Chapter 11

aDNA: an investigation of uniparental genetic heritage in Neolithic Malta

Bruno Ariano, Valeria Mattiangeli, Rowan McLaughlin, Ronika K. Power, Jay T. Stock, Bernardette Mercieca-Spiteri, Simon Stoddart, Caroline Malone & Dan Bradley

11.1. Introduction

11.1.1. The genome and ancient DNA

Modern genetic studies and the use of the biological component offer considerable potential in the study of past individuals and populations. The genome refers to the full genetic component of an organism which is passed through generation from parents to offspring. The key components of this structure are molecules called nucleotide base which are codified in four different letters: A(adenine), T(thymine), G(guanine) and C(cytosine). These molecules are bonded together within the DNA (Deoxyribonucleic acid) structure. The way these bases are ordered in a genome is, with the exception of monozygotic twins, unique to each individual; for example, two unrelated individuals differ for 3 million nucleotides over approximately 3 billion that compose a human genome. The study of DNA can help to estimate how different people are related, past and present. Using modern biological techniques, DNA can be extracted intact from living persons without losing information. However, when dealing with samples that date far back into the past, dead cells cannot preserve the integrity of the genetic information, and therefore old/ancient DNA is difficult to reassemble. Nevertheless, the ancient DNA (aDNA) field has an important role in dealing and using this type of information to study the history and evolution of ancient organisms. In recent years, thanks to the advancement of a new generation of DNA sequencing techniques, aDNA studies have revolutionized most of the previous concepts about genetics and history and shed light on the origin of different species.

11.1.2. Ancient DNA

The field of aDNA emerged in 1984 when Russ Higuchi and colleagues (Higuchi *et al*. 1984) extracted a fragment of DNA from a dry tissue of a quagga, an historical relative of the horse family. Soon after, Pääbo (1985) reported the first aDNA extraction from an ancient mummy. aDNA became even more prolific, giving the opportunity to analyse a multitude of material, such as bones (Hagelberg *et al*. 1989), hair (Gilbert *et al*. 2004) and even parchment (Teasdale *et al*. 2015).

Many hundreds of ancient genomes from different periods and parts of the world have been screened with high resolution, making it possible in particular, to shed light on human migrations (Mathieson *et al*. 2015) and on animal domestication (Daly *et al*. 2018; Zeder *et al*. 2006). Despite this recent progress, there are still some challenges that arise when dealing with ancient DNA samples.

Due to spontaneous damage that occurs after death, the DNA in ancient samples is usually present in short fragments with a size range between 50 to 70 nucleotides. With smaller and more numerous fragments, it is more difficult to assemble the DNA molecule in its original form. Moreover, due to the lack of a repair system in dead cells, spontaneous mutations in nucleotide base pairs accumulate. A study published by Skoglund *et al*. (2014) showed that the amount of a particular type of DNA mutations, deamination, in a sample is proportional to its age. If not taken in consideration, these damages can lead to erroneous interpretation of DNA results during population and evolutionary genetic analyses. In recent years, the deamination problem has been partially solved thanks to particular software that can target and quantify these specific patterns of postmortem damages (Jónsson *et al*. 2013).

A third problem that emerges when working with ancient samples is the low quantity of endogenous DNA present. These values can be as low as 0.1% (Stoneking & Krause 2011), posing a problem from bacterial and human genome contamination. For this reason, it is important that the extraction of DNA from ancient samples is carried out in special cleanroom

facilities, where particular procedures are adopted to keep the bacterial and human contamination levels as low as possible (MacHugh *et al.* 2000). Once the DNA has been extracted, two common approaches are used for obtaining the sequence data, shotgun genome sequencing and targeted capture.

The first method consists of fragmenting and sequencing the available genome of a sample. This technique has been extensively used for modern DNA analysis and can also be applied to ancient genomes as long as the samples are of good quality. The main advantage of this method is the opportunity to cover every position in a genome and study mutations that are still unknown or present in low frequency in a comparator population.

