damgaard, P. and et al (2018*a*), '137 ancient human genomes from across the Eurasian steppes', *Nature* **557**, 369–374.

de Barros Damgaard, P. and et al (2018*b*), 'The first horse herders and the impact of early bronze age steppe expansions into asia', *Science* **360**(6396), eaar7711.

de Manuel, M., Barnett, R., Sandoval-Velasco, M., Yamaguchi, N., Garrett Vieira, F., Zepeda Mendoza, M. L., Liu, S., Martin, M. D., Sinding, M.-H. S., Mak, S. S. T., Carøe, C., Liu, S., Guo, C., Zheng, J., Zazula, G., Baryshnikov, G., Eizirik, E., Koepfli, K.-P., Johnson, W. E., Antunes, A., Sicheritz-Ponten, T., Gopalakrishnan, S., Larson, G., Yang, H., O'Brien, S. J., Hansen, A. J., Zhang, G., Marques-Bonet, T. and Gilbert, M. T. P. (2020), 'The evolutionary history of extinct and living lions', *Proceedings of the National Academy of Sciences* **117**(20), 10927–10934.

Deng, L. and Xu, S. (2018), 'Adaptation of human skin color in various populations', **155**, 1.

Diamond, J. and Bellwood, P. (2003), 'Farmers and their languages: The first expansions'.

Doce, E. G. (2006), 'Sobre la función y el significado de la cerámica Campaniforme a la luz de los análises de contenidos', *Trabajos de Prehistoria* **63**(1), 69–84.

Dominguez Monedero, A. J. (1996), *Los Griegos en la Península Ibérica*, Arco Libros.

Ebenesersdóttir, S. S., Sandoval-Velasco, M., Gunnarsdóttir, E. D., Jagadeesan, A., Guðmundsdóttir, V. B., Thordardóttir, E. L., Einarsdóttir, M. S., Moore, K. H. S., Sigurðsson, Á., Magnúsdóttir, D. N., Jónsson, H., Snorradóttir, S., Hovig, E., Møller, P., Kockum, I., Olsson, T., Alfredsson, L., Hansen, T. F., Werge, T., Cavalleri, G. L., Gilbert, E., Lalueza-Fox, C., Walser, J. W., Kristjánsdóttir, S., Gopalakrishnan, S., Árnadóttir, L., Magnússon, Ó. ., Gilbert, M. T. P., Stefánsson, K. and Helgason, A. (2018), 'Ancient genomes from Iceland reveal the making of a human population.', *Science* **360**(6392), 1028–1032.

Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M. B. and Rollo, F. (2008), 'Complete Mitochondrial Genome Sequence of the Tyrolean Iceman', *Current Biology* **18**(21), 1687–1693.

Evans, D. M. (2018), 'Elucidating the genetics of craniofacial shape', *Nature Genetics* **50**(3), 319–321.

Feldman, M., Fernández-Domínguez, E., Reynolds, L., Baird, D., Pearson, J., Hershkovitz, I., May, H., Goring-Morris, N., Benz, M., Gresky, J., Bianco, R. A., Fairbairn, A., Mustafaoğlu, G., Stockhammer, P. W., Posth, C., Haak, W., Jeong, C. and Krause, J. (2019), 'Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia', *Nature Communications* **10**(1), 1218.

Fernandes, D. M., Strapagiel, D., Borówka, P., Marciniak, B., Żądzińska, E., Sirak, K., Siska, V., Grygiel, R., Carlsson, J., Manica, A., Lorkiewicz, W. and Pinhasi, R. (n.d.), 'A genomic Neolithic time transect of hunter-farmer admixture in central Poland', *Scientific Reports* **8**(1), 14879.

Fernandes, V., Alshamali, F., Alves, M., Costa, M. D., Pereira, J. B., Silva, N. M., Cherni, L., Harich, N., Cerny, V., Soares, P., Richards, M. B. and Pereira, L. (2012), 'The Arabian cradle: Mitochondrial relicts of the first steps along the Southern route out of Africa', *American Journal of Human Genetics* **90**(2), 347–355.

Finlayson, C., Giles Pacheco, F., Rodríguez-Vidal, J., Fa, D. A., María Gutierrez López, J., Santiago Pérez, A., Finlayson, G., Allue, E., Baena Preysler, J., Cáceres, I., Carrión, J. S., Fernández Jalvo, Y., Gleed-Owen, C. P., Jimenez Espejo, F. J., López, P., Antonio López Sáez, J., Antonio Riquelme Cantal, J., Sánchez Marco, A., Giles Guzman, F., Brown, K., Fuentes, N., Valarino, C. A., Villalpando, A., Stringer, C. B., Martinez Ruiz, F. and Sakamoto, T. (2006), 'Late survival of Neanderthals at the southernmost extreme of Europe', *Nature* **443**(7113), 850–853.

Finocchio, A., Trombetta, B., Messina, F., D'Atanasio, E., Akar, N., Loutradis, A., Michalodimitrakis, E. I., Cruciani, F. and Novelletto, A. (2018), 'A finely resolved phylogeny of y chromosome Hg J illuminates the processes of Phoenician and Greek colonizations in the Mediterranean', *Scientific Reports* **8**(1), 7465.

Fitzpatrick, A. P. (2011), *The Amesbury Archer and the Boscombe Bowmen : Bell Beaker burials on Boscombe Down, Amesbury, Wiltshire*, number 27, Wessex Archaeology.

Fitzpatrick, T. (1975), 'Soleil et peaus', *Journal de Medecine Esthetique* **2**, 33–34.

Fregel, R., Méndez, F. L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M. D., Santana, J., Morales, J., Ávila-Arcos, M. C., Underhill, P. A., Shapiro, B., Wojcik, G., Rasmussen, M., Soares, A. E. R., Kapp, J., Sockell, A., Rodríguez-Santos, F. J., Mikdad, A., Trujillo-Mederos, A. and Bustamante, C. D. (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* **115**(26), 6774–6779.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., Nickel, B., Peltzer, A., Rohland, N., Slon, V., Talamo, S., Lazaridis, I., Lipson, M., Mathieson, I., Schiffels, S., Skoglund, P., Derevianko, A. P., Drozdov, N., Slavinsky, V., Tsybankov, A., Cremonesi, R. G., Mallegni, F., Gély, B., Vacca, E., Morales, M. R., Straus, L. G., Neugebauer-Maresch, C., Teschler-Nicola, M., Constantin, S., Moldovan, O. T., Benazzi, S., Peresani, M., Coppola, D., Lari, M., Ricci, S., Ronchitelli, A., Valentin, F., Thevenet, C., Wehrberger, K., Grigorescu, D., Rougier, H., Crevecoeur, I., Flas, D., Semal, P., Mannino, M. A., Cupillard, C., Bocherens, H., Conard, N. J., Harvati, K., Moiseyev, V., Drucker, D. G., Svoboda, J., Richards, M. P., Caramelli, D., Pinhasi, R., Kelso, J., Patterson, N., Krause, J., Pääbo, S. and Reich, D. (2016), 'The genetic history of Ice Age Europe', **534**(7606), 200–205.

Furtwängler, A., Rohrlach, A. B., Lamnidis, T. C., Papac, L., Neumann, G. U., Siebke, I., Reiter, E., Steuri, N., Hald, J., Denaire, A., Schnitzler, B., Wahl, J., Ramstein, M., Schuenemann, V. J., Stockhammer, P. W., Hafner, A., Lösch, S., Haak, W., Schiffels, S. and Krause, J. (2020), 'Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland', *Nature Communications* **11**(1), 1915.

Gabarda, M. V. and Aguilella, G. (2016), *Sima del Pozo Cerdaña (Pina de Montalgrao, Alto Palancia, Castellón): una cavidad sepulcral eneolítica*, Diputació de Castelló.

Gamba, C., Fernandez, E., Tirado, M., Deguilloux, M. F., Pemonge, M. H., Utrilla, P., Edo, M., Molist, M., Rasteiro, R., Chikhi, L. and Arroyo-Pardo, E. (2012), 'Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers', *Molecular Ecology* **21**, 45–56.

Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., Bradley, D. and Orlando, L. (2016), 'Comparing the Performance of Three Ancient DNA Extraction Methods for High-Throughput Sequencing', *Molecular Ecology Resources* **16**(2), 459–69.

Gamble, C., Davies, W., Pettitt, P. and Richards, M. (2004), 'Climate change and evolving human diversity in Europe during the last glacial', *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 243–254.

Gandini, F., Achilli, A., Pala, M., Bodner, M., Brandini, S., Huber, G., Egyed, B., Ferretti, L., Gómez-Carballa, A., Salas, A., Scozzari, R., Cruciani, F., Coppa, A., Parson, W., Semino, O., Soares, P., Torroni, A., Richards, M. B. and Olivieri, A. (2016), 'Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes', *Scientific Reports* **6**, 25472.

Gangal, K., Sarson, G. R. and Shukurov, A. (2014), 'The near-eastern roots of the neolithic in South Asia', *PLoS ONE* **9**(5), e95714.

García-Bellido, P. (2014), *Grafitos sobre un shekel de plata del Tesoro de Mogente (Valencia)*, Anejos de Erytheia, Madrid.

García-Martínez de Lagrán, Í., Fernández-Domínguez, E. and Rojo-Guerra, M. A. (2018), 'Solutions or illusions? An analysis of the available palaeogenetic evidence from the origins of the Neolithic in the Iberian Peninsula', *Quaternary International* **470**, 353–368.

García-Prósper, E. (2016), *Los ritos funerarios de la necrópolis romana de la calle Quart de Valencia (Siglos II a.C-III d.C). PhD thesis.*, Universitat de València.

García-Prósper, E. and Polo-Cerdá, M. (2016), 'Los primeros pobladores de Valentia. Un proyecto transversal de ritual funerario y bioantropología.', *Arqueología Somos Todos* **5**, 18–20.

García-Prósper, E. and Polo-Cerdá, M. (2020), 'Memory written in bones from funus rites to the osteobiography of valentia', *Metode* **2020**(10).

García-Prósper, E., Polo-Cerdá, M. and Guérin, P. (2003), 'Rituales funerarios ibéricos en la necrópolis fundacional de Valentia', *Anales de Arqueología Cordobesa.* **13-14**, 279–310.

García-Prósper, E., Polo-Cerdá, M., Romero, A. and Iborra, P. (2010), 'Rituales alimentarios y economía de subsistencia en las tumbas de cámara de la necrópolis romana de la calle Quart de Valentia (ss. II a. C - III d. C).', *Saguntum extra* **9**, 233–242.

García Sanjuán, L. (1999), *Los origenes de la estratificacion social: patrones de desigualdad en la Edad del Bronce del suroeste de la Peninsula Iberica (Sierra Morena Occidental c. 1700-1100 A.N.E. - 2100-1300 A.N.E.)*, Archaeopress.

Gibbons, A. (2017), 'How Africans evolved a palette of skin tones', **358**(6360), 157–158.

Gilman, A. (1976), 'Bronze Age Dynamics in Southeast Spain', **1**, 307–319.

Gleize, Y., Mendisco, F., Pemonge, M. H., Hubert, C., Groppi, A., Houix, B., Deguilloux, M. F. and Breuil, J. Y. (2016), 'Early medieval Muslim graves in France: First archaeological, anthropological and palaeogenomic evidence', *PLoS ONE* **11**(2), e0148583.

Glick, T. F. (1970), 'The Crusader Kingdom of Valencia: Reconstruction on a Thirteenth-Century Frontier. Robert Ignatius Burns', *The American Historical Review* **75**(5), 1442.

Glick, T. F. (1977), ' Islam under the Crusaders: Colonial Survival in the Thirteenth-Century Kingdom of Valencia . Robert Ignatius Burns ', *Speculum* **52**(1), 128–129.

Goldberg, A., Günther, T., Rosenberg, N. A. and Jakobsson, M. (2016), 'Familial migration of the Neolithic contrasts massive male migration during Bronze Age in Europe inferred from ancient X chromosomes', *bioRxiv* .

Gómez-Sánchez, D., Olalde, I., Pierini, F., Matas-Lalueza, L., Gigli, E., Lari, M., Civit, S., Lozano, M., Vergès, J. M., Caramelli, D., Ramírez, O. and Lalueza-Fox, C. (2014), 'Mitochondrial DNA from El Mirador cave (Atapuerca, Spain) reveals the heterogeneity of Chalcolithic populations', *PLoS ONE* **9**, e105105.

González-Fortes, G., Jones, E. R., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., Garralda, M. D., Drak, L., Siska, V., Simalcsik, A., Boroneanţ, A., Vidal Romaní, J. R., Vaqueiro Rodríguez, M., Arias, P., Pinhasi, R., Manica, A. and Hofreiter, M. (2017), 'Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin', *Current Biology* **27**, 1801–1810.

González-Fortes, G., Tassi, F., Trucchi, E., Henneberger, K., Paijmans, J. L., Díez-Del-Molino, D., Schroeder, H., Susca, R. R., Barroso-Ruíz, C., Bermudez, F. J., Barroso-Medina, C., Bettencourt, A. M., Sampaio, H. A., Grandal-D'Anglade, A., Salas, A., De Lombera-Hermida, A., Fabregas Valcarce, R., Vaquero, M., Alonso, S., Lozano, M., Rodríguez-Alvarez, X. P., Fernández-Rodríguez, C., Manica, A., Hofreiter, M. and Barbujani, G. (2019), 'A western route of prehistoric human migration from Africa into the Iberian Peninsula', *Proceedings of the Royal Society B: Biological Sciences* **286**(1895), 20182288.