The targeted capture method, on the other hand, usually focuses on a predefined set of high frequency variants (referred as SNPs) that are enriched using a custom-built probe. This technique has the advantage of obtaining more data compared with the WGS approach, especially when dealing with samples with low DNA quantity. With more than 1.2 million SNPs covered (Mathieson *et al.* 2015), this technique has become frequently used for ancient DNA analysis. However, the main drawback of this approach resides in the limited number of analyses that can be performed using these variant positions. For example, the majority of rare mutations that are important for Mendelian diseases are not covered by the capture method and therefore cannot be directly studied.

11.1.3. Background: the genetic context of the Mesolithic in Europe

The Mesolithic period dates from the end of the Epipalaeolithic period, around 12,000 years ago, and it was heralded by rapidly rising temperatures accompanied by the establishment of a Holocene forest biome across Europe. These conditions contrasted with the preceding tundra and glacial conditions (Clark *et al.* 2009). During the Mesolithic, human populations were scattered in groups around Europe, living in small groups, and following a typical hunter-gatherer (HG) existence. Different published studies have investigated the genetic background of these populations, dividing them into three main groups. On the western side of Europe, individuals from Spain, Hungary and Luxembourg have been reported as genetically similar, and for this reason they have been identified as the Western Hunter-Gatherer (WHG) group. Also included in this group are individuals from eastern Europe that displayed a similar pattern of genetic affinity (González-Fortes *et al.* 2017; Jones *et al.* 2017). On the eastern side of Europe, two Mesolithic individuals from Russia were found to have some marked genetic distinctions from the WHG group (Haak *et al.* 2015). These individuals, who lived approximately 8000 years ago, are now considered part of a genetically distinct cluster identified as Eastern Hunter-Gatherer (EHG). This group can be considered a mix between WHG populations and Upper Palaeolithic individuals from Siberia (Mal'ta and Afontova Gora) (Raghavan *et al.* 2014; Fu *et al.* 2016). The influence of this group on other populations has been detected in hunter-gatherer individuals from Sweden and the Balkans (Gonzales Fortes *et al.* 2017; Lazaridis *et al.* 2014; Lazaridis 2018) and in populations from the steppe during the Bronze Age period (Haak *et al.* 2015).

A third genetic cluster is formed by two individuals found in western Georgia that are now identified as members of a Caucasus Hunter Gatherer (CHG) group. This population diverged from the WHG group long before the Last Glacial Maximum, approximately between 40 and 50 thousand years ago. It is a population that had a strong influence in both Mesolithic and Neolithic populations from Iran, and its influence is still present in the genomes of modern populations from Southern Caucasus (Jones *et al.* 2015).

11.1.4 The genetic impact of the agricultural revolution

The adoption of agriculture was a turning point in human history which occurred in different parts of Eurasia and the Middle East between 12,000 and 7000 BC. In the Levant and Southern Anatolia between 11,000 and 9600 BC, local hunter-gatherer populations began to adopt a farming and sedentary lifestyle, accompanied by animal and plant domestication. With the help of aDNA studies it was discovered in 2016 that the origin of Near Eastern farming had two genetically distinct roots, one residing in Anatolia and the other in Iran (Broushaki *et al.* 2016). Between *c.* 6,600 and 6,500 BC Iranian farmers spread genetically towards eastern Eurasia whilst the Anatolian farming communities became well-established in north-western Anatolia and had begun to move into Europe via Greece and the Balkans (Lazaridis 2018; Lazaridis *et al*. 2014). The arrival of farmers in Europe represented a genetic replacement with limited admixture from the local hunter-gatherer populations. This admixture became evident in 2009, when aDNA showed a genetic discontinuity between these two populations in Europe during the Neolithic period (Malstrom *et al.* 2009). More recent studies have emphasized this observation, giving a better view of the phenomenon. From the lower part of the Danube, the Anatolian farming culture reached the Hungarian plain by 5500 BC and gave birth to different farming groups