Grau-Mira, I. (2019), 'Power on the hills. Warfare, symbolic violence and landscape in the eastern Iberian Iron Age', *Journal of Anthropological Archaeology* **53**, 147–160.

Graullera Sanz, V. (1978), *La esclavitud en Valencia en los siglos XVI y XVII*, Instituto Valenciano de Estudios Históricos, Institución Alfonso el Magnánimo, Diputación Provincial, Consejo Superior de Investigaciones Científicas, Valencia.

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

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H. Y., Hansen, N. F., Durand, E. Y., Malaspinas, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Ž., Gušic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., De La Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D. and Pääbo, S. (2010), 'A draft sequence of the neandertal genome', *Science* **328**(5979), 710–722.

Guardia, J., Maragall, M., Mercadal, O., Olesti i Vila, O., Galbany i Casals, J. and Nadal Lorenzo, J. (2007), 'Enterrament d'epoca tardoromana d'un macaco amb aixovar al jaciment de Les Colomines', *Empúries* **55**, 199–227.

Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınç, G. M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., Munters, A. R., Coutinho, A., Simões, L. G., Vicente, M., Sjölander, A., Jansen Sellevold, B., Jørgensen, R., Claes, P., Shriver, M. D., Valdiosera, C., Netea, M. G., Apel, J., Lidén, K., Skar, B., Storå, J., Götherström, A. and Jakobsson, M. (2018), 'Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation', *PLoS Biology* **16**, e2003703.

Günther, T., Valdiosera, C., Malmström, H., Ureña, I., Rodriguez-Varela, R., Sverrisdóttir, Ó. O., Daskalaki, E. A., Skoglund, P., Naidoo, T., Svensson, E. M., Bermúdez de Castro, J. M., Carbonell, E., Dunn, M., Storå, J., Iriarte, E., Arsuaga, J. L., Carretero, J.-M., Götherström, A. and Jakobsson, M. (2015), 'Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques', *Proceedings of the National Academy of Sciences* **112**, 11917–11922.

Haak, W., Balanovsky, O., Sanchez, J. J., Koshel, S., Zaporozhchenko, V., Adler, C. J., der Sarkissian, C. S., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., Villems, R., Meller, H., Alt, K. W. and Cooper, A. (2010), 'Ancient DNA from European early Neolithic farmers reveals their near eastern affinities', *PLoS Biology* **8**.

Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K. W. and Burger, J. (2005), 'Ancient DNA from the first European farmers in 7500-year-old neolithic sites', *Science* **310**(5750), 1016–1018.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S. L., Risch, R., Rojo Guerra, M. A., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K. W. and Reich, D. (2015), 'Massive migration from the steppe was a source for Indo-European languages in Europe', *Nature* **522**(7555), 207–211.

Haber, M., Doumet-Serhal, C., Scheib, C. L., Xue, Y., Mikulski, R., Martiniano, R., Fischer-Genz, B., Schutkowski, H., Kivisild, T. and Tyler-Smith, C. (2019), 'A Transient Pulse of Genetic Admixture from the Crusaders in the Near East Identified from Ancient Genome Sequences', *American Journal of Human Genetics* **104**(5), 977–984.

Haber, M., Doumet-Serhal, C., Scheib, C., Xue, Y., Danecek, P., Mezzavilla, M., Youhanna, S., Martiniano, R., Prado-Martinez, J., Szpak, M., Matisoo-Smith, E., Schutkowski, H., Mikulski, R., Zalloua, P., Kivisild, T. and Tyler-Smith, C. (2017), 'Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences', *American Journal of Human Genetics* **101**, 274–282.

Haber, M., Mezzavilla, M., Xue, Y. and Tyler-Smith, C. (2016), 'Ancient DNA and the rewriting of human history: Be sparing with Occam's razor', *Genome Biology* **17**, 1–8.

Haber, M., Nassar, J., Almarri, M., Saupe, T., Saag, L., Griffith, S., Doumet-Serhal, C., Chanteau, J., Saghieh-Beydoun, M., Xue, Y., Scheib, C. and Tyler-Smith, C. (2020), 'A Genetic History of the Near East from an aDNA Time Course Sampling Eight Points in the Past 4,000 Years', *The American Journal of Human Genetics* **107**(1), 149–157.

Hagelberg, E., Hofreiter, M. and Keyser, C. (2015), 'Ancient DNA: the first three decades', *Philosophical Transactions of the Royal Society B* **370**, 20130371.

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

Halperin Donghi, T. (1980), *Un conflicto nacional: Moriscos y cristianos viejos en Valencia*, Institución Alfonso El Magnánimo, Valencia.

Hammer, M. F., Karafet, T., Rasanayagam, A., Wood, E. T., Altheide, T. K., Jenkins, T., Griffiths, R. C., Templeton, A. R. and Zegura, S. L. (1998), 'Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation', *Molecular Biology and Evolution* **15**(4), 427–441.

Hammer, Ø., Harper, D. A. and Ryan, P. D. (2001), 'Past: Paleontological statistics software package for education and data analysis', *Palaeontologia Electronica* **4**(1), 1–9.

Hansen, H. B., Damgaard, P. B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E. and Allentoft, M. E. (2017), 'Comparing ancient DNA preservation in petrous bone and tooth cementum', *PLoS ONE* **12**, e170940.

Harney, É., Patterson, N., Reich, D. and Wakeley, J. (2020), 'Assessing the performance of qpadm: A statistical tool for studying population admixture', *bioRxiv* .

Harrison, R. J. (1974), 'Origins of the Bell Beaker cultures', *Antiquity* **48**, 99–109.

Hellenthal, G., Busby, G. B., Band, G., Wilson, J. F., Capelli, C., Falush, D. and Myers, S. (2014), 'A genetic atlas of human admixture history', *Science* **343**, 747–751.

Henn, B. M., Botigué, L. R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J. K., Fadhlaoui-Zid, K., Zalloua, P. A., Moreno-Estrada, A., Bertranpetit, J., Bustamante, C. D. and Comas, D. (2012), 'Genomic ancestry of North Africans supports back-to-Africa migrations', *PLoS Genetics* **8**(1), e1002397.

Hernández, C. L., Dugoujon, J. M., Novelletto, A., Rodríguez, J. N., Cuesta, P. and Calderón, R. (2017), 'The distribution of mitochondrial DNA haplogroup H in southern Iberia indicates ancient human genetic exchanges along the western edge of the Mediterranean', *BMC Genetics* **18**, 46–60.

Hervella, M., Izagirre, N., González-Morales, M., Straus, L. G., Fregel, R. I. and De La Rúa, C. (2014), 'El ADN mitocondrial de los cazadores-recolectores de la región cantábrica: Nueva evidencia de la cueva de El Mirón (Ramales de la Victoria, Cantabria, España)', *Revista Española de Antropologia Fisica* **35**, 11–21.

Hervella, M., Rotea, M., Izagirre, N., Constantinescu, M., Alonso, S., Ioana, M., Lazar, C., Ridiche, F., Soficaru, A. D., Netea, M. G. and De-La-Rua, C. (2015), 'Ancient DNA from South-East Europe reveals different events during early and middle neolithic influencing the European genetic heritage', *PLoS ONE* **10**, e0128810.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A. and Wilson, A. C. (1984), 'DNA sequences from the quagga, an extinct member of the horse family', *Nature* **312**(5991), 282–284.

Hillgarth, J. (1980), *Popular Religion in Visigothic Spain*, Clarendon Press.

Hoffecker, J. F., Holliday, V. T., Anikovich, M. V., Sinitsyn, A. A., Popov, V. V., Lisitsyn, S. N., Levkovskaya, G. M., Pospelova, G. A., Forman, S. L. and Giaccio, B. (2008), 'From the Bay of Naples to the River Don: the Campanian Ignimbrite eruption and the Middle to Upper Paleolithic transition in Eastern Europe', *Journal of Human Evolution* **55**(5), 858–870.

Holt, R. D. (2009), 'Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives', *Proceedings of the National Academy of Sciences of the United States of America* **106**(2), 19659–19665.

Hutchinson, G. E. (1957), 'Concluding Remarks', *Cold Spring Harbor Symposia on Quantitative Biology* **22**, 415–427.

Immel, A., Terna, S., Simalcsik, A., Susat, J., Sarov, O., Sirbu, G., Hofmann, R., Muller, J., Nebel, A. and Krause-Kyora, B. (2020), 'Gene-flow from steppe individuals into Cucuteni-Trypillia associated populations indicates long-standing contacts and gradual admixture', *Scientific Reports* **10**(1), 4253.

Isern, N., Zilhão, J., Fort, J. and Ammerman, A. J. (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), 897–902.

Jantzen, D., Brinker, U., Orschiedt, J., Heinemeier, J., Piek, J., Hauenstein, K., Krüger, J., Lidke, G., Lübke, H., Lampe, R., Lorenz, S., Schult, M. and Terberger, T. (2011), 'A Bronze Age battlefield? Weapons and trauma in the Tollense Valley, north-eastern Germany', *Antiquity* **328**, 417–433.

Jensen, T. Z., Niemann, J., Iversen, K. H., Fotakis, A. K., Gopalakrishnan, S., Vågene, Å. J., Pedersen, M. W., Sinding, M. H. S., Ellegaard, M. R., Allentoft, M. E., Lanigan, L. T., Taurozzi, A. J., Nielsen, S. H., Dee, M. W., Mortensen, M. N., Christensen, M. C., Sørensen, S. A., Collins, M. J., Gilbert, M. T. P., Sikora, M., Rasmussen, S. and Schroeder, H. (2019), 'A 5700 year-old human genome and oral microbiome from chewed birch pitch', *Nature Communications* **10**(1), 5520.

Jeong, C., Wilkin, S., Amgalantugs, T., Bouwman, A. S., Taylor, W. T. T., Hagan, R. W., Bromage, S., Tsolmon, S., Trachsel, C., Grossmann, J., Littleton, J., Makarewicz, C. A., Krigbaum, J., Burri, M., Scott, A., Davaasambuu, G., Wright, J., Irmer, F., Myagmar, E., Boivin, N., Robbeets, M., Rühli, F. J., Krause, J., Frohlich, B., Hendy, J. and Warinner, C. (2018), 'Bronze age population dynamics and the rise of dairy pastoralism on the eastern eurasian steppe', *Proceedings of the National Academy of Sciences* **115**(48), E11248–E11255.

Jesse, R., Véla, E. and Pfenninger, M. (2011), 'Phylogeography of a Land Snail suggests Trans-Mediterranean Neolithic transport', *PLoS ONE* **6**, e020734.

Jeunesse, C. (2015), 'The dogma of the Iberian origin of the Bell Beaker: attempting its deconstruction', *Journal of Neolithic Archaeology* **16**, 158–166.

Jiménez Salvador, J. L., Ribera i Lacomba, A. V. and Rosselló Mesquida, M. (2014), Valentia y su territorium desde época romana imperial hasta la Antigüedad Tardía, *in* 'Ciudad y territorio: transformaciones materiales e ideológicas entre la época clásica y el Altomedioevo', Universidad de Córdoba.

Jobling, M. A. and Tyler-Smith, C. (2003), 'The human Y chromosome: An evolutionary marker comes of age', **4**(8), 598–612.

Jobling, M. A. and Tyler-Smith, C. (2017), 'Human Y-chromosome variation in the genome-sequencing era', *Nature Reviews Genetics* **18**(8), 485–497.

Jobling, M., Tyler-Smith, C., Hollox, E. and Kivisild, T. (2013), *Human Evolutionary Genetics*, Garland Science.

Jones, E. R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R. L., Gallego Llorente, M., Cassidy, L. M., Gamba, C., Meshveliani, T., Bar-Yosef, O., Müller, W., Belfer-Cohen, A., Matskevich, Z., Jakeli, N., Higham, T. F., Currat, M., Lordkipanidze, D., Hofreiter, M., Manica, A., Pinhasi, R. and Bradley, D. G. (2015), 'Upper Palaeolithic genomes reveal deep roots of modern Eurasians', *Nature Communications* **6**, 8912.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. and Orlando, L. (2013), MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters, *in* 'Bioinformatics', Vol. 29, pp. 1682–1684.

Jukes, T. and Cantor, C. (1969), *Mammalian Protein Metabolism*, Academic Press, New York.

Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M. A., Järve, M., Talas, U. G., Rootsi, S., Ilumäe, A. M., Mägi, R., Mitt, M., Pagani, L., Puurand, T., Faltyskova, Z., Clemente, F., Cardona, A., Metspalu, E., Sahakyan, H., Yunusbayev, B., Hudjashov, G., DeGiorgio, M., Loogväli, E. L., Eichstaedt, C., Eelmets, M., Chaubey, G., Tambets, K., Litvinov, S., Mormina, M., Xue, Y., Ayub, Q., Zoraqi, G., Korneliussen, T. S., Akhatova, F., Lachance, J., Tishkoff, S., Momynaliev, K., Ricaut, F. X., Kusuma, P., Razafindrazaka, H., Pierron, D., Cox, M. P., Sultana, G. N. N., Willerslev, R., Muller, C., Westaway, M., Lambert, D., Skaro, V., Kovačević, L., Turdikulova, S., Dalimova, D., Khusainova, R., Trofimova, N., Akhmetova, V., Khidiyatova, I., Lichman, D. V., Isakova, J., Pocheshkhova, E., Sabitov, Z., Barashkov, N. A., Nymadawa, P., Mihailov, E., Seng, J. W. T., Evseeva, I., Migliano, A. B., Abdullah, S., Andriadze, G., Primorac, D., Atramentova, L., Utevska, O., Yepiskoposyan, L., Marjanović, D., Kushniarevich, A., Behar, D. M., Gilissen, C., Vissers, L., Veltman, J. A., Balanovska, E., Derenko, M., Malyarchuk, B., Metspalu, A., Fedorova, S., Eriksson, A., Manica, A., Mendez, F. L., Karafet, T. M., Veeramah, K. R., Bradman, N., Hammer, M. F., Osipova, L. P., Balanovsky, O., Khusnutdinova, E. K., Johnsen, K., Remm, M., Thomas, M. G., Tyler-Smith, C., Underhill, P. A., Willerslev, E., Nielsen, R., Metspalu, M., Villems, R. and Kivisild, T. (2015), 'A recent bottleneck of Y chromosome diversity coincides with a global change in culture', *Genome Research* **25**(4), 459–466.

Kashuba, N., Kırdök, E., Damlien, H., Manninen, M. A., Nordqvist, B., Persson, P. and Götherström, A. (2019), 'Ancient DNA from mastics solidifies connection between material culture and genetics of mesolithic hunter–gatherers in Scandinavia', *Communications Biology* **2**(1), 185.

Kayser, M. (2015), 'Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes', *Forensic Science International: Genetics* **18**, 33–48.

Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., Stade, B., Franke, A., Mayer, J., Spangler, J., McLaughlin, S., Shah, M., Lee, C., Harkins, T. T., Sartori, A., Moreno-Estrada, A., Henn, B., Sikora, M., Semino, O., Chiaroni, J., Rootsi, S., Myres, N. M., Cabrera, V. M., Underhill, P. A., Bustamante, C. D., Vigl, E. E., Samadelli, M., Cipollini, G., Haas, J., Katus, H., O'Connor, B. D., Carlson, M. R., Meder, B., Blin, N., Meese, E., Pusch, C. M. and Zink, A. (2012), 'New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing', *Nature Communications* **3**, 698.

Keller, M., Spyrou, M. A., Scheib, C. L., Neumann, G. U., Kröpelin, A., Haas-Gebhard, B., Päffgen, B., Haberstroh, J., Lacomba, A. R. I., Raynaud, C., Cessford, C., Durand, R., Stadler, P., Nägele, K., Bates, J. S., Trautmann, B., Inskip, S. A., Peters, J., Robb, J. E., Kivisild, T., Castex, D., McCormick, M., Bos, K. I., Harbeck, M., Herbig, A. and Krause, J. (2019), 'Ancient Yersinia pestis genomes from across Western Europe reveal early diversification during the First Pandemic (541–750)', *Proceedings of the National Academy of Sciences of the United States of America* **116**(25), 12363–12372.

Kibblewhite, M., Tóth, G. and Hermann, T. (2015), 'Predicting the preservation of cultural artefacts and buried materials in soil', *Science of the Total Environment* **529**, 249–263.

Kimura, M. (1980), 'A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences', *J. Mol. Evol* **16**(1330), 111–120.

Kimura, M. (1981), 'Estimation of evolutionary distances between homologous nucleotide sequences', *Proceedings of the National Academy of Sciences of the United States of America* **78**(1), 454–458.

Klein, R. G. (2009), 'Darwin and the recent African origin of modern humans', **106**(38), 16007–16009.

Kılınç, G. M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A. M., Bıçakçı, E., Baird, D., Dönertaş, H. M., Ghalichi, A., Yaka, R., Koptekin, D., Açan, S. C., Parvizi, P., Krzewińska, M., Daskalaki, E. A., Yüncü, E., Dağtaş, N. D., Fairbairn, A., Pearson, J., Mustafaoğlu, G., Erdal, Y. S., Çakan, Y. G., Togan, n., Somel, M., Storå, J., Jakobsson, M. and Götherström, A. (2016), 'The Demographic Development of the First Farmers in Anatolia', *Current Biology* **26**, 2659–2666.

Knipper, C., Mittnik, A., Massy, K., Kociumaka, C., Kucukkalipci, I., Maus, M., Wittenborn, F., Metz, S. E., Staskiewicz, A., Krause, J. and Stockhammer, P. W. (2017), 'Female exogamy and gene pool diversification at the transition from the Final Neolithic to the Early Bronze Age in central Europe', *Proceedings of the National Academy of Sciences* **114**(38), 10083–10088.

Koch, J. and Cunliffe, B. (2013), *Celtic from the West 2: rethinking the Bronze Age and the arrival of Indo-European in Atlantic Europe*, Oxbow Books.

Kristiansen, K., Allentoft, M. E., Frei, K. M., Iversen, R., Johannsen, N. N., Kroonen, G., Pospieszny, Ł., Price, T. D., Rasmussen, S., Sjögren, K.-G., Sikora, M. and Willerslev, E. (2017), 'Re-theorising mobility and the formation of culture and language among the Corded Ware Culture in Europe', *Antiquity* **91**(356), 334–347.

Kuhn, J. M. M., Jakobsson, M. and Günther, T. (2018), 'Estimating genetic kin relationships in prehistoric populations', *PLoS ONE* **13**(4), e0195491.

Lacan, M., Keyser, C., Ricaut, F.-X., Brucato, N., Duranthon, F., Guilaine, J., Crubezy, E. and Ludes, B. (2011), 'Ancient DNA reveals male diffusion through the Neolithic Mediterranean route', *Proceedings of the National Academy of Sciences* **108**(24), 9788–9791.

Lalueza-Fox, C. (2013), *Palabras en el tiempo*, Editorial Crítica.

Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., De La Rasilla, M., Fortea, J., Rosas, A., Stoneking, M., Schöneberg, T., Bertranpetit, J. and Hofreiter, M. (2007), 'A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals', *Science* **318**(5855), 1453–1455.

Larruga, J. M., Díez, F., Pinto, F. M., Flores, C. and González, A. M. (2001), 'Mitochondrial DNA characterisation of European isolates: The Maragatos from Spain', *European Journal of Human Genetics* **9**(9), 708–716.

Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwängler, A., Peltzer, A., Posth, C., Vasilakis, A., McGeorge, P. J., Konsolaki-Yannopoulou, E., Korres, G., Martlew, H., Michalodimitrakis, M., Özsait, M., Özsait, N., Papathanasiou, A., Richards, M., Roodenberg, S. A., Tzedakis, Y., Arnott, R., Fernandes, D. M., Hughey, J. R., Lotakis, D. M., Navas, P. A., Maniatis, Y., Stamatoyannopoulos, J. A., Stewardson, K., Stockhammer, P., Pinhasi, R., Reich, D., Krause, J. and Stamatoyannopoulos, G. (2017), 'Genetic origins of the Minoans and Mycenaeans', *Nature* **548**(7666), 214–218.

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., Connell, S., Stewardson, K., Harney, E., Fu, Q., Gonzalez-Fortes, G., Jones, E. R., Roodenberg, S. A., Lengyel, G., Bocquentin, F., Gasparian, B., Monge, J. M., Gregg, M., Eshed, V., Mizrahi, A. S., Meiklejohn, C., Gerritsen, F., Bejenaru, L., Bliher, M., Campbell, A., Cavalleri, G., Comas, D., Froguel, P., Gilbert, E., Kerr, S. M., Kovacs, P., Krause, J., McGettigan, D., Merrigan, M., Merriwether, D. A., O'Reilly, S., Richards, M. B., Semino, O., Shamoon-Pour, M., Stefanescu, G., Stumvoll, M., Tinjes, A., Torroni, A., Wilson, J. F., Yengo, L., Hovhannisyan, N. A., Patterson, N., Pinhasi, R. and Reich, D. (2016), 'Genomic insights into the origin of farming in the ancient Near East', *Nature* **536**, 419–424.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P. H., Schraiber, J. G., Castellano, S., Lipson, M., Berger, B., Economou, C., Bollongino, R., Fu, Q., Bos, K. I., Nordenfelt, S., Li, H., De Filippo, C., Prüfer, K., Sawyer, S., Posth, C., Haak, W., Hallgren, F., Fornander, E., Rohland, N., Delsate, D., Francken, M., Guinet, J. M., Wahl, J., Ayodo, G., Babiker, H. A., Bailliet, G., Balanovska, E., Balanovsky, O., Barrantes, R., Bedoya, G., Ben-Ami, H., Bene, J., Berrada, F., Bravi, C. M., Brisighelli, F., Busby, G. B., Cali, F., Churnosov, M., Cole, D. E., Corach, D., Damba, L., Van Driem, G., Dryomov, S., Dugoujon, J. M., Fedorova, S. A., Gallego Romero, I., Gubina, M., Hammer, M., Henn, B. M., Hervig, T., Hodoglugil, U., Jha, A. R., Karachanak-Yankova, S., Khusainova, R., Khusnutdinova, E., Kittles, R., Kivisild, T., Klitz, W., Kučinskas, V., Kushniarevich, A., Laredj, L., Litvinov, S., Loukidis, T., Mahley, R. W., Melegh, B., Metspalu, E., Molina, J., Mountain, J., Näkkäläjärvi, K., Nesheva, D., Nyambo, T., Osipova, L., Parik, J., Platonov, F., Posukh, O., Romano, V., Rothhammer, F., Rudan, I., Ruizbakiev, R., Sahakyan, H., Sajantila, A., Salas, A., Starikovskaya, E. B., Tarekegn, A., Toncheva, D., Turdikulova, S., Uktveryte, I., Utevska, O., Vasquez, R., Villena, M., Voevoda, M., Winkler, C. A., Yepiskoposyan, L., Zalloua, P., Zemunik, T., Cooper, A., Capelli, C., Thomas, M. G., Ruiz-Linares, A., Tishkoff, S. A., Singh, L., Thangaraj, K., Villems, R., Comas, D., Sukernik, R., Metspalu, M., Meyer, M., Eichler, E. E., Burger, J., Slatkin, M., Pääbo, S., Kelso, J., Reich, D. and Krause, J. (2014), 'Ancient human genomes suggest three ancestral populations for present-day Europeans', *Nature* **513**, 409–413.

Leonardi, M., Librado, P., Der Sarkissian, C., Schubert, M., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A., Gamba, C., Willerslev, E. and Orlando, L. (2017), Evolutionary patterns and processes: Lessons from ancient DNA, *in* 'Systematic Biology', Vol. 66, pp. e1–e29.

Lewis, A. R. (1968), 'The Crusader Kingdom of Valencia. Reconstruction on a Thirteenth-Century Frontier. Burns, R.I', *The Hispanic American Historical Review* **48**(4), 685–687.

Li, W., Cerise, J. E., Yang, Y. and Han, H. (2017), 'Application of t-SNE to human genetic data', *Journal of Bioinformatics and Computational Biology* **15**, 1750017.

Li, Y., Zhao, W., Li, D., Tao, X., Xiong, Z., Liu, J., Zhang, W., Ji, A., Tang, K., Liu, F. and Li, C. (2019), 'EDAR, LYPLAL1, PRDM16, PAX3, DKK1, TNFSF12, CACNA2D3, and SUPT3H gene variants influence facial morphology in a Eurasian population', *Human Genetics* **138**(6).

Liebert, A., López, S., Jones, B. L., Montalva, N., Gerbault, P., Lau, W., Thomas, M. G., Bradman, N., Maniatis, N. and Swallow, D. M. (2017), 'World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection', *Human Genetics* **138**(6), 681–689.

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

Lindahl, T. (1996), 'The Croonian Lecture, 1996: Endogenous damage to DNA', **351**(1347), 1529–1538.

Linderholm, A., Kılınç, G. M., Szczepanek, A., Włodarczak, P., Jarosz, P., Belka, Z., Dopieralska, J., Werens, K., Górski, J., Mazurek, M., Hozer, M., Rybicka, M., Ostrowski, M., Bagińska, J., Koman, W., Rodríguez-Varela, R., Storå, J., Götherström, A. and Krzewińska, M. (2020), 'Corded Ware cultural complexity uncovered using genomic and isotopic analysis from south-eastern Poland', *Scientific Reports* **10**(1), 6885.

Lippert, C., Sabatini, R., Maher, M. C., Kang, E. Y., Lee, S., Arikan, O., Harley, A., Bernal, A., Garst, P., Lavrenko, V., Yocum, K., Wong, T., Zhu, M., Yang, W. Y., Chang, C., Lu, T., Lee, C. W., Hicks, B., Ramakrishnan, S., Tang, H., Xie, C., Piper, J., Brewerton, S., Turpaz, Y., Telenti, A., Roby, R. K., Och, F. J. and Venter, J. C. (2017), 'Identification of individuals by trait prediction using whole-genome sequencing data', *Proceedings of the National Academy of Sciences of the United States of America* **114**(38), 10166–10171.