(Starčevo, Körös and Criş). Some centuries later, from the same region, another cultural movement started to spread into north-west Europe with a new form of decorated pottery called the Linearbandkeramik (LBK) (Cunliffe 2015). A second culturally different wave of Neolithicization moved from the Adriatic Balkans through to the Mediterranean coast where it is associated with the pottery of the Impressed and Cardial traditions pottery style. The Impressed Ware culture was more closely associated with regions across Italy towards the Ligurian coast, whilst a variant of this pottery group, the Cardial Ware culture, arrived in Provence and extended towards the Atlantic and Portugal (Price 2000). It is important however, to point that these cultures were different, even though they were all close genetically to the same Anatolian Neolithic source (Olalde *et al.* 2015, Mathieson *et al.* 2018). The earliest Neolithic settlements in Italy, which date from about 6200 BC, are located along the lowland coastal areas of south-east Italy (the Apulian Salento peninsula and Tavoliere) (Malone 2003; Natali & Forgia 2018). Very high densities (*c.* one site per 3 km²) of ditched settlements across the area signal a major population increase (Whitehouse 2013). Adoption of the Neolithic economy then rapidly spread westward into Calabria (Morter & Robb 2010) and Sicily (Leighton 1999; Natali & Forgia 2018), reaching Malta by at least 5800 BC, based on environmental evidence (see Volume 1, Chapters 3 & 4), with clear archaeological traces present in the archipelago by 5500 BC (see Volume 2, Chapter 2*).*

11.1.5. Arrival in Malta

The evidence supplied by archaeology, particularly the affinities between Għar Dalam and early Neolithic Impressed Wares of Southern Italy, strongly suggest that the source population of the Neolithic expansion into the Maltese Islands were located in Southern Italy and Sicily (see Volume 2, Chapter 10). Theories of an earlier colonization of Malta have been debated, but since hunter-gatherer populations require a large space for foraging, it seems unlikely that Malta would have been a viable long-term home before the advent of agriculture (Malone 1997–8). From the first evidence of human settlement, the early Maltese society evolved through different cultural phases: Għar Dalam, Grey Skorba, Red Skorba and finally Żebbuġ, signalling the start of the Temple Period and an increasingly distinctive island culture. In this last phase, the use of rock-cut tombs, containing collective burials and distinctive pottery defined the island culture (Malone *et al.* 1995).

Subsequent cultural phases (the Temple Period) witnessed an unprecedented development in Maltese society, culminating in the Tarxien phase between 2800 and 2400 BC (Volume 2, Chapter 2). During the Tarxien phase, collective burial in the elaborate Circle cave