Lippold, S., Xu, H., Ko, A., Li, M., Renaud, G., Butthof, A., Schröder, R. and Stoneking, M. (2014), 'Human paternal and maternal demographic histories: Insights from high-resolution Y chromosome and mtDNA sequences', *Investigative Genetics* **13**, 1–17.

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Harney, E., Nordenfelt, S., Llamas, B., Gusztáv, B. M., Köhler, K., Oross, K., Bondár, M., Marton, T., Osztás, A., Jakucs, J., Paluch, T., Horváth, F., Csengeri, P., Koós, J., Sebok, K., Anders, A., Raczky, P., Regenye, J., Barna, J. P., Fábián, S., Serlegi, G., Toldi, Z., Nagy, E. G., Dani, J., Molnár, E., Pálfi, G., Márk, L., Melegh, B., Bánfai, Z., Domboróczki, L., Fernández-Eraso, J., Mujika-Alustiza, J. A., Fernández, C. A., Echevarría, J. J., Bollongino, R., Orschiedt, J., Schierhold, K., Meller, H., Cooper, A., Burger, J., Bánffy, E., Alt, K. W., Lalueza-Fox, C., Haak, W. and Reich, D. (2017), 'Parallel palaeogenomic transects reveal complex genetic history of early European farmers', *Nature* **551**(7680), 368–372.

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., Wollstein, A., Lao, O., de Bruijne, M., Ikram, M. A., van der Lugt, A., Rivadeneira, F., Uitterlinden, A. G.,

Hofman, A., Niessen, W. J., Homuth, G., de Zubicaray, G., McMahon, K. L., Thompson, P. M., Daboul, A., Puls, R., Hegenscheid, K., Bevan, L., Pausova, Z., Medland, S. E., Montgomery, G. W., Wright, M. J., Wicking, C., Boehringer, S., Spector, T. D., Paus, T., Martin, N. G., Biffar, R. and Kayser, M. (2012), 'A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans', *PLoS Genetics* **8**(9), e1002932.

Llamas, B., Willerslev, E. and Orlando, L. (2017), 'Human evolution: A tale from ancient genomes', *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**(1713).

Lucotte, G. (2015), 'The Major Y-Chromosome Haplogroup R1b-M269 in West-Europe, Subdivided by the Three SNPs S21/U106, S145/L21 and S28/U152, Shows a Clear Pattern of Geographic Differentiation', *Advances in Anthropology* **5**(1), 22–30.

Lugli, G. A., Milani, C., Mancabelli, L., Turroni, F., Ferrario, C., Duranti, S., van Sinderen, D. and Ventura, M. (2017), 'Ancient bacteria of the ötzi's microbiome: A genomic tale from the Copper Age', *Microbiome* **5**(1), 1–18.

Maca-Meyer, N., González, A. M., Pestano, J., Flores, C., Larruga, J. M. and Cabrera, V. M. (2003), 'Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography', *BMC Genetics* **4**(15), 1–11.

Maca-Meyer, N., Sánchez-Velasco, P., Flores, C., Larruga, J. M., González, A. M., Oterino, A. and Leyva-Cobián, F. (2003), 'Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain)', *Annals of Human Genetics* **67**(4), 329–339.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-Tamir, B., Sykes, B. and Torroni, A. (1999), 'The emerging tree of west Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs', *American Journal of Human Genetics* **64**(1), 232–249.

Machancoses López, M. (2015), *Topografía urbana de la Valentia romana altoimperial: ciudad y suburbio. PhD thesis*, Universitat de València.

Machancoses López, M. (2016), 'Actualización de la topografía de las necrópolis de Valentia siglos I-III d.C.', *Anales de arqueología cordobesa* **27**, 183–214.

MacHugh, D., Edwards, C., Bailey, J., Bancroft, D. and Bradley, D. (2000), 'The extraction and analysis of ancient DNA from bone and teeth: a survey of current methodologies', *Ancient Biomolecules* **3**, 81–102.

Mackintosh-Smith, T. (2019), *Arabs: A 3,000-Year History of Peoples, Tribes and Empires*, Yale University Press.

Malaspinas, A. S., Tange, O., Moreno-Mayar, J. V., Rasmussen, M., DeGiorgio, M., Wang, Y., Valdiosera, C. E., Politis, G., Willerslev, E. and Nielsen, R. (2014), 'bammds: a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS)', *Bioinformatics (Oxford, England)* **30**(20), 2962–2964.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q., Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., Spence, J. P., Song, Y. S., Poletti, G., Balloux, F., Van Driem, G., De Knijff, P., Romero, I. G., Jha, A. R., Behar, D. M., Bravi, C. M., Capelli, C., Hervig, T., Moreno-Estrada, A., Posukh, O. L., Balanovska, E., Balanovsky, O., Karachanak-Yankova, S., Sahakyan, H., Toncheva, D., Yepiskoposyan, L., Tyler-Smith, C., Xue, Y., Abdullah, M. S., Ruiz-Linares, A., Beall, C. M., Di Rienzo, A., Jeong, C., Starikovskaya, E. B., Metspalu, E., Parik, J., Villems, R., Henn, B. M., Hodoglugil, U., Mahley, R., Sajantila, A.,

Stamatoyannopoulos, G., Wee, J. T., Khusainova, R., Khusnutdinova, E., Litvinov, S., Ayodo, G., Comas, D., Hammer, M. F., Kivisild, T., Klitz, W., Winkler, C. A., Labuda, D., Bamshad, M., Jorde, L. B., Tishkoff, S. A., Watkins, W. S., Metspalu, M., Dryomov, S., Sukernik, R., Singh, L., Thangaraj, K., Paäbo, S., Kelso, J., Patterson, N. and Reich, D. (2016), 'The Simons Genome Diversity Project: 300 genomes from 142 diverse populations', *Nature* **538**(7624), 201–206.

Mallory, J. P. (1991), *In Search of the Indo-Europeans; Language, Archaeology and Myth*, Thames and Hudson Ltd.

Manica, A., Amos, W., Balloux, F. and Hanihara, T. (2007), 'The effect of ancient population bottlenecks on human phenotypic variation', *Nature* **538**(7624), 201–206.

Maroñas, O., Phillips, C., Söchtig, J., Gomez-Tato, A., Cruz, R., Alvarez-Dios, J., De Cal, M. C., Ruiz, Y., Fondevila, M., Carracedo, Á. and Lareu, M. V. (2014), 'Development of a forensic skin colour predictive test', *Forensic Science International: Genetics* **13**, 34–44.

Maroñas, O., Söchtig, J., Ruiz, Y., Phillips, C., Carracedo and Lareu, M. V. (2015), 'The genetics of skin, hair, and eye color variation and its relevance to forensic pigmentation predictive tests', **27**(1), 13–40.

Martin, A. R., Lin, M., Granka, J. M., Myrick, J. W., Liu, X., Sockell, A., Atkinson, E. G., Werely, C. J., Möller, M., Sandhu, M. S., Kingsley, D. M., Hoal, E. G., Liu, X., Daly, M. J., Feldman, M. W., Gignoux, C. R., Bustamante, C. D. and Henn, B. M. (2017), 'An Unexpectedly Complex Architecture for Skin Pigmentation in Africans', *Cell* **171**(6), 1340–1353.

Martínez Cortizas, A., López-Merino, L., Bindler, R., Mighall, T. and Kylander, M. E. (2016), 'Early atmospheric metal pollution provides evidence for Chalcolithic/Bronze Age mining and metallurgy in Southwestern Europe', *Science of the Total Environment* **545-546**, 398–406.

Martínez-Cruz, B., Harmant, C., Platt, D. E., Haak, W., Manry, J., Ramos-Luis, E., Soria-Hernanz, D. F., Bauduer, F., Salaberria, J., Oyharçabal, B., Quintana-Murci, L. and Comas, D. (2012), 'Evidence of pre-roman tribal genetic structure in basques from uniparentally inherited markers', *Molecular Biology and Evolution* **29**(9), 2211–2222.

Martínez Valle, A. (1998), 'Nuevos hallazgos de inscripciones romanas en la provincia de Valencia', *SAGVNTVM: Papeles del Laboratorio de Arqueología de Valencia* **31**(31), 263–268.

Martiniano, R., Caffell, A., Holst, M., Hunter-Mann, K., Montgomery, J., Müldner, G., McLaughlin, R. L., Teasdale, M. D., Van Rheenen, W., Veldink, J. H., Van Den Berg, L. H., Hardiman, O., Carroll, M., Roskams, S., Oxley, J., Morgan, C., Thomas, M. G., Barnes, I., McDonnell, C., Collins, M. J. and Bradley, D. G. (2016), 'Genomic signals of migration and continuity in Britain before the Anglo-Saxons', *Nature Communications* **19**, 10326.

Martiniano, R., Cassidy, L. M., Ó'Maoldúin, R., McLaughlin, R., Silva, N. M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M. J., Serra, M., Burger, J., Parreira, R., Moran, E., Valera, A. C., Porfirio, E., Boaventura, R., Silva, A. M. and Bradley, D. G. (2017), 'The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods', *PLoS Genetics* **13**, e1006852.

Martínez Valle, R. and Guérin Fockedey, P. (1987), 'Inhumaciones infantiles en poblados ibéricos del área valenciana)', **21**, 231–266.

Marzal Palacios, F. J. (2002), 'Una presencia constante: los esclavos sarracenos en Valencia (siglos XIII-XVI)', *Sharq Al-Andalus* **16-17**.

Marzal Palacios, F. J. (2006), *La esclavitud en Valencia durante la Baja Edad Media (1375-1425). PhD thesis*, Universitat de Valencia.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., Fernandes, D., Ferry, M., Gamarra, B., Fortes, G. G., Haak, W., Harney, E., Jones, E., Keating, D., Krause-Kyora, B., Kucukkalipci, I., Michel, M., Mittnik, A., Nägele, K., Novak, M., Oppenheimer, J., Patterson, N., Pfrengle, S., Sirak, K., Stewardson, K., Vai, S., Alexandrov, S., Alt, K. W., Andreescu, R., Antonović, D., Ash, A., Atanassova, N., Bacvarov, K., Gusztáv, M. B., Bocherens, H., Bolus, M., Boroneanţ, A., Boyadzhiev, Y., Budnik, A., Burmaz, J., Chohadzhiev, S., Conard, N. J., Cottiaux, R., Čuka, M., Cupillard, C., Drucker, D. G., Elenski, N., Francken, M., Galabova, B., Ganetsovski, G., Gély, B., Hajdu, T., Handzhyiska, V., Harvati, K., Higham, T., Iliev, S., Janković, I., Karavanić, I., Kennett, D. J., Komšo, D., Kozak, A., Labuda, D., Lari, M., Lazar, C., Leppek, M., Leshtakov, K., Vetro, D. L., Los, D., Lozanov, I., Malina, M., Martini, F., McSweeney, K., Meller, H., Mentušić, M., Mirea, P., Moiseyev, V., Petrova, V., Douglas Price, T., Simalcsik, A., Sineo, L., Šlaus, M., Slavchev, V., Stanev, P., Starović, A., Szeniczey, T., Talamo, S., Teschler-Nicola, M., Thevenet, C., Valchev, I., Valentin, F., Vasilyev, S., Veljanovska, F., Venelinova, S., Veselovskaya, E., Viola, B., Virag, C., Zaninović, J., Zaüner, S., Stockhammer, P. W., Catalano, G., Krauß, R., Caramelli, D., Zariņa, G., Gaydarska, B., Lillie, M., Nikitin, A. G., Potekhina, I., Papathanasiou, A., Borić, D., Bonsall, C., Krause, J., Pinhasi, R. and Reich, D. (2018), 'The genomic history of southeastern Europe', *Nature* **555**, 197–203.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., De Castro, J. M. B., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M. A., Roodenberg, J., Vergès, J. M., Krause, J., Cooper, A., Alt, K. W., Brown, D., Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R. and Reich, D. (2015), 'Genome-wide patterns of selection in 230 ancient Eurasians', *Nature* **528**(7583), 499–503.

Mathieson, S. and Mathieson, I. (2018), 'FADS1 and the timing of human adaptation to agriculture', *Molecular Biology and Evolution* **35**(12), 2957–2970.

Matisoo-Smith, E., Gosling, A. L., Platt, D., Kardailsky, O., Prost, S., Cameron-Christie, S., Collins, C. J., Boocock, J., Kurumilian, Y., Guirguis, M., Pla Orquín, R., Khalil, W., Genz, H., Abou Diwan, G., Nassar, J. and Zalloua, P. (2018), 'Ancient mitogenomes of Phoenicians from Sardinia and Lebanon: A story of settlement, integration, and female mobility', *PLoS ONE* **13**(1), e0190169.

McConnell, J. R., Wilson, A. I., Stohl, A., Arienzo, M. M., Chellman, N. J., Eckhardt, S., Thompson, E. M., Pollard, A. M. and Steffensen, J. P. (2018), 'Lead pollution recorded in Greenland ice indicates European emissions tracked plagues, wars, and imperial expansion during antiquity.', *Proceedings of the National Academy of Sciences of the United States of America* **115**(22), 5726–5731.

McEvoy, B., Beleza, S. and Shriver, M. D. (2006), 'The genetic architecture of normal variation in human pigmentation: An evolutionary perspective and model', *Human Molecular Genetics* **15**(2), R176–181.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M. A. (2010), 'The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research* **20**(9), 1297–1303.

Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., De Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., Briggs, A. W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M. F., Shunkov, M. V., Derevianko, A. P., Patterson, N., Andrés, A. M., Eichler, E. E., Slatkin, M., Reich, D., Kelso, J. and Pääbo, S. (2012), 'A high-coverage genome sequence from an archaic Denisovan individual', *Science* **338**(6104), 222–226.

Mittnik, A., Wang, C. C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., Furtwängler, A., Andrades Valtueña, A., Feldman, M., Economou, C., Oinonen, M., Vasks, A., Balanovska, E., Reich, D., Jankauskas, R., Haak, W., Schiffels, S. and Krause, J. (2018), 'The genetic prehistory of the Baltic Sea region', *Nature Communications* **9**, 442.

Mullis, K. B. and Faloona, F. A. (1987), 'Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction', *Methods in Enzymology* **115**, 335–350.

Murillo-Barroso, M. and Montero-Ruiz, I. (2012), 'Copper Ornaments in the Iberian Chalcolithic: Technology versus Social Demand', *Journal of Mediterranean Archaeology* **25**(1), 53–73.

Myres, N. M., Rootsi, S., Lin, A. A., Järve, M., King, R. J., Kutuev, I., Cabrera, V. M., Khusnutdinova, E. K., Pshenichnov, A., Yunusbayev, B., Balanovsky, O., Balanovska, E., Rudan, P., Baldovic, M., Herrera, R. J., Chiaroni, J., Di Cristofaro, J., Villems, R., Kivisild, T. and Underhill, P. A. (2011), 'A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe', *European Journal of Human Genetics* **19**(1), 95–101.

Narasimhan, V., Patterson, N., Moorjani, P., Lazaridis, I., Lipson, M., Mallick, S., Rohland, N., Bernardos, R., Kim, A., Nakatsuka, N., Olalde, I., Coppa, A., Mallory, J., Moiseyev, V., Monge, J., Olivieri, L., Adamski, N., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., Culleton, B., Ferry, M., Fernandes, D., Gamarra, B., Gaudio, D., Hajdinjak, M., Harney, É., Harper, T., Keating, D., Lawson, A. M., Michel, M., Novak, M., Oppenheimer, J., Rai, N., Sirak, K., Slon, V., Stewardson, K., Zhang, Z., Akhatov, G., Bagashev, A., Baitanayev, B., Bonora, G. L., Chikisheva, T., Derevianko, A., Dmitry, E., Douka, K., Dubova, N., Epimakhov, A., Freilich, S., Fuller, D., Goryachev, A., Gromov, A., Hanks, B., Judd, M., Kazizov, E., Khokhlov, A., Kitov, E., Kupriyanova, E., Kuznetsov, P., Luiselli, D., Maksudov, F., Meiklejohn, C., Merrett, D., Micheli, R., Mochalov, O., Muhammed, Z., Mustafokulov, S., Nayak, A., Petrovna, R., Pettener, D., Potts, R., Razhev, D., Sarno, S., Sikhymbaeva, K., Slepchenko, S., Stepanova, N., Svyatko, S., Vasilyev, S., Vidale, M., Voyakin, D., Yermolayeva, A., Zubova, A., Shinde, V., Lalueza-Fox, C., Meyer, M., Anthony, D., Boivin, N., Thangaraj, K., Kennett, D., Frachetti, M., Pinhasi, R. and Reich, D. (2019), 'The Genomic Formation of South and Central Asia', *Science* **365**(6457), eaa7487.

Negrete, J. (2009), *La gran aventura de los griegos*, La Esfera de los Libros.

Núñez, C., Baeta, M., Cardoso, S., Palencia-Madrid, L., García-Romero, N., Llanos, A. and M. de Pancorbo, M. (2016), 'Mitochondrial dna reveals the trace of the ancient settlers of a violently devastated late bronze and iron ages village', *PLOS ONE* **11**(5), 1–16.

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S. and Willerslev, E. (2017), 'Tracing the peopling of the world through genomics', **541**(7637), 302–310.

Nielsen, R., Paul, J. S., Albrechtsen, A. and Song, Y. S. (2011), 'Genotype and SNP calling from next-generation sequencing data', **12**(6), 443–451.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., Indap, A., King, K. S., Bergmann, S., Nelson, M. R., Stephens, M. and Bustamante, C. D. (2008), 'Genes mirror geography within Europe', *Nature* **456**(7218), 98–101.

Olalde, I., Allentoft, M. E., Sánchez-Quinto, F., Santpere, G., Chiang, C. W. K., DeGiorgio, M., Prado-Martinez, J., Rodríguez, J. A., Rasmussen, S., Quilez, J., Ramírez, O., Marigorta, U. M., Fernández-Callejo, M., Prada, M. E., Encinas, J. M. V., Nielsen, R., Netea, M. G., Novembre, J., Sturm, R. A., Sabeti, P., Marquès-Bonet, T., Navarro, A., Willerslev, E. and Lalueza-Fox, C. (2014), 'Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European', *Nature* **507**, 225–228.

Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., Altena, E., Lipson, M., Lazaridis, I., Harper, T. K., Patterson, N., Broomandkhoshbacht, N., Diekmann, Y., Faltyskova, Z., Fernandes, D., Ferry, M., Harney, E., De Knijff, P., Michel, M., Oppenheimer, J., Stewardson, K., Barclay, A., Alt, K. W., Liesau, C., Rios, P., Blasco, C., Miguel, J. V., Garcia, R. M., Fernandez, A. A., Banffy, E., Bernabo-Brea, M., Billoin, D., Bonsall, C., Bonsall, L., Allen, T., Buster, L., Carver, S., Navarro, L. C., Craig, O. E., Cook, G. T., Cunliffe, B., Denaire, A., Dinwiddy, K. E., Dodwell, N., Ernee, M., Evans, C., Kucharik, M., Farre, J. F., Fowler, C., Gazenbeek, M., Pena, R. G., Haber-Uriarte, M., Haduch, E., Hey, G., Jowett, N., Knowles, T., Massy, K., Pfrengle, S., Lefranc, P., Lemercier, O., Lefebvre, A., Martinez, C. H., Olmo, V. G., Ramirez, A. B., Maurandi, J. L., Majo, T., McKinley, J. I., McSweeney, K., Mende, B. G., Mod, A., Kulcsar, G., Kiss, V., Czene, A., Patay, R., Endrodi, A., Kohler, K., Hajdu, T., Szeniczey, T., Dani, J., Bernert, Z., Hoole, M., Cheronet, O., Keating, D., Veleminsky, P., Dobe, M., Candilio, F., Brown, F., Fernandez, R. F., Herrero-Corral, A. M., Tusa, S., Carnieri, E., Lentini, L., Valenti, A., Zanini, A., Waddington, C., Delibes, G., Guerra-Doce, E., Neil, B., Brittain, M., Luke, M., Mortimer, R., Desideri, J., Besse, M., Brucken, G., Furmanek, M., Hauszko, A., Mackiewicz, M., Rapinski, A., Leach, S., Soriano, I., Lillios, K. T., Cardoso, J. L., Pearson, M. P., Wodarczak, P., Price, T. D., Prieto, P., Rey, P. J., Risch, R., Guerra, M. A., Schmitt, A., Serralongue, J., Silva, A. M., Smrcka, V., Vergnaud, L., Zilhao, J., Caramelli, D., Higham, T., Thomas, M. G., Kennett, D. J., Fokkens, H., Heyd, V., Sheridan, A., Sjogren, K. G., Stockhammer, P. W., Krause, J., Pinhasi, R., Haak, W., Barnes, I., Lalueza-Fox, C. and Reich, D. (2018), 'The Beaker phenomenon and the genomic transformation of northwest Europe', *Nature* **555**, 190–196.

Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C. J., Gandini, F., Pala, M., Soares, P., Ferrando-Bernal, M., Adamski, N., Broomandkhoshbacht, N., Cheronet, O., Culleton, B. J., Fernandes, D., Lawson, A. M., Mah, M., Oppenheimer, J., Stewardson, K., Zhang, Z., Arenas, J. M. J., Moyano, I. J. T., Salazar-García, D. C., Castanyer, P., Santos, M., Tremoleda, J., Lozano, M., Borja, P. G., Fernández-Eraso, J., Mujika-Alustiza, J. A., Barroso, C., Bermúdez, F. J., Mínguez, E. V., Burch, J., Coromina, N., Vivó, D., Cebrià, A., Fullola, J. M., García-Puchol, O., Morales, J. I., Xavier Oms, F., Majó, T., Vergès, J. M., Díaz-Carvajal, A., Ollich-Castanyer, I., Javier López-Cachero, F., Silva, A. M., Alonso-Fernández, C., De Castro, G. D., Echevarría, J. J., Moreno-Márquez, A., Berlanga, G. P., Ramos-García, P., Ramos-Muñoz, J., Vila, E. V., Arzo, G. A., Arroyo, Á. E., Lillios, K. T., Mack, J., Velasco-Vázquez, J., Waterman, A., De Lugo Enrich, L. B., Sánchez, M. B., Agustí, B., Codina, F., De Prado, G., Estalrrich, A., Flores, Á. F., Finlayson, C., Finlayson, G., Finlayson, S., Giles-Guzmán, F., Rosas, A., González, V. B., Atiénzar, G. G., Hernández Pérez, M. S., Llanos, A., Marco, Y. C., Beneyto, I. C., López-Serrano, D., Tormo, M. S., Valera, A. C., Blasco, C., Liesau, C., Ríos, P., Daura, J., De Pedro Michó, M. J., Diez-Castillo, A. A., Fernández, R. F., Farré, J. F., Garrido-Pena, R., Gonçalves, V. S., Guerra-Doce, E., Herrero-Corral, A. M., Juan-Cabanilles, J., López-Reyes, D., McClure, S. B., Pérez, M. M., Foix, A. O., Borràs, M. S., Sousa, A. C., Encinas, J. M. V., Kennett, D. J., Richards, M. B., Alt, K. W., Haak, W., Pinhasi, R., Lalueza-Fox, C. and Reich, D. (2019), 'The genomic history of the Iberian Peninsula over the past 8000 years', *Science* **363**(6432), 1230–1234.

Olalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobón, I., Ramirez, O., Civit, S., Borja, P. G., Salazar-García, D. C., Talamo, S., Fullola, J. M., Oms, F. X., Pedro, M., Martínez,

P., Sanz, M., Daura, J., Zilhão, J., Marquès-Bonet, T., Gilbert, M. T. P. and Lalueza-Fox, C. (2015), 'A common genetic origin for early farmers from mediterranean cardial and central european LBK cultures', *Molecular Biology and Evolution* **32**(12), 3132–3142.

Oliver, A, A.-P. E. and Fernández, E. (2008), 'Secuencias genéticas matrilineales de los restos óseos humanos de la Costa Lloguera', *Verdolay* **11**, 37–48.

Olivieri, A., Sidore, C., Achilli, A., Angius, A., Posth, C., Furtwängler, A., Brandini, S., Capodiferro, M. R., Gandini, F., Zoledziewska, M., Pitzalis, M., Maschio, A., Busonero, F., Lai, L., Skeates, R., Gradoli, M. G., Beckett, J., Marongiu, M., Mazzarello, V., Marongiu, P., Rubino, S., Rito, T., Macaulay, V., Semino, O., Pala, M., Abecasis, G. R., Schlessinger, D., Conde-Sousa, E., Soares, P., Richards, M. B., Cucca, F. and Torroni, A. (2017), 'Mitogenome Diversity in Sardinians: A Genetic Window onto an Island's Past', *Molecular Biology and Evolution* **34**, 1230–1239.

Olmos, R. and Marcos, M. L. (2000), 'El vaso del "Ciclo de la Vida de Valencia": una reflexión sobre la imagen metamórfica en época iberohelenística', *Archivo español de arqueología* **73**(181), 59–86.

O'Sullivan, N. J., Teasdale, M. D., Mattiangeli, V., Maixner, F., Pinhasi, R., Bradley, D. G. and Zink, A. (2016), 'A whole mitochondria analysis of the Tyrolean Iceman's leather provides insights into the animal sources of Copper Age clothing', *Scientific Reports* **6**, 31279.

Otto-Dorn, K. (1957), 'Grabung Im Ummayadischen Rusafah', *Ars Orientalis* **2**, 119–133.

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

Pääbo, S. (1993), 'Ancient DNA', *Scientific American* **269**(5), 86–92.

Pääbo, S., Gifford, J. A. and Wilson, A. C. (1988), 'Mitochondrial DNA sequences from a 7000-year old brain', *Nucleic Acids Research* **16**(20), 9775–9787.

Pala, M., Chaubey, G., Soares, P. and Richards, M. B. (2014), The Archaeogenetics of European Ancestry, *in* 'Encyclopedia of Life Sciences', Wiley.

Pardiñas, A. F., Roca, A., García-Vazquez, E. and López, B. (2012), 'Assessing the Genetic Influence of Ancient Sociopolitical Structure: Micro-differentiation Patterns in the Population of Asturias (Northern Spain)', *PLoS ONE* **7**(11), e50206.

Parra, E. J., Kittles, R. A. and Shriver, M. D. (2004), 'Implications of correlations between skin color and genetic ancestry for biomedical research', **36**(11 SUPPL. 1), S54–60.

Paternoster, L., Zhurov, A. I., Toma, A. M., Kemp, J. P., St. Pourcain, B., Timpson, N. J., McMahon, G., McArdle, W., Ring, S. M., Smith, G. D., Richmond, S. and Evans, D. M. (2012), 'Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position', *American Journal of Human Genetics* **90**(3), 478–485.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. and Reich, D. (2012), 'Ancient admixture in human history', *Genetics* **192**, 1065–1093.

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

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J. and Nieselt, K. (2016), 'EAGER: efficient ancient genome reconstruction', *Genome Biology* **17**, 60.

Pereira, J. B., Costa, M. D., Vieira, D., Pala, M., Bamford, L., Harich, N., Cherni, L., Alshamali, F., Hatina, J., Rychkov, S., Stefanescu, G., King, T., Torroni, A., Soares, P., Pereira, L. and Richards, M. B. (2017), 'Reconciling evidence from ancient and contemporary genomes: A major source for the European Neolithic within Mediterranean Europe', *Proceedings of the Royal Society B: Biological Sciences* **284**(1851), 201619976.

Pereira, L., Freitas, F., Fernandes, V., Pereira, J. B., Costa, M. D., Costa, S., Máximo, V., Macaulay, V., Rocha, R. and Samuels, D. C. (2009), 'The Diversity Present in 5140 Human Mitochondrial Genomes', *American Journal of Human Genetics* **84**, 628–640.

Pereira, L., Richards, M., Goios, A., Alonso, A., Albarrán, C., Garcia, O., Behar, D. M., Gölge, M., Hatina, J., Al-Gazali, L., Bradley, D. G., Macaulay, V. and Amorim, A. (2005), 'High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium', *Genome Research* **15**(1), 19–24.

Pereira Menaut, G. (1979), 'Inscripciones Romanas de Valentia', *Serie de Trabajos Varios del Servicio de Investigación Prehistrica de Valencia* **64**.

Pérez Jordà, G., Bernabeu Aubán, J., Carrión Marco, Y., García Puchol, O., Molina Balaguer, L. and Gómez Puche, M. (2011), 'La Vital (Gandia, Valencia): vida y muerte en la desembocadura del Serpis durante el III y el I milenio a.C', *Serie de Trabajos Varios del Servicio de Investigación Prehistrica de Valencia* **113**.

Peter, B. M. (2016), 'Admixture, population structure, and f-statistics', *Genetics* **202**(4), 1485–1501.

Phillips, W. D. (1985), *Slavery from Roman Times to the Early Transatlantic Trade*, Manchester University Press.

Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y. and Hinds, D. A. (2016), 'Detection and interpretation of shared genetic influences on 42 human traits', *Nature Genetics* **48**(7), 709–717.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., Anders, A., Pietrusewsky, M., Rollefson, G., Jovanovic, M., Trinhhoang, H., Bar-Oz, G., Oxenham, M., Matsumura, H. and Hofreiter, M. (2015), 'Optimal ancient DNA yields from the inner ear part of the human petrous bone', *PLoS ONE* **10**(6), e0129102.

Pinhasi, R., Thomas, M. G., Hofreiter, M., Currat, M. and Burger, J. (2012), 'The genetic history of Europeans', *Trends. Genet.* **28**, 496–505.

Platt, D. E., Haber, M., Dagher-Kharrat, M. B., Douaihy, B., Khazen, G., Ashrafian Bonab, M., Salloum, A., Mouzaya, F., Luiselli, D., Tyler-Smith, C., Renfrew, C., Matisoo-Smith, E. and Zalloua, P. A. (2017), 'Mapping Post-Glacial expansions: The Peopling of Southwest Asia', *Scientific Reports* **7**, 40338.

Posth, C., Renaud, G., Mittnik, A., Drucker, D. G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., Francken, M., Malina, M., Bolus, M., Lari, M., Gigli, E., Capecchi, G., Crevecoeur, I., Beauval, C., Flas, D., Germonpré, M., Van Der Plicht, J., Cottiaux, R., Gély, B., Ronchitelli, A., Wehrberger, K., Grigorescu, D., Svoboda, J., Semal, P., Caramelli, D., Bocherens, H., Harvati, K., Conard, N. J., Haak, W., Powell, A. and Krause, J. (2016), 'Pleistocene mitochondrial genomes suggest a single major dispersal of non-africans and a late glacial population turnover in Europe', *Current Biology* **26**(6), 827–833.

Poznik, G. D., Xue, Y., Mendez, F. L., Willems, T. F., Massaia, A., Wilson Sayres, M. A., Ayub, Q., McCarthy, S. A., Narechania, A., Kashin, S., Chen, Y., Banerjee, R., Rodriguez-Flores, J. L., Cerezo, M., Shao, H., Gymrek, M., Malhotra, A., Louzada, S., Desalle, R., Ritchie, G. R., Cerveira, E., Fitzgerald, T. W., Garrison, E., Marcketta, A., Mittelman, D., Romanovitch, M., Zhang, C., Zheng-Bradley, X., Abecasis, G. R., McCarroll, S. A., Flicek, P., Underhill, P. A., Coin, L., Zerbino, D. R., Yang, F., Lee, C., Clarke, L., Auton, A., Erlich, Y., Handsaker, R. E., Bustamante, C. D. and Tyler-Smith, C. (2016), 'Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences', *Nature Genetics* **48**(6), 593–599.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000), 'Inference of population structure using multilocus genotype data', *Genetics* **155**(2), 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J. and Sham, P. C. (2007), 'PLINK: A tool set for whole-genome association and population-based linkage analyses', *American Journal of Human Genetics* **81**(3), 559–575.

Racimo, F., Sikora, M., Vander Linden, M., Schroeder, H. and Lalueza-Fox, C. (2020), 'Beyond broad strokes: sociocultural insights from the study of ancient genomes', **21**(6), 355–366.

Racimo, F., Woodbridge, J., Fyfe, R. M., Sikora, M., Sjögren, K. G., Kristiansen, K. and Linden, M. V. (2020), 'The spatiotemporal spread of human migrations during the European Holocene', *Proceedings of the National Academy of Sciences of the United States of America* **117**(16), 8989–9000.

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T. W., Orlando, L., Metspalu, E., Karmin, M., Tambets, K., Rootsi, S., Mägi, R., Campos, P. F., Balanovska, E., Balanovsky, O., Khusnutdinova, E., Litvinov, S., Osipova, L. P., Fedorova, S. A., Voevoda, M. I., Degiorgio, M., Sicheritz-Ponten, T., Brunak, S., Demeshchenko, S., Kivisild, T., Villems, R., Nielsen, R., Jakobsson, M. and Willerslev, E. (2014), 'Upper palaeolithic Siberian genome reveals dual ancestry of native Americans', *Nature* **505**(7481), 87–91.

Ralf, A., Montiel González, D., Zhong, K. and Kayser, M. (2018), 'Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data', *Molecular biology and evolution* **35**(5), 1291–1294.

Rasmussen, D. A. and Noor, M. A. (2009), 'What can you do with 0.1× genome coverage? A case study based on a genome survey of the scuttle fly Megaselia scalaris (Phoridae)', *BMC Genomics* **10**(1), 1–9.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M. T. P., Wang, Y., Raghavan, M., Campos, P. F., Kamp, H. M., Wilson, A. S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E. D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Pierre, T. L., Gronnow, B., Meldgaard, M., Andreasen, C., Fedorova, S. A., Osipova, L. P., Higham, T. F., Ramsey, C. B., Hansen, T. V., Nielsen, F. C., Villems, R., Nielsen, R., Krogh, A., Wang, J. and Willerslev, E. (2010), 'Ancient human genome sequence of an extinct Palaeo-Eskimo', *Nature* **463**(7282), 757–762.

Recio, C. (2011), *Algar de Paláncia. Historia general. Desde la fundación hasta principios del Siglo XXI*, Ayuntamiento de Algar de Pálancia.

Reich, D. (2018), *Who We Are and How We Got Here*, Oxford University Press.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J. J., Kelso, J., Slatkin, M. and Pääbo, S. (2010), 'Genetic history of an archaic hominin group from Denisova cave in Siberia', *Nature* **468**(7327), 1053–1060.

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., Ko, A. M. S., Ko, Y. C., Jinam, T. A., Phipps, M. E., Saitou, N., Wollstein, A., Kayser, M., Pääbo, S. and Stoneking, M. (2011), 'Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania', *American Journal of Human Genetics* **89**(4), 516–528.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L. and Singh, L. (2009), 'Reconstructing Indian population history', *Nature* **461**(7263), 489–494.

Reidla, M., Kivisild, T., Metspalu, E., Kaldma, K., Tambets, K., Tolk, H. V., Parik, J., Loogväli, E. L., Derenko, M., Malyarchuk, B., Bermisheva, M., Zhadanov, S., Pennarun, E., Gubina, M., Golubenko, M., Damba, L., Fedorova, S., Gusar, V., Grechanina, E., Mikerezi, I., Moisan, J. P., Chaventré, A., Khusnutdinova, E., Osipova, L., Stepanov, V., Voevoda, M., Achilli, A., Rengo, C., Rickards, O., De Stefano, G. F., Papiha, S., Beckman, L., Janicijevic, B., Rudan, P., Anagnou, N., Michalodimitrakis, E., Koziel, S., Usanga, E., Geberhiwot, T., Herrnstadt, C., Howell, N., Torroni, A. and Villems, R. (2003), 'Origin and Diffusion of mtDNA Haplogroup X', *American Journal of Human Genetics* **73**(5), 1178–1190.

Renka, R. J. (1988), 'Algorithm 660: QSHEP2D: Quadratic Shepard Method for Bivariate Interpolation of Scattered Data', *ACM Transactions on Mathematical Software (TOMS)* **14**(2), 149–150.

Ribera i Lacomba, A. (1998), *La fundació de València: la ciutat a l'època romanorepublicana, segles II-I a. de C.*, Institució Alfons el Magnànim, Diputació de València.

Ribera i Lacomba, A. (2003), *El papel militar de la fundación de Valentia (138 a.C.): historia y arqueología*, Casa de Velázquez.

Ribera i Lacomba, A. (2008), Depositos rituales de Valentia (Hispania). De la primera fundación republicana (138 a.C.) A la segunda augustea, *in* H. DiGiuseppe and M. Serlorenzi, eds, 'I riti del costruire nelle acque violate', Scienze e Lettere, Roma, Palazzo Massimo, pp. 269–294.

Ribera i Lacomba, A., Escrivà Chover, M. I. and Vioque Hellín, J. (2010), *Guía del centro arqueológico de l'Almoina*, Ayuntamiento de Valencia.

Richards, M. (2003), 'The Neolithic Invasion of Europe', *Annual Review of Anthropology* **32**, 135–162.

Richards, M. B. (2005), *Archaeometry and the Antique Analysis. Organic and Biological Materials*, Encyclopedia of Analytical Science (Second Edition).

Richards, M. B., Macaulay, V. A., Bandelt, H. and Sykes, B. C. (1998), 'Phylogeography of mitochondrial DNA in western Europe', *Annals of Human Genetics* **62**(3), 241–260.

Richards, M. B., Soares, P. and Torroni, A. (2016), 'Palaeogenomics: Mitogenomes and migrations in Europe's past', *Current Biology* **26**(6), R243–246.

Richards, M. and Macaulay, V. (2001), 'The Mitochondrial Gene Tree Comes of Age', *The American Journal of Human Genetics* **68**(6), 1315–1320.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Gölge, M., Dimitrov, D., Hill, E., Romano, V., Calì, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Nørby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A. and Hans-Jürgen, B. (2000), 'Tracing european founder lineages in the near eastern mtDNA pool', *American Journal of Human Genetics* **67**(5), 1251–1276.

Richards, M., Macaulay, V., Torroni, A. and Bandelt, H. J. (2002), 'In search of geographical patterns in European mitochondrial DNA', *American Journal of Human Genetics* **71**(5), 1168–1174.

Ripollès i Alegre, P. P. (2002), La Ceca de Valentia y las monedas de su época, *in* 'Valencia y las primeras ciudades romanas de Hispania', Ajuntament de Valencia, pp. 335–348.

Rivollat, M., Jeong, C., Schiffels, S., Küçükkalıpçı, i., Pemonge, M.-H., Rohrlach, B. A., Alt, K. W., Binder, D., Deguilloux, M.-F. and Haak, W. (2020), 'Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers', *Science Advances* **6**(22), eaaz5344.

Rodríguez-Varela, R., Günther, T., Krzewińska, M., Storå, J., Gillingwater, T. H., MacCallum, M., Arsuaga, J. L., Dobney, K., Valdiosera, C., Jakobsson, M., Götherström, A. and Girdland-Flink, L. (2017), 'Genomic Analyses of Pre-European Conquest Human Remains from the Canary Islands Reveal Close Affinity to Modern North Africans', *Current Biology* **27**(21), 3396–3402.e5.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. and Reich, D. (2015), 'Partial uracil – DNA – glycosylase treatment for screening of ancient DNA', *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**(1660), 20130624.

Rohland, N. and Hofreiter, M. (2007), 'Ancient dna extraction from bones and teeth', *Nature Protocols* **2**(7), 1756–1762.

Rojo Guerra, M., Garrido-Pena, R. and García-Martinez de Lagrán, I. (2005), 'Bell Beakers in the Iberian Peninsula and their european context', *Serie Arte y Arqueología, Universidad de Valladolid* **21**.