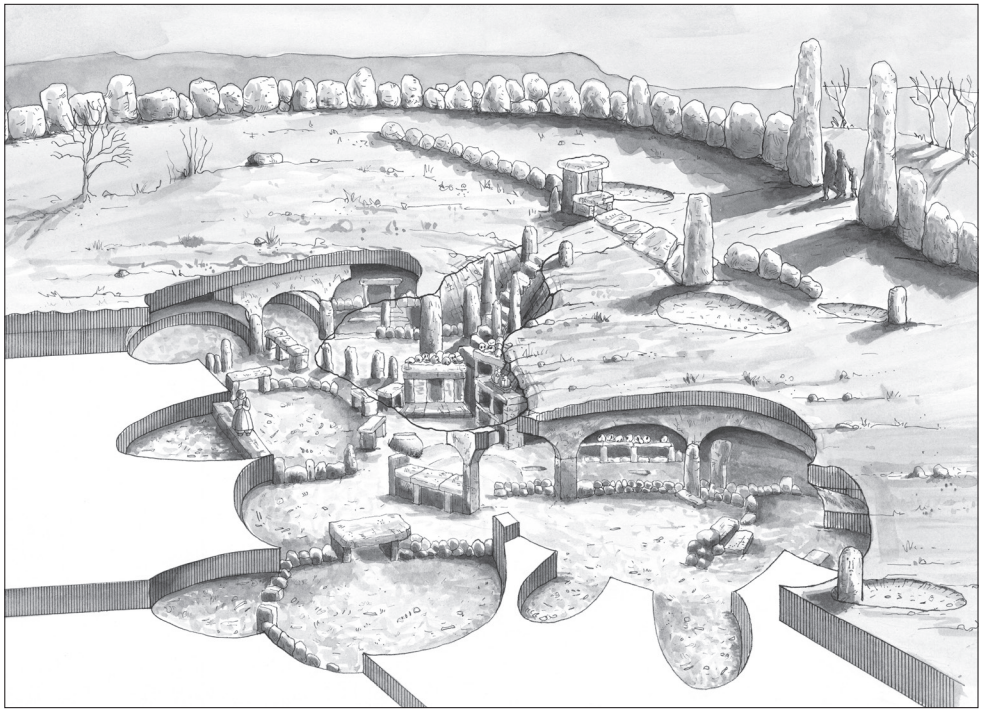

Figure 11.1. *Reconstruction of the Circle (Malone* et al*. 2009d).*

complex on Gozo (Fig. 11.1) and at the Ħal Saflieni Hypogeum in Malta represent exceptional mortuary sites. The Circle excavations unearthed the individuals analysed for this study in the early 1990s (Malone *et al.* 2009d) and are the subject of additional study in this volume. The ancient DNA work we report here was undertaken in collaboration with the *FRAGSUS Project* (2013–2018) as part of a programme of environmental and archaeological research, including an extensive re-assessment of the Circle, applying additional radiocarbon dating and stable isotope studies. The overall aim of this research has been to understand better the cultural, economic and environmental dynamics of prehistoric Malta (Malone *et al.* 2019; Ariano *et al*. 2022).

11.2. Research questions

Since ancient times the Mediterranean Sea has represented one of the most important routes for migration in southern Europe. For example, during the late Neolithic period there is proof of both a cultural and a direct genetic connection between Portuguese and Greek Neolithic populations (Hofmanova *et al.* 2016). Despite this evidence, the prehistoric population history of South Europe remains under-explored in terms of genetic studies. In contrast, most aDNA publications have focused on the history of Central and Northern European populations, with little attention paid to southern Europe. The reason for this absence is because of the particularly warm climate conditions that tend to accelerate the degradation process of aDNA samples. Importantly, the Maltese work we are reporting here is the genetic analysis of one of the most southerly archipelagos of the Mediterranean. Specifically, we obtained uniparental genetic data (mitochrondrial DNA and Y-chromosome haplotypes) from 3 ancient individuals that lived in Malta during the transition between the Neolithic and Bronze Age periods. Thanks to this data we addressed the question of whether the Maltese were genetically more similar to Neolithic or to Bronze Age populations in Eurasia.

11.3. Methods

11.3.1. aDNA data collection and mitochondrial analysis