Roostalu, U., Kutuev, I., Loogväli, E. L., Metspalu, E., Tambets, K., Reidla, M., Khusnutdinova, E. K., Usanga, E., Kivisild, T. and Villems, R. (2007), 'Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in west Eurasia: The Near Eastern and Caucasian perspective', *Molecular Biology and Evolution* **24**(2), 436–448.

Rubiera Mata, M. (1985), 'Valencia en el pacto de Tudmir', *Sharq Al-Andalus* **2**, 119–120.

Rubio Vela, A. (1980), La ciudad de Valencia en 1348: la peste negra, *in* 'Primer Congreso de Historia del País Valenciano', Vol. 2, Prehistoria, Edades Antigua y Media, celebrado en Valencia del 14 al 18 de abril de 1971, pp. 519–552.

Ruggles, D. F. (2000), *Gardens, Landscape, and Vision in the Palaces of Islamic Spain*, Pennsylvania State University Press.

Ruggles, D. F. (2008), *Islamic gardens and landscapes*, University of Pennsylvania Press.

Ruiz-de Loizaga, S. (2009), *La peste en los reinos peninsulares, segun documentacion del archivo vaticano (1348-1460)*, Museo Vasco de Historia de la Medicina y de la Ciencia.

Ruiz, I. M. (1993), 'Bronze Age metallurgy in southeast Spain', *Antiquity* **67**(254), 46–57.

Ruiz Zapatero, G. and Almagro Gorbea, M. (1992), 'Paleoetnología de la Península Ibérica. Reflexiones y perspectivas de futuro', *Complutum* **2**(2), 469–499.

Saag, L., Varul, L., Scheib, C. L., Stenderup, J., Allentoft, M. E., Saag, L., Pagani, L., Reidla, M., Tambets, K., Metspalu, E., Kriiska, A., Willerslev, E., Kivisild, T. and Metspalu, M. (2017), 'Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe', *Current Biology* **27**(14), 2185–2193.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988), 'Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase', *Science* **239**(4839), 487–491.

Saiki, R., Scharf, S., Faloona, F., Mullis, K. and Horn, G. (1985), 'Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia', *Science* **230**(4732), 1350–1354.

Sánchez-Quinto, F., Schroeder, H., Ramirez, O., Ávila-Arcos, M., Pybus, M., Olalde, I., Velazquez, A., Marcos, M., Encinas, J., Bertranpetit, J., Orlando, L., Gilbert, M. and Lalueza-Fox, C. (2012), 'Genomic affinities of two 7,000-year-old Iberian hunter-gatherers', *Current Biology* **22**(16), 1494–1499.

Sánchez-Velasco, P., Escribano de Diego, J., Paz-Miguel, J. E., Ocejo-Vinyals, G. and Leyva-Cobián, F. (1999), 'HLA-DR, DQ nucleotide sequence polymorphisms in the Pasiegos (Pas valleys, Northern Spain) and comparison of the allelic and haplotypic frequencies with those of other European populations', *Tissue Antigens* **53**(1), 65–73.

Sawyer, S., Renaud, G., Viola, B., Hublin, J. J., Gansauge, M. T., Shunkov, M. V., Derevianko, A. P., Prüfer, K., Kelso, J. and Pääbo, S. (2015), 'Nuclear and mitochondrial DNA sequences from two Denisovan individuals', *Proceedings of the National Academy of Sciences of the United States of America* **112**(51), 15696–15700.

Schönemann, P. H. (1966), 'A generalized solution of the orthogonal procrustes problem', *Psychometrika* **31**(1), 1–10.

Schroeder, H., Margaryan, A., Szmyt, M., Theulot, B., Włodarczak, P., Rasmussen, S., Gopalakrishnan, S., Szczepanek, A., Konopka, T., Jensen, T. Z., Witkowska, B., Wilk, S., Przybyła, M. M., Pospieszny, Ł., Sjögren, K. G., Belka, Z., Olsen, J., Kristiansen, K., Willerslev, E., Frei, K. M., Sikora, M., Johannsen, N. N. and Allentoft, M. E. (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* **166**(22), 10705–10710.

Schubert, M., Lindgreen, S. and Orlando, L. (2016), 'AdapterRemoval v2: Rapid adapter trimming, identification, and read merging', *BMC Research Notes* **9**(1), 8.

Schuenemann, V. J., Peltzer, A., Welte, B., Van Pelt, W. P., Molak, M., Wang, C. C., Furtwängler, A., Urban, C., Reiter, E., Nieselt, K., Teßmann, B., Francken, M., Harvati, K., Haak, W., Schiffels, S. and Krause, J. (2017), 'Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods', *Nature Communications* **8**, 15694.

Schulz Paulsson, B. (2019), 'Radiocarbon dates and Bayesian modeling support maritime diffusion model for megaliths in Europe', *Proceedings of the National Academy of Sciences of the United States of America* **116**(9), 3460–3465.

Secher, B., Fregel, R., Larruga, J. M., Cabrera, V. M., Endicott, P., Pestano, J. J. and González, A. M. (2014), 'The history of the North African mitochondrial DNA haplogroup U6 gene flow into the African, Eurasian and American continents', *BMC Evolutionary Biology* **14**(1), 109.

Sero, D., Zaidi, A., Li, J., White, J. D., Zarzar, T. B., Marazita, M. L., Weinberg, S. M., Suetens, P., Vandermeulen, D., Wagner, J. K., Shriver, M. D. and Claes, P. (2019), 'Facial recognition from DNA using face-to-DNA classifiers', *Nature Communications* **10**(1), 2557.

Shinde, V., Narasimhan, V. M., Rohland, N., Mallick, S., Mah, M., Lipson, M., Nakatsuka, N., Adamski, N., Broomandkhoshbacht, N., Ferry, M., Lawson, A. M., Michel, M., Oppenheimer, J., Stewardson, K., Jadhav, N., Kim, Y. J., Chatterjee, M., Munshi, A., Panyam, A., Waghmare, P., Yadav, Y., Patel, H., Kaushik, A., Thangaraj, K., Meyer, M., Patterson, N., Rai, N. and Reich, D. (2019), 'An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers', *Cell* **179**(3), 729–735.

Shriver, M. D., Parra, E. J., Dios, S., Bonilla, C., Norton, H., Jovel, C., Pfaff, C., Jones, C., Massac, A., Cameron, N., Baron, A., Jackson, T., Argyropoulos, G., Jin, L., Hoggart, C. J., McKeigue, P. M. and Kittles, R. A. (2003), 'Skin pigmentation, biogeographical ancestry and admixture mapping', *Human Genetics* **112**(4), 387–399.

Sikora, M., Carpenter, M. L., Moreno-Estrada, A., Henn, B. M., Underhill, P. A., Sánchez-Quinto, F., Zara, I., Pitzalis, M., Sidore, C., Busonero, F., Maschio, A., Angius, A., Jones, C., Mendoza-Revilla, J., Nekhrizov, G., Dimitrova, D., Theodossiev, N., Harkins, T. T., Keller, A., Maixner, F., Zink, A., Abecasis, G., Sanna, S., Cucca, F. and Bustamante, C. D. (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), e1004353.

Silva, M., Justeau, P., Rodrigues, S., Oteo-Garcia, G., Dulias, K., Foody, G., Fichera, A., Yau, B., Rito, T., Wilson, J., Gandini, F., Edwards, C., Pala, M., Soares, P. and Richards, M. (2019), 'Untangling Neolithic and Bronze Age mitochondrial lineages in South Asia', *Annals of Human Biology* **46**(2), 140–144.

Silva, M., Oliveira, M., Vieira, D., Brandão, A., Rito, T., Pereira, J. B., Fraser, R. M., Hudson, B., Gandini, F., Edwards, C., Pala, M., Koch, J., Wilson, J. F., Pereira, L., Richards, M. B. and Soares, P. (2017), 'A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals', *BMC evolutionary biology* **17**(1).

Simon, L. J., ed. (1996), *Iberia and the Mediterranean World of the Middle Ages: Studies in Honor of Robert I. Burns, S.J.*, BRILL.

Skoglund, P., Storå, J., Götherström, A. and Jakobsson, M. (2013), 'Accurate sex identification of ancient human remains using DNA shotgun sequencing', *Journal of Archaeological Science* **40**(12), 1427–1432.

Skourtanioti, E., Erdal, Y. S., Frangipane, M., Balossi Restelli, F., Yener, K. A., Pinnock, F., Matthiae, P., Özbal, R., Schoop, U. D., Guliyev, F., Akhundov, T., Lyonnet, B., Hammer, E. L., Nugent, S. E., Burri, M., Neumann, G. U., Penske, S., Ingman, T., Akar, M., Shafiq, R., Palumbi, G., Eisenmann, S., D'Andrea, M., Rohrlach, A. B., Warinner, C., Jeong, C., Stockhammer, P. W., Haak, W. and Krause, J. (2020), 'Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus', *Cell* **181**(5), 1158–1175.

Soares, P. A., Trejaut, J. A., Rito, T., Cavadas, B., Hill, C., Eng, K. K., Mormina, M., Brandão, A., Fraser, R. M., Wang, T. Y., Loo, J. H., Snell, C., Ko, T. M., Amorim, A., Pala, M., Macaulay, V., Bulbeck, D., Wilson, J. F., Gusmão, L., Pereira, L., Oppenheimer, S., Lin, M. and Richards, M. B. (2016), 'Resolving the ancestry of Austronesian-speaking populations', *Human Genetics* **135**(3), 309–326.

Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H. J., Torroni, A. and Richards, M. B. (2010), 'The Archaeogenetics of Europe', *Current Biology* **20**(4), R174–183.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V. and Richards, M. B. (2009), 'Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock', *American Journal of Human Genetics* p. 740–759.

Sokal, R. R. (2001), 'Archaeogenetics: DNA and the Population Prehistory of Europe', *The American Journal of Human Genetics* **69**(1), 243–244.

Soldevila, F., Bruguera, J. and Ferrer i Mallol, M. T. (2007), *Les quatre grans cròniques*, Institut d'Estudis Catalans.

Solé-Morata, N., Villaescusa, P., García-Fernández, C., Font-Porterias, N., Illescas, M. J., Valverde, L., Tassi, F., Ghirotto, S., Férec, C., Rouault, K., Jiménez-Moreno, S., Martínez-Jarreta, B., Pinheiro, M. F., Zarrabeitia, M. T., Carracedo, Á., De Pancorbo, M. M. and Calafell, F. (2017), 'Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ', *Scientific Reports* **7**(1), 7341.

Stevens, R. (1999), 'The history of haemophilia in the royal families of europe', *British Journal of Haematology* **105**(1), 25–32.

Stokowski, R. P., Pant, P. V., Dadd, T., Fereday, A., Hinds, D. A., Jarman, C., Filsell, W., Ginger, R. S., Green, M. R., Van Der Ouderaa, F. J. and Cox, D. R. (2007), 'A genomewide association study of skin pigmentation in a South Asian population', *American Journal of Human Genetics* **81**(6), 1119–1132.

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M. H. Y., Konkel, M. K., Malhotra, A., Stütz, A. M., Shi, X., Casale, F. P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., Malig, M., Chaisson, M. J., Walter, K., Meiers, S., Kashin, S., Garrison, E., Auton, A., Lam, H. Y., Mu, X. J., Alkan, C., Antaki, D., Bae, T., Cerveira, E., Chines, P., Chong, Z., Clarke, L., Dal, E., Ding, L., Emery, S., Fan, X., Gujral, M., Kahveci, F., Kidd, J. M., Kong, Y., Lameijer, E. W., McCarthy, S., Flicek, P., Gibbs, R. A., Marth, G., Mason, C. E., Menelaou, A., Muzny, D. M., Nelson, B. J., Noor, A., Parrish, N. F., Pendleton, M., Quitadamo, A., Raeder, B., Schadt, E. E., Romanovitch, M., Schlattl, A., Sebra, R., Shabalin, A. A., Untergasser, A., Walker, J. A., Wang, M., Yu, F., Zhang, C., Zhang, J., Zheng-Bradley, X., Zhou, W., Zichner, T., Sebat, J., Batzer, M. A., McCarroll, S. A., Mills, R. E., Gerstein, M. B., Bashir, A., Stegle, O., Devine, S. E., Lee, C., Eichler, E. E. and Korbel, J. O. (2015), 'An integrated map of structural variation in 2,504 human genomes', *Nature* **526**(7571), 75–81.