For this project we used data submitted by Ariano *et al*. (2022) from 3 petrous bones from the Circle. Reads obtained for each sample were aligned to the human reference genome (hg19/GRCh37). Both private and Haplogroup defining mutations were taken from the software **Haplofind** (Vianello *et al.* 2013) output. For each individual, these mutations were then used to measure the number of mismatches with the consensus *fasta* sequence. The contamination rate was calculated as the ratio of the number of mismatches over the total count of positions in the consensus sequence. When the mismatches included deaminated bases, these were counted as an upper limit value of contamination. Fastq files were aligned to the human Revised Cambridge Reference Sequence, (rCRS, NC_012920.1) using the tool **mpileup** from the software **samtools** (Li *et al.* 2009). Only SNP calls with a base quality above 30 (parameter -Q30) were then retained for further analyses. The genome coverage of each sample was calculated using the tool **qualimap** (Okonechnikov *et al.* 2016). A consensus mitochondrial Fasta sequence was first obtained for each sample using **bcftools** software (Li *et al.* 2011) (parameter -c) and then given to the software Haplofind (Vianello *et al.* 2013) for the haplogroup assignment (Table 11.2). From this analysis, we considered as valid only the haplogroups that were at the most terminal part of a branch and had an assignment score of at least 0.9 and where the assignment did not derive from a transition SNP.

11.3.2. Contamination
There are two common ways of checking for sample contamination in ancient DNA samples; the first method consists of checking for the presence of molecular damage at the 5′ and 3′ end of aligned reads. The second method is used also in modern DNA analyses and involves checking for the haploid state of the mitochondrial and X-chromosome DNA in male individuals. Given that all our samples were already treated for *postmortem* damages, we concentrated upon this last method for our contamination analyses.

11.3.3. Y-chromosome haplogroup determination
Samples that were identified as male were evaluated for Y-chromosome haplogroup lineage. This task was executed using the software Yleaf v2 (Ralf *et al*. 2018) and the ISOGG (International Society of Genetic Genealogy) 2019 database as reference (https://isogg.org/tree/ISOGG_YDNA_SNP_Index.html). SNPs annotated with the '~' label were excluded from this analysis (Table 11.3).