Szécsényi-Nagy, A., Roth, C., Brandt, G., Rihuete-Herrada, C., Tejedor-Rodríguez, C., Held, P., García-Martínez-De-Lagrán, Í., Arcusa Magallón, H., Zesch, S., Knipper, C., Bánffy, E., Friederich, S., Meller, H., Bueno Ramírez, P., Barroso Bermejo, R., De Balbín Behrmann, R., Herrero-Corral, A. M., Flores Fernández, R., Alonso Fernández, C., Jiménez Echevarria, J., Rindlisbacher, L., Oliart, C., Fregeiro, M. I., Soriano, I., Vicente, O., Micó, R., Lull, V., Soler Díaz, J., López Padilla, J. A., Roca De Togores Muñoz, C., Hernández Pérez, M. S., Jover Maestre, F. J., Lomba Maurandi, J., Avilés Fernández, A., Lillios, K. T., Silva, A. M., Magalhães Ramalho, M., Oosterbeek, L. M., Cunha, C., Waterman, A. J., Roig Buxó, J., Martínez, A., Ponce Martínez, J., Hunt Ortiz, M., Mejías-García, J. C., Pecero Espín, J. C., Cruz-Auñón Briones, R., Tomé, T., Carmona Ballestero, E., Cardoso, J. L., Araújo, A. C., Liesau Von Lettow-Vorbeck, C., Blasco Bosqued, C., Ríos Mendoza, P., Pujante, A., Royo-Guillén, J. I., Esquembre Beviá, M. A., Dos Santos Goncalves, V. M., Parreira, R., Morán Hernández, E., Méndez Izquierdo, E., Vega Y Miguel, J., Menduiña García, R., Martínez Calvo, V., López Jiménez, O., Krause, J., Pichler, S. L., Garrido-Pena, R., Kunst, M., Risch, R., Rojo-Guerra, M. A., Haak, W. and Alt, K. W. (2017), 'The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age', *Scientific Reports* **7**(1), 15644.

Tallavaara, M., Luoto, M., Korhonen, N., Järvinen, H. and Seppä, H. (2015), 'Human population dynamics in Europe over the Last Glacial Maximum', *Proceedings of the National Academy of Sciences of the United States of America* **112**(27), 8232–8237.

Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogväli, E. L., Tolk, H. V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova, E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J. P., Barać, L., Peričić, M., Rudan, P., Terzić, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, G. F., Anagnou, N., Pappa, K. I., Michalodimitrakis, E., Ferák, V., Füredi, S., Komel, R., Beckman, L. and Villems, R. (2004), 'The Western and Eastern Roots of the Saami - The Story of Genetic "Outliers" Told by Mitochondrial DNA and Y Chromosomes', *American Journal of Human Genetics* **74**(4), 661–682.

Tang, H. and Barsh, G. S. (2017), 'Skin color variation in Africa', *Science* **358**(6365), 867–868.

Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P. (2013), 'Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration', *Briefings in Bioinformatics* **14**(2), 178–192.

Torres Balbás, L. (1951), 'La población musulmana de Valencia en 1238', *Al-Andalus* **XVI**, 345–348.

Torró, J. (2012), *La conquista de Al-Andalus en el siglo XIII: La conquista del reino de Valencia. Un proceso de colonización medieval desde la arqueología del territorio*, Servicio de Publicaciones de la Universidad de Murcia.

Torroni, A., Achilli, A., Macaulay, V., Richards, M. and Bandelt, H. J. (2006), 'Harvesting the fruit of the human mtDNA tree', **22**(6), 339–345.

Torroni, A., Bandelt, H. J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M. L., Bonné-Tamir, B. and Scozzari, R. (1998), 'mtDNA analysis reveals a major late paleolithic population expansion from southwestern to northeastern Europe', *American Journal of Human Genetics* **62**(5), 1137–1152.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L. and Wallace, D. C. (1996), 'Classification of european mtDNAs from an analysis of three European populations', *Genetics* **144**(4), 1835–1850.

Torroni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V., Larsen, M., Smith, D. G., Vullo, C. M. and Wallace, D. C. (1993), 'Asian affinities and continental radiation of the four founding Native American mtDNAs', *American Journal of Human Genetics* **53**(3), 563–590.

Tyler-Smith, C. and Krausz, C. (2009), 'The will-o'-the-wisp of genetics — hunting for the azoospermia factor gene', *New England Journal of Medicine* **360**(9), 925–927.

Ubieto, A. (1975), *Orígenes del reino de Valencia: cuestiones cronológicas sobre su reconquista*, ANUBAR Ediciones.

Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazón Lahr, M., Foley, R. A., Oefner, P. J. and Cavalli-Sforza, L. L. (2001), 'The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations', *Annals of Human Genetics* **65**(1), 43–62.

Valdiosera, C., Günther, T., Vera-Rodríguez, J. C., Ureña, I., Iriarte, E., Rodríguez-Varela, R., Simões, L. G., Martínez-Sánchez, R. M., Svensson, E. M., Malmström, H., Rodríguez, L., Bermúdez de Castro, J.-M., Carbonell, E., Alday, A., Hernández Vera, J. A., Götherström, A.,

Carretero, J.-M., Arsuaga, J. L., Smith, C. I. and Jakobsson, M. (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* **115**(13), 3428–3433.

Valverde, L., Illescas, M. J., Villaescusa, P., Gotor, A. M., Garc'a, A., Cardoso, S., Algorta, J., Catarino, S., Rouault, K., Férec, C., Hardiman, O., Zarrabeitia, M., Jiménez, S., Pinheiro, M. F., Jarreta, B. M., Olofsson, J., Morling, N. and De Pancorbo, M. M. (n.d.), *European Journal of Human Genetics* **24**, 437–441.

van de Loosdrecht, M., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E. H., El Hajraoui, M. A., Amzazi, S., Hublin, J. J., Pääbo, S., Schiffels, S., Meyer, M., Haak, W., Jeong, C. and Krause, J. (2018), 'Pleistocene north african genomes link near eastern and sub-saharan african human populations', *Science* **360**(6388), 548–552.

van de Loosdrecht, M. S., Mannino, M. A., Talamo, S., Villalba-Mouco, V., Posth, C., Aron, F., Brandt, G., Burri, M., Freund, C., Radzeviciute, R., Stahl, R., Wissgott, A., Klausnitzer, L., Nagel, S., Meyer, M., Tagliacozzo, A., Piperno, M., Tusa, S., Collina, C., Schimmenti, V., Salvo, R. D., Prüfer, K., Hublin, J.-J., Schiffels, S., Jeong, C., Haak, W. and Krause, J. (2020), 'Genomic and dietary transitions during the Mesolithic and Early Neolithic in Sicily', *bioRxiv* .

van der Maaten, L. and Hinton, G. (2008), 'Visualizing data using t-SNE', *Journal of Machine Learning Research* **9**, 2579–2605.

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. and Thermes, C. (2018), 'The Third Revolution in Sequencing Technology', *Trends in Genetics* **34**(9), 666–681.

van Oven, M. (2015), 'PhyloTree Build 17: Growing the human mitochondrial DNA tree', *Forensic Science International: Genetics Supplement Series* **5**, E392–E394.

Vander Linden, M. (2007), 'What linked the Bell Beakers in third millennium BC Europe?', *Antiquity* **81**(312), 343–352.

Verdugo, M. P., Mullin, V. E., Scheu, A., Mattiangeli, V., Daly, K. G., Delser, P. M., Hare, A. J., Burger, J., Collins, M. J., Kehati, R., Hesse, P., Fulton, D., Sauer, E. W., Mohaseb, F. A., Davoudi, H., Khazaeli, R., Lhuillier, J., Rapin, C., Ebrahimi, S., Khasanov, M., Farhad Vahidi, S. M., MacHugh, D. E., Ertuğrul, O., Koukouli-Chrysanthaki, C., Sampson, A., Kazantzis, G., Kontopoulos, I., Bulatovic, J., Stojanović, I., Mikdad, A., Benecke, N., Linstädter, J., Sablin, M., Bendrey, R., Gourichon, L., Arbuckle, B. S., Mashkour, M., Orton, D., Horwitz, L. K., Teasdale, M. D. and Bradley, D. G. (2019), 'Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent', *Science* **365**(6449), 173–176.

Villalba-Mouco, V., van de Loosdrecht, M. S., Posth, C., Mora, R., Martínez-Moreno, J., Rojo-Guerra, M., Salazar-García, D. C., Royo-Guillén, J. I., Kunst, M., Rougier, H., Crevecoeur, I., Arcusa-Magallón, H., Tejedor-Rodríguez, C., García-Martínez de Lagrán, I., Garrido-Pena, R., Alt, K. W., Jeong, C., Schiffels, S., Utrilla, P., Krause, J. and Haak, W. (2019), 'Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula', *Current Biology* **29**(7), 1169–1177.

Wallace, D. C. and Torroni, A. (1992), 'American Indian prehistory as written in the mitochondrial DNA: a review.', *Human Biology* **64**(3), 403–416.

Walsh, S., Chaitanya, L., Breslin, K., Muralidharan, C., Bronikowska, A., Pospiech, E., Koller, J., Kovatsi, L., Wollstein, A., Branicki, W., Liu, F. and Kayser, M. (2017), 'Global skin colour prediction from DNA', *Human Genetics* **136**(7), 847–863.

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W. and Kayser, M. (2013), 'The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA', *Forensic Science International: Genetics* **7**(1), 98–115.

Watt, W. M. and Cachia, P. (1965), *A History of Islamic Spain*, Edinburgh University Press.

Wei, W., Ayub, Q., Chen, Y., McCarthy, S., Hou, Y., Carbone, I., Xue, Y. and Tyler-Smith, C. (2013), 'A calibrated human Y-chromosomal phylogeny based on resequencing', *Genome Research* **23**(2), 388–395.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H. J., Kronenberg, F., Salas, A. and Schönherr, S. (2016), 'HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing', *Nucleic Acids Research* **44**(W1), W58–W63.

Wetterstrand, K. (2018), 'DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)', *National Human Research Institute* .
URL: *www.genome.gov/sequencingcostsdata*

White, J., Indencleef, K., Naqvi, S., Eller, R. J., Roosenboom, J., Lee, M. K., Li, J., Mohammed, J., Richmond, S., Quillen, E. E., Norton, H. L., Feingold, E., Swigut, T., Marazita, M. L., Peeters, H., Hens, G., Shaffer, J. R., Wysocka, J., Walsh, S., Weinberg, S. M. and Shriver, M. (2020), 'Insights into the genetic architecture of the human face', *bioRxiv* .

Willerslev, E. and Cooper, A. (2005), 'Ancient DNA', *Proceedings of the Royal Society B: Biological Sciences* **272**(1558), 3–16.

Wilson, R. T., Roff, A. N., Dai, P. J., Fortugno, T., Douds, J., Chen, G., Nikiforova, S. O., Chinchilli, V. M., Hartman, T. J., Demers, L. M., Canfield, V. A., Cheng, K. C., Grove, G. L., Barnholtz-Sloan, J., Frudakis, T., Shriver, M. D., Cheng, K. C. and Wilson, R. T. (2011), 'Genetic ancestry, skin reflectance and pigmentation genotypes in association with serum vitamin D metabolite balance', *Hormone Molecular Biology and Clinical Investigation* **7**(1), 279–293.

Wright, R. (2012), 'Late and Vulgar Latin in Muslim Spain: the African connection', *MOM Éditions* **49**(1), 35–54.

Xing, S., Martinón-Torres, M. and Bermúdez de Castro, J. M. (2019), 'Late middle pleistocene hominin teeth from tongzi, southern china', *Journal of Human Evolution* **130**, 96–108.

Xiong, Z., Dankova, G., Howe, L. J., Lee, M. K., Hysi, P. G., De Jong, M. A., Zhu, G., Adhikar, K., Li, D., Li, Y., Pan, B., Feingold, E., Marazita, M. L., Shaffer, J. R., McAloney, K., Xu, S., Jin, L., Wang, S., De Vri, F. M., Lendemeije, B., Richmond, S., Zhurov, A., Lewis, S., Sharp, G., Paternoster, L., Thompson, H., Gonzalez-Jose, R., Catira Bortolini, M., Canizales-Quinteros, S., Gallo, C., Poletti, G., Bedoya, G., Rothhammer, F., Uitterlinden, A. G., Ikram, M. A., Wolvius, E. B., Kushner, S. A., Nijsten, T., Palstra, R. J., Boehringer, S., Medland, S. E., Tang, K., Ruiz-Linares, A., Martin, N. G., Spector, T. D., Stergiakouli, E., Weinberg, S. M., Liu, F. and Kayser, M. (2019), 'Novel genetic loci affecting facial shape variation in humans', *eLife* **8**, e49898.

Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C. and Saunders, S. R. (1998), 'Improved DNA extraction from ancient bones using silica based spin columns', *American Journal of Physical Anthropology* **105**(4), 539–543.

Zalloua, P., Collins, C. J., Gosling, A., Biagini, S. A., Costa, B., Kardailsky, O., Nigro, L., Khalil, W., Calafell, F. and Matisoo-Smith, E. (2018), 'Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza', *Scientific Reports* **8**(1), 17567.

Zaorska, K., Zawierucha, P. and Nowicki, M. (2019), 'Prediction of skin color, tanning and freckling from DNA in Polish population: linear regression, random forest and neural network approaches', *Human Genetics* **138**(6), 635–647.

Zheng, H. X., Yan, S., Qin, Z. D. and Jin, L. (2012), 'MtDNA analysis of global populations support that major population expansions began before Neolithic Time', *Scientific Reports* **2**, 745.

Zilhão, J., Anesin, D., Aubry, T., Badal, E., Cabanes, D., Kehl, M., Klasen, N., Lucena, A., Martín-Lerma, I., Martínez, S., Matias, H., Susini, D., Steier, P., Wild, E. M., Angelucci, D. E., Villaverde, V. and Zapata, J. (2017), 'Precise dating of the Middle-to-Upper Paleolithic transition in Murcia (Spain) supports late Neandertal persistence in Iberia', *Heliyon* **3**(11), e00435.