11.3.4. Collection of publicly available data
To contextualize our haplogroup results with other published ancient samples, we downloaded a well curated dataset of ancient DNA metadata from AmtDB (Ehler *et al.* 2018). We then used this resource to compare the geographical distribution of all sample haplogroups (both mitochondrial and Y-chromosome), focusing in particular on Neolithic, and Bronze Age periods. The samples were finally filtered for latitude and longitude thus restricting our analysis to Eurasia.

11.4. Results

11.4.1. Mitochondrial contamination and history

A common method for estimating DNA contamination of a sample is to check the rate of heterozygous sites present in the mitochondrial DNA. The contamination percentages of our high coverage samples, not considering sites that can derive from transition, range from values of 0.3% to 0.78% (Table 11.1). These values can be considered as acceptable for a no-contamination hypothesis. Once assured about the quality of our samples, we used the software **Haplofind** to investigate mitochondrial haplogroups, with the following results (Table 11.2):

- MLT5 belongs to the haplogroup K1a which is a subgroup of the major branch K. This branch has already been described in individuals that come from Anatolia during the Pottery and pre-Pottery Neolithic period (Mathieson *et al.* 2015).
- The individual MLT6 belongs to the haplogroup V which, although low in frequency, has been found in populations from central Europe associated with LBK, Únětice and Pitted ware culture, and from Neolithic populations in Portugal (Haak *et al.* 2015).
- MLT9 belongs to the haplogroup H4a1, which is a derived branch of haplogroup H. This major group evolved first in the Near East during the Neolithic period and afterward spread into western Europe (Torroni *et al.* 1998). It appears in fact to be frequent in France during Middle Neolithic period and Iberia during the Epi-Cardial Neolithic period.

By inspecting the distribution of ancient haplogroups, it appears that the Maltese belonged to mitochondrial branches that were particularly widespread during the Neolithic period. Interestingly, samples that matched the Maltese haplogroups during the Bronze Age period (details in Fig. 11.2) tended to come from central Europe and the Iberian Peninsula and belonged to the Bell Beaker culture.

11.4.2. Y-chromosome contamination and lineages

The results from Y chromosome screening indicate that two of our samples (MLT5 and MLT9) were male. We then used SNP information from the ISOGG database to define haplogroups and we found that the two individuals each belonged to one of two common European Neolithic haplogroup branches. MLT5 belongs to haplogroup H2. This haplogroup is rarely found in modern European populations and its earliest evidence dates back to a pre–pottery sample

Table 11.1. *Results from the contamination analysis. No sample shows significant traces of contamination, both excluding and including Transition sites (MD).*

Sample ID	Mean coverage	Site contamination %	Site contamination no-MD %
MLT5	128.26	1.422	0.533
MLT6	106.8	1.548	0.787
MLT9	184.87	0.563	0.340

Table 11.2. *Haplogroup assignment from Haplofind. The assignment score gives a probability of a sequence to be part of an haplogroup. The Haploscore gives an assignment score taking into account the previous major haplogroup from the same branch.*

Sample ID	Mitochondrial coverage	Haplo-group	Haplo-score	Assignment score
MLT5	128.26	K1a	0.8	0.96
MLT6	106.8	V	1	0.98
MLT9	184.87	H4a1	1	0.99

Table 11.3. *Sex assignment for each sample. When a sample did not reach a sufficient confidence interval it is indicated as 'Not Assigned'. For male individuals also the Haplogroup is assigned using the ISOGG database as reference.*

Sample ID	Only ChrY	Ratio ChrY/ChrY+ChrX	SE	95% CI	Sex assignment	Haplogroup
MLT5	208312	0.1162	0.0002	0.115-0.116	Male	H2
MLT6	43469	0.0178	0.0001	0.017-0.018	Not assigned	-
MLT9	177879	0.1224	0.0003	0.121-0.122	Male	G2a2a1a3

Table 11.4. *Values associated with contamination level using the X chromosome in male individuals.*

Sample ID	Contamination %	SE	P-value
MLT5	0.6	0.0014	6.789e-11
MLT9	1.1	0.0017	1.128e-08

in the Levant between 7300–6750 BC (Lazaridis *et al.* 2016). In more recent times this haplogroup was found in an Anatolian farmer and a European Neolithic sample belonging to the Starcevo culture. MLT9 has the haplogroup G2a2a1a3, one of the subclades of the major branch G commonly present in Europe during the Neolithic period (Broushaki *et al.* 2016). From examination of the incidence these haplogroups in ancient Eurasia, their prevalence during the Neolithic period compared with later times is clear (details in Fig. 11.3). There is a trend for matches to follow a more southern distribution. In the post-Neolithic comparison, only two H2 matches were found, in an Early Bronze Age sample from Bulgaria. Haplogroup G2a2a1a3 was interestingly found in 3 samples from Neolithic-Copper Age in Spain and Portugal. Other close subclades are common among Early European farmers and rarely feature in the Bronze period sample where they are mostly replaced by haplogroups R1a and R1b (Haak *et al.* 2015).

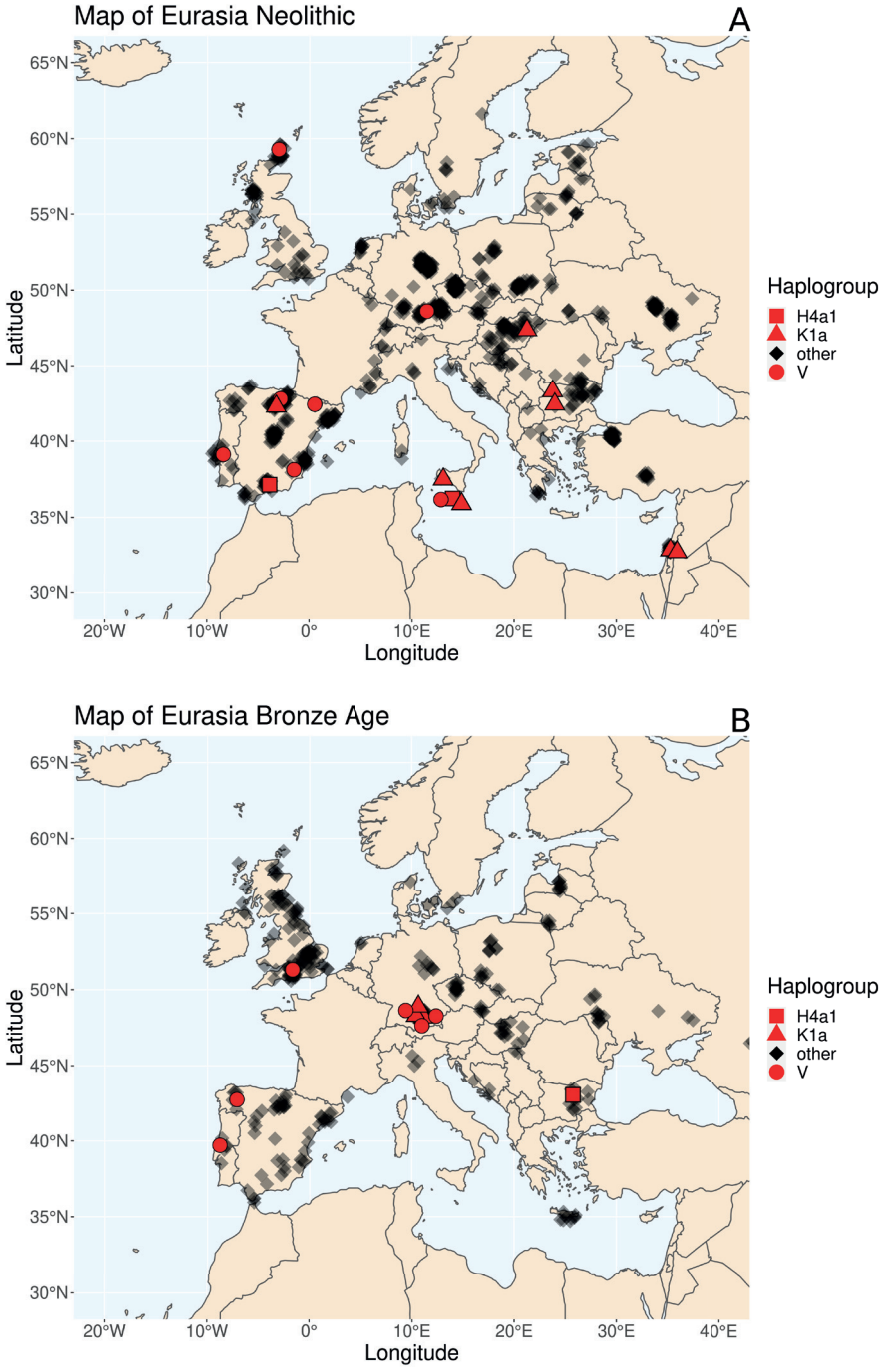

Figure 11.2. *Distribution of ancient mitochondrial haplogroups in Eurasia. Each point is a sample with the shape representing the haplogroup to which it belongs. A red colour indicates a match with one of the Maltese haplogroups encountered in this work, dark grey points show the geographical distribution of unmatched samples. Panel A: distribution of haplogroups during the Neolithic. Panel B: distribution of haplogroups in Bronze and Iron Age samples.*

11.5. Discussion

Mitochondrial DNA and Y-chromosome sequences from Neolithic Maltese individuals from the Temple Period (3rd millennium cal. BC) were analysed. Y chromosome haplogroup information showed that MLT5 and MLT9 are both part of Neolithic haplogroups common during the Neolithic period. Interestingly the MLT9 haplogroup was also found in samples from Copper Age Iberia pointing to a possible connection with the Cardial culture. These haplogroups almost disappeared during the Bronze Age and Iron Age periods with the only three matches found in Bronze Age individuals from Eastern Europe and Central Asia. Mitochondrial haplogroups results mirrored these findings with samples that matched the Maltese mostly as Neolithic farmers and Bell Beaker samples from Western Europe. Both these results point to a Western European Neolithic or Bell Beaker ancestry of our ancient Maltese and we believe further analysis

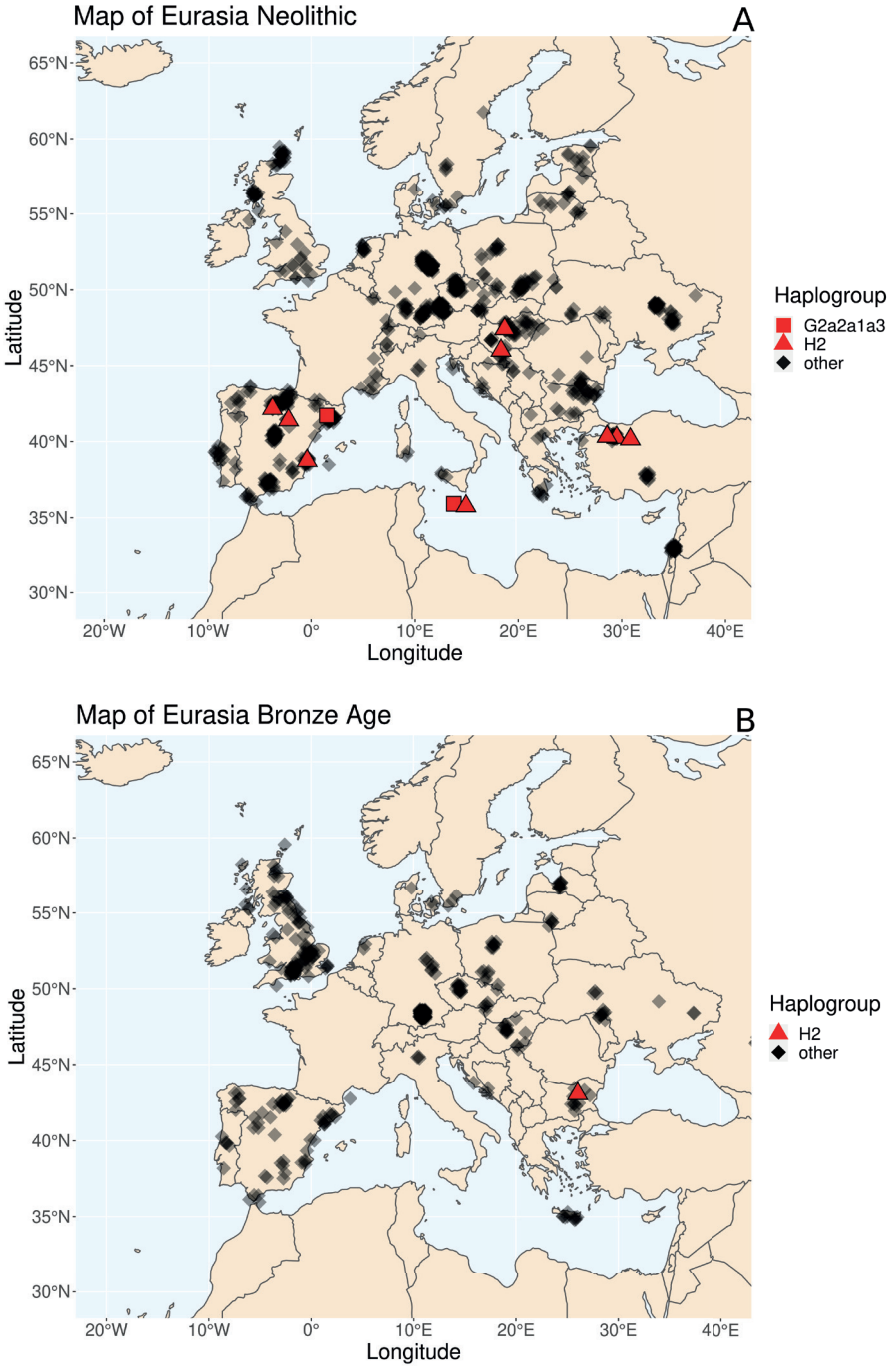

Figure 11.3. *Distribution of ancient Y haplogroups in Eurasia. Each point is a sample with the shape representing haplogroup. A red symbol indicates a match with one of the Maltese haplogroups encountered in this work. Panel A: distribution of haplogroups during Neolithic. Panel B: distribution of haplogroups in Bronze Age samples.*

of autosomal markers will clarify and refine estimates of their ancestry.

11.6. Conclusion

The populations of the Maltese islands, located in the south of the Mediterranean Sea, were shaped by a succession of different cultures during the Neolithic period. The first group settled on the islands just after 6000 BC, probably as an Early Neolithic population. After an initial oscillation between growth and decline (see Volume 1, Chapter 2) an apogee of culture and population density was reached during Temple Period, especially in the Tarxien phase between *c*. 2800 and 2400 BC, which saw the construction of unparalleled sophisticated megalithic structures. Then this culture seemingly collapsed, and a number of questions have vexed scholars of early Malta ever sense: who were these ancient inhabits of Malta, and which ancient population did they resemble the most? To answer these questions, we offer here a first assessment of Maltese ancient DNA data using three individuals that lived during the Tarxien phase of the Temple Period.

11.6.1. The Neolithic routes

The culture of Neolithic farming spread from northwest Anatolia into western Europe following two main routes. One route was associated with the Linearbandkeramik culture (LBK) and followed the Danube valley and spread northwest towards northern Europe. The other route was associated with Impressa-Cardial pottery culture and followed a westward Mediterranean route reaching the Atlantic in France and Iberia. Malta's early settlers were likely part of this latter route with their uniparental markers resembling other southern European Neolithic samples most strongly.

11.6.2. The eastern influence

By the 2nd millennium BC, the Bronze Age period populations from the steppe migrated from eastern to western Europe, displacing preceding local cultures (Olalde *et al*. 2018). Exotic pottery coming from eastern Europe, even before the Bronze Age period, could suggest a connection between the Maltese and other populations (for example, Thermi, Bell Beakers and the potential Balkan Cetina style). No genetic evidence in our samples implies contact with eastern populations.

11.7. Future perspectives

The field of ancient DNA study is in continuous development, especially as the financial cost of sequencing analysis reduces. Although haploid lineage markers can give hints about ancestry, using autosomal markers will help us to answer more important questions about migration and admixture. Therefore, our first next step will be to deepen our investigations by using methods to detect admixture, kinship and population structure from autosomal markers.